PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Deep brain stimulation for locomotion in incomplete human spinal cord injury (DBS-SCI) – protocol of a prospective one-armed multi-centre study
AUTHORS	Stieglitz, Lennart; Hofer, Anna-Sophie; Bolliger, Marc; Oertel, Markus; Filli, Linard; Willi, Romina; Cathomen, Adrian; Meyer, Christian; Schubert, Martin; Hubli, Michèle; Kessler, Thomas; Baumann, Christian; Imbach, Lukas; Krüsi, Iris; Prusse, Andrea; Schwab, Martin; Regli, Luca; Curt, Armin

VERSION 1 – REVIEW

REVIEWER	Winfried Mayr Medical University of Vienna
REVIEW RETURNED	01-Mar-2021

GENERAL COMMENTS	The manuscript describes long lists of standard test and definitions in a clear and well readable manner. Also, parts of the study are explained in details, but essential parts are missing and critical discussion of the study design is not sufficient in the provided version. Main shortcomings are the absence of clear strategies for setting stimulation parameters in the intraoperative testing and the application through the observation period, the lack of coordinative principles for comparable movement rehabilitation programs for the preparation and the intervention period, and the non-consideration of medication, in particular anti-spasticity medication. To adjust a stimulation setup, in general and in particular in the concrete application, is a multivariant problem with essential influence of electrode position, configuration (active contacts), polarity, intensity (pulse width and amplitude) and frequency. To find an optimum requires clear strategies and criteria for the initial electrode placement under response recording (page 11: “constant-frequency stimulation performed” – frequency strongly influences neural processing and responses), intervention setup (page 17: cannot just be trial and error for “optimal setting”, but requires clear and comparable criteria and systematic variation). Rehabilitation training (same page) requires standardization, at least in clear strategies for individualized programs, to get comparable study results. Not at least medication needs to be considered, e.g. anti-spasticity medication has strong influence on augmentation or inhibition in motor control and requires consideration in the interpretation of study results. The primary endpoint 6MWT is a simple integrative criterion, but the huge amount of additional assessment procedures listed potentially contain the relevant information for scientific interpretation of the outcome. These assessments should follow a clear protocol for minimizing the burden for the study participants and avoid bias on results through long assessment sessions. Expected outcome

	should be considered in the discussion section. The one subject, that has already passed the study should be presented in more detail. Obviously, the expected results have not been accomplished in the primary endpoint, but there must be more results in the many secondary assessments and those should be considered relevant for further planning. Just to apply for an earlier start in the subacute recovery phase will rather lead to a less clear separation of regeneration under standard treatment and intervention driven gain. This should be avoided in a small pilot study with just 5 participants with already inevitable hard to compare lesion profiles.

REVIEWER	Zhou Li Chinese Academy of Sciences, Institute of Nanoenergy and Nanosystems
REVIEW RETURNED	30-May-2021

GENERAL COMMENTS	In the manuscript entitled “Deep brain stimulation for locomotion in incomplete human spinal cord injury (DBS-SCI) – protocol of a prospective one-armed multi-centre study”, the author introduced a kind of protocol which could be used for deep brain stimulation. Besides, this work has sorted out the monitoring indexes of MLR-DBS in detail, which has certain guiding significance for the future work. However, the reviewer suggests that the manuscript requires few revisions before publication. 1. After presenting a large number of methods that can be used for evaluation, the author should further discuss the advantages and disadvantages of each method and why 6MWT is chosen in the discussion part. 2. Formatting issues such as font spacing/case should be corrected. 3. In recent years, there are many excellent works on deep brain stimulation and new devices for that, such as: Full activation pattern mapping by simultaneous deep brain stimulation and fMRI with graphene fiber electrodes, Nature communications, 2020, 11(1): 1-12. ; Electrical Stimulation for Nervous System Injury: Research Progress and Prospects, Acta Physico-Chimica Sinica, 2020, 36 (X), 2005038; Recent Development of Implantable and Flexible Nerve Electrodes, Smart Materials in Medicine, 2020, 1, 131-147; which are suggested to be cited.

