PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A randomized controlled pilot and feasibility study of multimodal agility-based exercise training (MAT) versus strength and endurance training (SET) to improve Multiple Sclerosis-related fatigue and fatigability during inpatient rehabilitation [ReFEx] – study protocol.
AUTHORS	Wolf, Florian; Nielsen, Jörn; Saliger, Jochen; Hennecken, Eva; Eschweiler, Mareike; Folkerts, Ann-Kristin; Karbe, Hans; Zimmer, Philipp

VERSION 1 – REVIEW

REVIEWER	Seebacher, Barbara Medical University of Innsbruck, Clinical Department of Neurology
REVIEW RETURNED	20-Mar-2022

GENERAL COMMENTS	A randomized controlled pilot and feasibility study of multimodal agility-based exercise training versus strength and endurance training to improve Multiple Sclerosis related fatigue and fatigability during inpatient rehabilitation [ReFEx] – study protocol Dear authors, I enjoyed reading your manuscript as it is well written, and the planned study addresses a relevant topic. I appreciate the detailed intervention manual being attached as a supplement as it allows the reproduction of the intervention. There are some points which I would like to draw your attention to, however. Please find my comments as follows. Major The planned study intervention will comprise either SET or land- and water-based MAT for 4 to 6 weeks during inpatient rehabilitation. Regarding the comparability of the interventions, one could expect that a longer intervention period may lead to more favourable results. This aspect should be described in more detail or considered in the Study limitations section. Randomisation: Stratification will be performed according to the EDSS score, WEIMuS score, age and MS disease course i.e., for recruiting 24 patients, there will be 16 strata. In my opinion, this is not at all feasible, particularly considering that gender is a significant predictor with respect to a change in motor function; it is further expected that there will be a difference with strength and endurance training between females and males. Please clarify. Inclusion criteria: Please justify why people with primary progressive MS will be excluded from the trial. Considering that there is a lack of drug treatment for this population the exclusion does not seem ethical. Intervention: The MAT group will receive a greater amount of therapy than the SET group (160-240 minutes in total). I suppose the reasoning behind this choice is one of practicability. It seems important to take the time difference into account, either by changing

	the approach or by stating this as a study limitation. For the main trial, this should be adjusted in any case and hence the pilot and main study data not pooled. Specifically trained exercise therapists will supervise the MAT intervention whereas strength training sessions will be supervised by exercise science students or therapists. Based on the diverging education levels of postgraduate therapists and students I wonder whether this could influence the results. Please clarify. Table 4: Please reorganise the table according to logical criteria e.g., eligibility rate first, followed by the recruitment rate. Please replace eligibility criteria by eligibility rate. If the eligibility and recruitment rates are known information on the refusal rate is somewhat redundant, right? I suggest formulating the terms in in the same direction i.e., sticking to retention rates instead of mixing up those with drop-out rates as this may be confusing for the reader. As for the treatment effect, please see my further comments with the Statistical analysis section. Please revise the last requirement as it is phrased negatively, contrasting to all other requirements. Please include references for all outcome measures if they are not already provided. Statistical analysis It has been stated that depression will be considered a confounder variable, but I could not find a description in the Statistical analysis section. Furthermore, with such a small sample size, the feasibility outcomes and effect size calculations appear useful. Estimation of preliminary results is justifiable but needs to be done employing appropriate statistical tests. Continuous data should be tested for a normal distribution and other test assumptions need to be examined. Performing paired t-tests followed by independent t-tests from change scores are not appropriate and should be replaced by a Mixed Model (mixed Design Factorial ANOVA), followed by post-hoc tests correcting for multiple comparisons if the overall analyses yield significant results. For ordinal scale data, the proposed approach seems useful, replacing t-tests by Mann Whitney-U Tests, followed by post hoc tests as described. With the very small sample size, the results are likely to be non-significant, however. For ordinal scale data, please clearly describe the tests used as these data should not be treated equal to continuous or nominal data. This also refers to the calculation of effect sizes – please specify the effect sizes estimated. Cohen's d (95% CI) may only be used for continuous data. Qualitative data analysis: Could you please describe the deductive coding in more detail i.e., on what preliminary considerations and on which specific hypotheses in the study planning it is based. Please also give 2-3 examples of reviews of relevant literature you will base your deductive coding upon. Discussion: Analysis of blood-based biomarkers is planned to be part of the main ReFEx trial but not the pilot study. This appears to be another reason for not pooling the data of the pilot study with those of the main trial. In addition, land- and water-based MAT are substantially different approaches, with likely varying mechanisms of action. Please include some discussion on that, including implications for the study. Minor: Abstract: please replace "conductance" by "conduct" throughout the manuscript; land- and water-based (instead of land-based and water-based). Subjective fatigue will also be assessed at 1, 4, and 12 weeks after discharge (include "at"). Funding: "This work was supported by the …" should be changed to "This work is supported by the…".

	Manual for the land-based and water-based MAT: “Transport chain”: Please replace “visiual” by “visual”; “Remote control”: Please replace “clsoed” by “closed”.

REVIEWER	Sessford, James Toronto Rehabilitation Institute - Lyndhurst Centre
REVIEW RETURNED	25-Mar-2022

GENERAL COMMENTS	This is a feasibility study comparing traditional strength and endurance training against balance training combined with agility training in people with multiple sclerosis in an inpatient setting. Primary outcomes of interest are subjective and objective fatigue/fatigability. Assessments of feasibility will also be examined using quantitative and qualitative means. Abstract: Pg 2, line 36. The way it is written, it could be interpreted that each individual will receive MAT as well as SET. Try changing “with” to “against” to clarify that MAT will be compared against SET. Also, is it MAT that is being compared or BMCT+MAT? Pg 2, line 38. “With the conductance of the ReFEx….” Conductance is a term relating to electrical currents, try “By conducting the ReFEx…”. This could also be changed in the top right field of Table 4. Pg 2, line 40-41. Point/objective (III) is unclear to me. I’m also surprised that the comparison between MAT (or BMCT + MAT) and SET is not listed as an objective. Abstract does not outline any analytical plan. Intro: Pg 5, line 102. Can the reasons/mechanisms for BMCT being described as “promising” be elaborated upon? In the hypotheses section, the use of MAT is confusing because it seems to be that the study will be looking at BMCT+MAT. So, it would be hard to isolate the effect of MAT alone from the effects of BMCT. Should the terminology throughout the abstract and paper better reflect that one group receives SET and the other receives BMCT+MAT? Methods In the methods, could repetitions maximum (RPM) be change to RM? That is more common in exercise vernacular and RPM can be confused with revolutions per minute, such as in cycling speed. Pg 13, line 286. Perhaps change “regenerate” to “recuperate”? I think it would be useful for many of the questionnaires used, to offer example items in the text. Data analysis (pg 20) refers to sociodemographic outcomes but I don’t see those variables clearly laid out anywhere. What is the rationale for multiple t-tests as opposed to methods that we be more robust to protecting against type 1 error such as MANOVA and repeated measures MANOVA? It's appreciated that effect sizes will be presented.

VERSION 1 – AUTHOR RESPONSE

Dear Dr. Barbara Seebacher,

Thank you very much for the very helpful comments on our manuscript. We revised the manuscript accordingly.

1. The planned study intervention will comprise either SET or land- and water-based MAT for 4 to 6 weeks during inpatient rehabilitation. Regarding the comparability of the interventions, one could expect that a longer intervention period may lead to more favourable results. This aspect should be described in more detail or considered in the Study limitations section.

We agree with the reviewer's point of view that a longer intervention period would likely result in more favorable results. However, as the setting of the study will be an inpatient clinic, a longer intervention period will not be possible, as this is predetermined by the national health care system in Germany. We added this statement in the text (page 24, line 545-47). Nevertheless, the (main) study will be able to make a statement on what to expect from an inpatient rehabilitation stay under the current conditions with regard to the targeted outcomes.

2. Randomisation: Stratification will be performed according to the EDSS score, WEIMuS score, age and MS disease course i.e., for recruiting 24 patients, there will be 16 strata. In my opinion, this is not at all feasible, particularly considering that gender is a significant predictor with respect to a change in motor function; it is further expected that there will be a difference with strength and endurance training between females and males. Please clarify.

As described in the manuscript (page 8, line 212) the minimization procedure will be applied. As defined in reference 23: “Unlike stratified randomization, minimization works toward minimizing the total imbalance for all factors together instead of considering mutually exclusive subgroups. […] The decision to allocate to the treatment or control group can be made by comparing the totals for each

category and choosing the group that gives most balance overall." Consequently, there are no 16 strata, as the reviewer assumed.