REVIEWER	Luka Milosevic University Health Network
REVIEW RETURNED	01-Jun-2021

GENERAL COMMENTS	The authors present a first-in-main Phase I/II clinical trial protocol for assessing cuneiform deep brain stimulation (DBS) in patients with incomplete spinal cord injury (SCI) in a prospective cohort of five patients. This is a novel DBS indication, and this study provides a means of establishing safety and feasibility. Hopefully the investigators can also be successful in achieving clinically significant functional improvement. Major points to consider: 1) The small sample size is daunting if the authors hope to achieve reproducible clinical benefit, particularly if effect sizes are not large; which may be expected based on PPN-DBS literature in PD. 2) The introduction would benefit from a bit more detail regarding

	clinical and preclinical studies (more importantly, their findings) which have investigated CNF-DBS. What have the stroke/SCI studies shown, in particular? Additionally, is there rodent literature to suggest that CNF may be a better target than PPN, or perhaps the peripeduncular nucleus (e.g. PMID: 17525137)? 3) It is unclear as to how the battery of intraoperative physiological tests during intraoperative recordings will inform decision making regarding electrode position. Moreover, I could foresee issues with patient/muscle fatigue and compliance (15mm of recording with testing at each 0.5mm, including real/imagined movements, cognitive testing, side effect screening at various frequencies and intensities, online analyses of kinematic and physiological measures, etc). This battery seems rather infeasible (or at least very costly in terms of time). This aspect of the surgery can/should be refined, procedurally, as it is difficult to foresee an efficient "decision-making" pipeline in the operating theater. Which of these features (or perhaps which combination) might inform optimal location (or location within the CNF at all; particularly considering potential issues with image-based targeting)? Are there expected intraoperative neurophysiological markers of CNF in animal literature (e.g. characteristic oscillatory frequency, neuronal firing rate or pattern, expected evoked phenomena, etc)? Side note: how many microelectrodes trajectories will be used? Minor points to consider: Abstract - Major functional recovery plateaus (add:)_occur_ three to four months - please define AIS C – or is this meant to say SCI? Introduction - The number of spared, descending fibres – remove comma - might be worthwhile to acknowledge that that clinical trials for PPN-DBS in Parkinson's disease have not exactly yielded the most promising results, to date. This could further serve to justify target selection in this work. Methods Study population - how/why was n=5 decided upon? [the answer to this question arises towards the end of the Methods section – perhaps these sections can be consolidated] Inclusion and exclusion criteria - is there a rationale as to why there is no defined maximum timepoint since incurred SCI? could this result in variable outcome results? i.e. differences between patients who sustained SCI 6 months preoperatively vs. those who have sustained SCI at, for example, 20 years prior? With a small cohort, it might be favourable to constrain this inclusion criterion. Surgery - which hemisphere will be targeted and how is that decision made? - could consider implanting Medtronic Percept PC for chronic LFP recording capability and future closed-loop application. - details about postoperative electrophysiological assessments are provided below; however the intraoperative assessments are likely methodologically different. How are MEPs and SSEPs evoked intraoperatively? What is ERP in response too? These details should be clarified (especially to be able to be differentiated from postoperative measures). Also, what is the purpose of obtaining these measures intraoperatively? Electrophysiological assessments - for MEPs, "All measures are recorded before and after interventions" - what does intervention refer to? Will these

	measurements be done in one session, e.g. postoperatively in the stimulation naïve condition, then again after a period of stimulation? Or, before and after surgery (in which case the measurements may not exactly be comparable due to differences in EMG/TMS sites). Please elaborate on the protocol and what is meant by "intervention". DBS during behavioural testing and rehabilitative training - "In the first two weeks after lead implantation, different stimulation parameters (frequency, Hz; pulse width, µs; amplitudes, mV) are tested" – what about potential insertional/microlesion effects? Study Endpoints - authors could consider splitting the "Clinical assessments" subsection into "Primary endpoint" and "Secondary endpoints" subsections. Also, Table 2 seems a bit superfluous, as it seems it is just a summary of the list of clinical assessments; each of which reappear in Table 3. Sample size - authors should consider that most PPN-DBS studies in PD were n≤7, and most failed to find statistically (/clinically) significant results.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Winfried Mayr , Medical University of Vienna
Comments to the Author:
The manuscript describes long lists of standard test and definitions in a clear and well readable manner. Also, parts of the study are explained in details, but essential parts are missing and critical discussion of the study design is not sufficient in the provided version.
Main shortcomings are the absence of clear strategies for setting stimulation parameters in the intraoperative testing and the application through the observation period, the lack of coordinative principles for comparable movement rehabilitation programs for the preparation and the intervention period, and the non-consideration of medication, in particular anti-spasticity medication. To adjust a stimulation setup, in general and in particular in the concrete application, is a multivariant problem with essential influence of electrode position, configuration (active contacts), polarity, intensity (pulse width and amplitude) and frequency. To find an optimum requires clear strategies and criteria for the initial electrode placement under response recording (page 11: "constant-frequency stimulation performed" – frequency strongly influences neural processing and responses), intervention setup (page 17: cannot just be trial and error for "optimal setting", but requires clear and comparable criteria and systematic variation). Rehabilitation training (same page) requires standardization, at least in clear strategies for individualized programs, to get comparable study results. Not at least medication needs to be considered, e.g. anti-spasticity medication has strong influence on augmentation or inhibition in motor control and requires consideration in the interpretation of study results.