Previous trials conducted by our group did not show that gender was a significant predictor for a change in motor function in pwMS. Previous trial data also does not indicate that gender is a significant predictor for responding to endurance training in pwMS (Schlagheck et al., 2021, DOI: 10.1055/a-1481-8639). However, there certainly are differences in disease characteristics between men and women, so we will include gender as a stratification factor in a full trial. With 15 participants recruited, there is one male in each group right now.

3. Inclusion criteria: Please justify why people with primary progressive MS will be excluded from the trial. Considering that there is a lack of drug treatment for this population the exclusion does not seem ethical.

As described in the discussion section (page 25 line 555-61), analysis of blood-based biomarkers is planned for a full trial. This would focus on the kynurenine pathway, which has been shown to differ between patients with RR/SPMS and PPMS (Lim et al., 2017, DOI: 10.1038/srep41473). For this reason, we also decided to exclude persons with PPMS in this pilot/feasibility study.

4. Intervention: The MAT group will receive a greater amount of therapy than the SET group (160-240 minutes in total). I suppose the reasoning behind this choice is one of practicability. It seems important to take the time difference into account, either by changing the approach or by stating this as a study limitation. For the main trial, this should be adjusted in any case and hence the pilot and main study data not pooled.

It is correct that there will be a difference in the amount of therapy time, which has practical reasons, as you suggested. In total, and as described in Table 3, MAT will receive 240min and SET will receive 200min per week (+150min 'MS-group' each). (We do not know how you calculated the 160min.) After the feasibility phase, we will analyze adherence to and fidelity of the interventions. Based on this, we will decide whether changes to the intervention schedules are necessary.

5. Specifically trained exercise therapists will supervise the MAT intervention whereas strength training sessions will be supervised by exercise science students or therapists. Based on the diverging education levels of postgraduate therapists and students I wonder whether this could influence the results. Please clarify.

The reviewer is correct at this point. However, as stated in line 282-84 (page 11), exercise science students are also specifically trained to perform thestrength training sessions and a manual is also provided. Furthermore, the strength training takes place in a one-on-one setting, making it easier to give instructions and communicate than in a group format. Finally, as mentioned, some strength sessions are also led by postgraduate therapists (page 11, line 281). So far, we have had no problems with students conducting the strength training sessions, as evidenced by analysis of training logs.

6. Table 4: Please reorganise the table according to logical criteria e.g., eligibility rate first, followed by the recruitment rate. Please replace eligibility criteria by eligibility rate. If the eligibility and recruitment rates are known information on the refusal rate is somewhat redundant, right? I suggest formulating the terms in in the same direction i.e., sticking to retention rates instead of mixing up those with drop-out rates as this may be confusing for the reader. As for the treatment effect, please see my further comments with the Statistical analysis section.

We followed the reviewer's suggestions and reorganized Table 4 as requested and replaced the term *eligibility criteria* with *eligibility rate*. The *refusal rate* gives explicit information on how many patients refused to participate despite being fully eligible. We do not consider this redundant, as the recruitment rate only provides information on the number of successfully randomized patients. We reduced the wording ‘drop-out rates’ in the table and the text.

7. Please revise the last requirement as it is phrased negatively, contrasting to all other requirements.

We have followed the reviewer's recommendation and rephrased the last progression requirement (page 23, line 501-02).

8. Please include references for all outcome measures if they are not already provided.

Thank you for drawing attention to the missing references, especially regarding the secondary outcome measures. We have added the missing references where appropriate (page 20, line 419-27).

9. Statistical analysis: It has been stated that depression will be considered a confounder variable, but I could not find a description in the Statistical analysis section.

Yes, that's right, we suspect that depressive mood as assessed with the CES could be a relevant confounder. However, depression will only be added to the statistical analysis in the full trial.

10. Furthermore, with such a small sample size, the feasibility outcomes and effect size calculations appear useful. Estimation of preliminary results is justifiable but needs to be done employing appropriate statistical tests. Continuous data should be tested for a normal distribution and other test assumptions need to be examined. Performing paired t-tests followed by independent t-tests from change scores are not appropriate and should be replaced by a Mixed Model (mixed Design Factorial ANOVA), followed by post-hoc tests correcting for multiple comparisons if the overall analyses yield significant results. For ordinal scale data, the proposed approach seems useful, replacing t-tests by Mann Whitney-U Tests, followed by post hoc tests as described. With the very small sample size, the results are likely to be non-significant, however.

The reviewer is correct when stating that, given the small number of cases in the feasibility study, no significant results are to be expected. For this reason, we decided not to perform the significance tests for the effects within and between groups and rewrote the paragraph (page 21-22, line 444-61). The main goal of the feasibility study is to determine whether the adapted MAT and SET are

feasible in the inpatient rehabilitation setting with a special emphasis on patients' acceptance (as stated in the introduction, page 6, line 164-67).

11. For ordinal scale data, please clearly describe the tests used as these data should not be treated equal to continuous or nominal data. This also refers to the calculation of effect sizes – please specify the effect sizes estimated. Cohen's d (95% CI) may only be used for continuous data.

For the same reasons, we decided against the calculation of effect sizes and described this in the analysis section (page 22, line 460-461).

12. Qualitative data analysis: Could you please describe the deductive coding in more detail i.e., on what preliminary considerations and on which specific hypotheses in the study planning it is based. Please also give 2-3 examples of reviews of relevant literature you will base your deductive coding upon.

The coding of the interviews follows the procedure of qualitative content analysis, using a combined model of deductive (a priori) and inductive coding (on the resulting text material) to identify themes and subthemes. In this regard, deductive coding is based on the two research guiding questions of this feasibility study, namely to what extent the study design can be implemented in the patients' rehabilitation setting and how patients experience the interventions. Accordingly, we aim to capture patients' subjective perspectives on acceptance, benefits, and satisfaction with study participation and also to consider patients' experiences with the respective study therapies. To this end, we designed a semi-structured interview based on the categories described in line 380-82 (page 19). For these categories, we formulated questions such as: 1. How do participants describe their fatigue? 2. What do MAT participants think about training in a group setting? 3. How are individual therapy sessions perceived to be exhausting (physically, cognitively?)? 4. Are there specific things that participants enjoy or do not enjoy about the individual therapy sessions? 5. What is the acute impact of the sessions on participants' feelings of fatigue (physical, cognitive?), including time to recover? 6. What

participants thought about their improvement in balance and other symptoms and to what they personally attributed this improvement, 7. Whether they were satisfied with their group assignment. This procedure was guided by relevant literature on (qualitative) feasibility studies in general (O´Cathain et al., 2015; doi: 10.1186/s40814-015-0026-y; Orsmond & Cohn, 2015; doi: 10.1177/1539449215578649) and in pwMS in particular (Hersche et al., 2019; doi.org/10.1016/j.msard.2019.06.034; Gunn et al., 2017; doi: 10.1186/s40814-017-0168-1; Smith, Williams, Barker, 2020; doi: 10.1136/Bmjopen-2019-035378; Learmonth, Kinnett-Hopkins, Motl, 2018; doi.org/10.1080/09638288.2018.1490823) .

13. Discussion: Analysis of blood-based biomarkers is planned to be part of the main ReFEx trial but not the pilot study. This appears to be another reason for not pooling the data of the pilot study with those of the main trial.

Blood-based biomarkers would be considered exploratory outcomes in the main trial. As such, a smaller sample size would be acceptable. We do not anticipate that the blood draws will have any other substantial impact on the results of the main trial.

14. In addition, land- and water-based MAT are substantially different approaches, with likely varying mechanisms of action. Please include some discussion on that, including implications for the study.

We agree with this statement and have included some discussion on this issue in the discussion section (page 24-25, line 547-554).

15. Minor issues:

- Abstract: please replace “conductance” by “conduct” throughout the manuscript;
- land- and water-based (instead of land-based and water-based).
- Subjective fatigue will also be assessed at 1, 4, and 12 weeks after discharge (include “at”).

- Funding: “This work was supported by the …” should be changed to “This work is supported by the…”.
- Manual for the land-based and water-based MAT: “Transport chain”:
- Please replace “visiual” by “visual”; “Remote control”: Please replace “clsoed” by “closed”.

We have followed all the reviewer's suggestions and have corrected or adjusted the wording, expressions, and terms accordingly.

Dear Dr. James Sessford,

Thank you for the many comments on our manuscript. We have taken all the comments into account and have revised the manuscript accordingly and hope that we have been able to remove the ambiguities about the content of the intervention group.