We agree with Prof. Mayr that the strategy for setting stimulation parameters during intraoperative testing and subsequently during training is highly relevant. As this is the first human study investigating stimulation of the cuneiform nucleus to induce locomotion, data on optimal stimulation parameters are currently lacking. Thus, we have to orient on the one hand on experience gained from preclinical studies, which suggest low (≤50 Hz) frequency stimulations with medium to broad pulse widths (200-1000 µs) in various mammalian species (e.g. Bachmann et al.[1] in rats, Opris et al.[2] in cats, Chang et al.[3] in pigs). On the other hand, we will test different stimulation parameters in all patients and establish efficacy and side effect profiles to ultimately identify a therapeutic window. However, so far we only have these data from one patient and thus cannot yet make any generalizable statement in this regard. Stimulation intensities or amplitudes required to induce locomotion are highly individual and depend on the chosen frequency and pulse width and thus cannot be predefined. We addressed this comment in the subsections "Surgery" (line 201-234 and

242-247) and “DBS during behavioural testing and rehabilitative training” (line 421-436) in the Methods section, and in the Discussion section (line 515-522).
We also share the opinion of Prof. Mayr that it would be ideal if rehabilitation programs prior to study inclusion where comparable. One limitation to achieving this is that we are recruiting patients internationally and thus cannot direct the rehabilitation schedules patients are completing. Additionally, we include patients that have already completed in-patient rehabilitation. After inclusion, study participants undergo identical assessments and trainings at the Balgrist University Hospital for two weeks in order to identify each patient’s stimulation settings and training capacity. This is followed by discharge either home or to a rehabilitation center located close to the patient’s home. After discharge, training intensity is monitored by regular follow-ups by phone and also by constant activity monitoring via wearable, wireless sensors (subsection “Long-term Monitoring of Physical Activity” in Methods section, line 286-290). While this does not ensure identical training intensity, it does allow us to give regular feedback and make adjustments as necessary. Nevertheless, a completely congruent training is not possible as chronic spinal cord injury patients, despite certain inclusion criteria, constitute a heterogeneous and complex patient group requiring individual treatment adapted to the respective needs. Furthermore, locomotion parameters, e.g. speed, stepping frequency, or hip height strongly depend on stimulation parameters used.[1,4] Since parameters required for induction of stepping movements are individual, training intensities cannot be completely identical. This comment was addressed in subsection “DBS during behavioural testing and rehabilitative training” (line 433-436) in the Methods section and in the Discussion section (line 522-534).
Thank you very much for raising the point about anti-spasticity medication. We record each patient’s medication as part of the screening and study inclusion process, including of course anti-spasticity medication. However, while anticoagulant drugs are paused prior to surgery, we are not intending to change already established symptomatic drug regimens, such as anti-spasticity therapy, as part of the study as we initially want to investigate whether the stimulation itself enables gait training and leads to functional improvement. It is true that different drug regimens are a potential confounder that complicates data analysis. However, we do not aim to compare the functional changes between patients, but rather to examine the functional changes within each patient over the course of the study. In case changes to established drug regimens, e.g. anti-spasticity medication, are required for medical reasons, their influence will of course be considered in data interpretation. While it is currently premature in this initial study, the impact of different drug regimens on the therapeutic efficacy of MLR stimulation will certainly need to be considered in future and larger follow-up studies. We addressed this comment in subsections “Study design” (line 145-147) and “Modified Ashworth Scale (MAS)” (line 297-303) in the Methods section, as well as in the Discussion section (line 534-539).

The primary endpoint 6MWT is a simple integrative criterion, but the huge amount of additional assessment procedures listed potentially contain the relevant information for scientific interpretation of the outcome. These assessments should follow a clear protocol for minimizing the burden for the study participants and avoid bias on results through long assessment sessions. Expected outcome should be considered in the discussion section.