1. Abstract: Pg 2, line 36. The way it is written, it could be interpreted that each individual will receive MAT as well as SET. Try changing “with” to “against” to clarify that MAT will be compared against SET. Also, is it MAT that is being compared or BMCT+MAT?

We agree with the reviewer that the wording is misleading. We have rewritten the abstract to make it clearer, which interventions are being compared (page 2, line 36-41). MAT is being compared to SET. However, MAT contains elements of BMCT. Because MAT is a new framework to be applied in pwMS, we build upon the BMCT, as it has been described before in the context of MS and fatigue.

2. Pg 2, line 38. “With the conductance of the ReFEx….” Conductance is a term relating to electrical currents, try “By conducting the ReFEx…”. This could also be changed in the top right field of Table 4.

We have followed the reviewer's comments and corrected these terms accordingly.

3. Pg 2, line 40-41. Point/objective (III) is unclear to me.

We have deleted the objectives in the abstract as they added to the confusion related to MAT and BMCT. The aims of the ReFEx project (page 5, line 125-36) and the feasibility study (page 6, line 164-67) are stated in the introduction.

4. I'm also surprised that the comparison between MAT (or BMCT + MAT) and SET is not listed as an objective.

See above.

5. Abstract does not outline any analytical plan.

The reviewer's criticism is valid; therefore, we have added a brief statement on quantitative data analysis in the abstract, noting that all of our data analyses involve primarily descriptive statistics (page 2, line 50).

6. Intro: Pg 5, line 102. Can the reasons/mechanisms for BMCT being described as "promising" be elaborated upon?

We followed the reviewer's recommendation and added a phrase to explain why BMCT was considered promising by Moss-Morris and colleagues (please see page 5, line 117).

7. In the hypotheses section, the use of MAT is confusing because it seems to be that the study will be looking at BMCT+MAT. So, it would be hard to isolate the effect of MAT alone from the effects of BMCT. Should the terminology throughout the abstract and paper better reflect that one group receives SET and the other receives BMCT+MAT?

As mentioned earlier, MAT is being compared to SET. However, MAT contains elements of BMCT. We have reworded both, the abstract, and the text in line 125-32 (page 5) to avoid confusion.

8. Methods: In the methods, could repetitions maximum (RPM) be change to RM? That is more common in exercise vernacular and RPM can be confused with revolutions per

minute, such as in cycling speed.

Pg 13, line 286. Perhaps change “regenerate” to “recuperate”?

Thank you for these comments, we have followed the comments and corrected these terms and designations accordingly.

9. I think it would be useful for many of the questionnaires used, to offer example items in the text.

We fully agree, but in view of the limited number of words, we will not include example items in the text.

10. Data analysis (pg 20) refers to sociodemographic outcomes but I don’t see those variables clearly laid out anywhere.

The reviewer's criticism is valid; we added a paragraph on baseline sample characteristics in the outcomes section (page 14, line 350-61).

11. What is the rationale for multiple t-tests as opposed to methods that we be more robust to protecting against type 1 error such as MANOVA and repeated measures MANOVA? It's appreciated that effect sizes will be presented.

We reconsidered our (quantitative) data analysis plan and decided to drop the significance testing of the within- and between-group effects in light of the small sample size and likely meaningless output. For the same reasons we decided against the calculation of effect sizes and described this in the analysis section. The main goal of the feasibility study is to determine whether the adapted MAT and SET are feasible in the inpatient rehabilitation setting with a special emphasis on patients’ acceptance (as stated in the introduction, page 6, line 164-67).

VERSION 2 – REVIEW

REVIEWER	Seebacher, Barbara Medical University of Innsbruck, Clinical Department of Neurology
REVIEW RETURNED	06-Jun-2022

GENERAL COMMENTS	I appreciate the changes you have made to your manuscript and

	your well-considered responses. Good luck with your research!

REVIEWER	Sessford, James Toronto Rehabilitation Institute - Lyndhurst Centre
REVIEW RETURNED	13-Jun-2022

GENERAL COMMENTS	Thank you, I believe the manuscript is very strong and all major questions have been adequately addressed. The following few issues may be adequately addressed by copy-editing at the discretion of the editor but I would ask these points: -Should “Akronym” be switched to “acronym” under title on title page? -Should the trial status update be more current at this point than Nov 2021? - Under Baseline sample characteristics “Height will be ascertained from participants” - This is unclear if it means height will be measured or self-reported by participants.