We agree with Prof. Mayr that the 6MWT is a rather gross test on the basis of which many details cannot be examined. Therefore, we perform a series of additional, secondary outcome assessments. We know from preclinical studies that MLR-DBS enables more intensive training and leads to improvement of, for example, walking quality and temporal parameters of the step cycle. However, since gait analyses in humans are significantly more complex, we have chosen a simple, international standard test that allows comparability with other patients and studies as primary readout. As depicted in Table 3, our study follows a precise schedule of individual assessments that are performed within certain short time intervals, preventing exhaustion and ensuring sufficient break times between tests. An exact determination of the timepoints is not possible as spinal cord injured patients fatigue at different rates based on their different lesion patterns, requiring some room for

individual adjustments. Nevertheless, the specified intervals are adhered to and at the end of the study period each patient will have completed the same type and number of assessments. Expected outcome was addressed in the Discussion section (line 543-562).

The one subject, that has already passed the study should be presented in more detail. Obviously, the expected results have not been accomplished in the primary endpoint, but there must be more results in the many secondary assessments and those should be considered relevant for further planning. Just to apply for an earlier start in the subacute recovery phase will rather lead to a less clear separation of regeneration under standard treatment and intervention driven gain. This should be avoided in a small pilot study with just 5 participants with already inevitable hard to compare lesion profiles.

Since we intend to publish the protocol of the presented study we deliberately do not include any further details about the already included patient. Of course, we are continuously performing analyses to facilitate and improve planning for the next patients and to learn from patient to patient, but it would be premature to publish any data at the present time after having tested only one patient. The key conclusions we were able to draw based on the first patient are reviewed in the Discussion section (from line 512 on).
We agree with Prof. Mayr that standard treatment vs. intervention driven functional recovery need to be observed independently at this stage in order to draw conclusions. However, due to emerging evidence from preclinical studies (e.g. studies conducted in the laboratory of Prof. Martin E. Schwab, manuscript in preparation) with promising results for both treatment initiation in the subchronic and chronic phase after injury, we extend the time window of inclusion from 6 to 3 months after injury. As we are not comparing two study groups with MLR-DBS enabled training vs. training alone but analyse intra-individual functional changes with MLR-DBS enabled training at 6 months after treatment initiation vs. baseline, a stable neurological condition at the time of implantation is required for stimulation-induced effects to become detectable. This was clarified in the Discussion (line 581-585). We are not comparing different patients with each other in this study as the disease of chronic spinal cord injury is extremely heterogeneous requiring large patient cohorts enabling subgroup analyses. We at first have to investigate whether a positive effect on recurrence of locomotor capacity can be achieved with MLR-DBS in spinal cord injured patients (studies involving the application of new neuromodulation techniques in spinal cord injured patients, e.g. epidural or transcutaneous stimulation, are typically small case series).The ideal time to initiate treatment can currently only be anticipated based on preclinical studies, but also has to be identified in human patients before initiating larger cohort studies.

Reviewer: 2
Dr. Zhou Li, Chinese Academy of Sciences
Comments to the Author:
In the manuscript entitled “Deep brain stimulation for locomotion in incomplete human spinal cord injury (DBS-SCI) – protocol of a prospective one-armed multi-centre study”, the author introduced a kind of protocol which could be used for deep brain stimulation. Besides, this work has sorted out the monitoring indexes of MLR-DBS in detail, which has certain guiding significance for the future work. However, the reviewer suggests that the manuscript requires few revisions before publication.

1. After presenting a large number of methods that can be used for evaluation, the author should further discuss the advantages and disadvantages of each method and why 6MWT is chosen in the discussion part.

We thank Dr. Li for this suggestion for improvement and addressed this comment in the Discussion section as proposed (line 543-562).

2. Formatting issues such as font spacing/case should be corrected.

We thank Dr. Li for this advice, we have again screened the entire document for formatting errors such as font spacing and case.

3. In recent years, there are many excellent works on deep brain stimulation and new devices for that, such as: Full activation pattern mapping by simultaneous deep brain stimulation and fMRI with graphene fiber electrodes, Nature communications, 2020, 11(1): 1-12. ; Electrical Stimulation for Nervous System Injury: Research Progress and Prospects, Acta Physico-Chimica Sinica, 2020, 36 (X), 2005038; Recent Development of Implantable and Flexible Nerve Electrodes, Smart Materials in Medicine, 2020, 1, 131-147; which are suggested to be cited.

We thank Dr. Li for this literature recommendations, the suggested references have been included in the Introduction section (line 122-124).

Reviewer: 3
Dr. Luka Milosevic, University Health Network
Comments to the Author:
The authors present a first-in-main Phase I/II clinical trial protocol for assessing cuneiform deep brain stimulation (DBS) in patients with incomplete spinal cord injury (SCI) in a prospective cohort of five patients. This is a novel DBS indication, and this study provides a means of establishing safety and feasibility. Hopefully the investigators can also be successful in achieving clinically significant functional improvement.

Major points to consider:
1) The small sample size is daunting if the authors hope to achieve reproducible clinical benefit, particularly if effect sizes are not large; which may be expected based on PPN-DBS literature in PD.

We thank Dr. Milosevic for raising the issue of small sample size. We are well aware that the study size does not allow generalizability of the data generated and conclusions drawn. However, since this is a proof-of-concept study where DBS of the cuneiform nucleus is performed for the first time in human patients, we have deliberately kept the number of study participants low. Even though the effect sizes may be small, the large number of different tests allows us to collect important data and insights into the effect of stimulation on various aspects that accompany an injury, on which larger future studies can be based. We addressed this comment in the Discussion section (line 540-543).

2) The introduction would benefit from a bit more detail regarding clinical and preclinical studies (more importantly, their findings) which have investigated CNF-DBS. What have the stroke/SCI studies shown, in particular? Additionally, is there rodent literature to suggest that CNF may be a better target than PPN, or perhaps the peripeduncular nucleus (e.g. PMID: 17525137)?

We thank Dr. Milosevic for this suggestion for improvement. This comment was addressed in the Introduction section (line 113-126).

3) It is unclear as to how the battery of intraoperative physiological tests during intraoperative recordings will inform decision making regarding electrode position. Moreover, I could foresee issues with patient/muscle fatigue and compliance (15mm of recording with testing at each 0.5mm, including real/imagined movements, cognitive testing, side effect screening at various frequencies and intensities, online analyses of kinematic and physiological measures, etc). This battery seems rather infeasible (or at least very costly in terms of time). This aspect of the surgery can/should be refined, procedurally, as it is difficult to foresee an efficient "decision-making" pipeline in the operating theater. Which of these features (or perhaps which combination) might inform optimal location (or location within the CNF at all; particularly considering potential issues with image-based targeting)? Are there expected intraoperative neurophysiological markers of CNF in animal literature (e.g. characteristic oscillatory frequency, neuronal firing rate or pattern, expected evoked phenomena, etc)? Side note: how many microelectrodes trajectories will be used?

Dr. Milosevic is absolutely right that the intraoperative recording phase is extensive and very time-consuming. Unfortunately, data on possible characteristic oscillatory patterns to guide the procedure, such as we know from pallidal, thalamic or subthalamic DBS, is scarce (e.g. Noga et al.[5], reference cited in line 522 in Discussion section). For this reason, we intend to collect as much data from the target region as possible to facilitate faster and easier surgeries in a possible larger future study. Using a ben's gun we plan to use 5 microelectrodes at once to cover as much area as possible in one pass. The only limitation to this is again the surgeon's regard of the patient's individual anatomy so the surgeon might decide to use fewer microelectrodes to avoid complications. During the recording process we will search for possible neuronal responsiveness to alternating movement or imagination of walking as indicator for promising locations for stimulation testing. In case no such response can be recorded, the selection of stimulation sites will be based solely on the preoperative imaging and target planning. The procedure is described in more detail in the subsection "Surgery" in the Methods section (line 201-225 and 242-247).

Minor points to consider:

The suggestions for minor revisions were implemented exactly or similar as proposed by Dr. Milosevic.

Abstract
- Major functional recovery plateaus (add:)_occur_ three to four months
- please define AIS C – or is this meant to say SCI?

Introduction
- The number of spared, descending fibres – remove comma
- might be worthwhile to acknowledge that that clinical trials for PPN-DBS in Parkinson's disease have not exactly yielded the most promising results, to date. This could further serve to justify target selection in this work.

Methods
Study population
- how/why was n=5 decided upon? [the answer to this question arises towards the end of the Methods section – perhaps these sections can be consolidated]

We thank Dr. Milosevic for addressing this. Sample size selection is described in the subsection "Sample size" in the Methods section (line 470-476) and discussed in the Discussion section (line 540-543).

Inclusion and exclusion criteria
- is there a rationale as to why there is no defined maximum timepoint since incurred SCI? could this result in variable outcome results? i.e. differences between patients who sustained SCI 6 months preoperatively vs. those who have sustained SCI at, for example, 20 years prior? With a small cohort, it might be favourable to constrain this inclusion criterion.

We thank Dr. Milosevic for raising this point. The reason for having a minimum timepoint for inclusion is that patients need to show a certain minimum level of functionality and neurological stability before being included in the study. The reason why we did not define a maximum timepoint is that we currently do not know yet based on which parameters to define proper timing as the right timepoint for initiation of MLR-DBS after SCI in human patients is entirely unknown. However, since we do not compare patients with each other but analyse intra-individual changes over the course of the study, differences in regeneration times prior to study participation are less of a concern despite a small sample size.

Surgery
- which hemisphere will be targeted and how is that decision made?

This has been addressed in the section “Study design” in line 150-153.

- could consider implanting Medtronic Percept PC for chronic LFP recording capability and future closed-loop application.

We agree with Dr. Milosevic, that the new Percept PC impulse generator offers some interesting opportunities especially in long-term recording. Unfortunately, the device was not available when the study was designed. Changing the investigational device after the study was already started seems unwise to us, but we totally agree, that the Percept PC should be considered for future studies, especially because it offers longer pulse widths.

- details about postoperative electrophysiological assessments are provided below; however the intraoperative assessments are likely methodologically different. How are MEPs and SSEPs evoked intraoperatively? What is ERP in response too? These details should be clarified (especially to be able to be differentiated from postoperative measures). Also, what is the purpose of obtaining these measures intraoperatively?

We thank Dr. Milosevic for this comment. ERP refers to lower extremity motor responses (clarified in line 241-242 of the subsection “Surgery” in the Methods section). Intraoperative MEPs and SSEPs are measured for monitoring purposes due to the close relationships with surrounding structures of the brainstem (addressed in line 241 of the “Surgery” subsection and in line 372-374 of “Electrophysiological assessments” subsection of the Methods section). As patients are awake during electrode implantation, MEPs and SSEPs measurements can be carried out in the same way as during pre- and postoperative measurements (no suppressive effect of anesthesia).

Electrophysiological assessments
- for MEPs, “All measures are recorded before and after interventions” - what does intervention refer to? Will these measurements be done in one session, e.g. postoperatively in the stimulation naïve condition, then again after a period of stimulation? Or, before and after surgery (in which case the

measurements may not exactly be comparable due to differences in EMG/TMS sites). Please elaborate on the protocol and what is meant by “intervention”.

We thank Dr. Milosevic for this advice. We clarified this in line 404-406 of the subsection “Motor evoked potentials (MEPs)” in the Methods section (and also for SSEPs in the respective subsection in line 386-388).

DBS during behavioural testing and rehabilitative training
- “In the first two weeks after lead implantation, different stimulation parameters (frequency, Hz; pulse width, µs; amplitudes, mV) are tested” – what about potential insertional/microlesion effects?

In contrast to the actual clinical applications for DBS in Parkinson’s disease and tremor, we expect to be activating instead of inhibiting the target region’s neuronal function. Therefore, we don’t expect microlesioning effects to affect improvement of the patient’s condition. We agree with Dr. Milosevic that microlesioning might, on the other hand, reduce responsiveness to stimulation in the early postoperative phase, however, such effects were not observed in animal studies.

Study Endpoints
- authors could consider splitting the “Clinical assessments” subsection into “Primary endpoint” and “Secondary endpoints” subsections. Also, Table 2 seems a bit superfluous, as it seems it is just a summary of the list of clinical assessments; each of which reappear in Table 3.

Our primary endpoint is an increased distance covered during the 6 Minute Walking Test comparing performance at the 6 months timepoint with and without DBS with performance at baseline. However, the 6 Minute Walking Test is also measured at other timepoints, where it is considered as secondary endpoint parameter. We have therefore chosen the current format instead of a division into “Primary endpoint” and “Secondary endpoint” subsections in order not to have to describe the 6 Minute Walking Test twice. We agree with Dr. Milosevic that Table 2 und 3 show a certain overlap, however, we think that the current description is clearer in consideration of the various tests and we therefore prefer to maintain the current structure.

Sample size
- authors should consider that most PPN-DBS studies in PD were n≤7, and most failed to find statistically (/clinically) significant results.

This comment was considered in the Discussion section (line 540-543).

VERSION 2 – REVIEW

REVIEWER	Luka Milosevic University Health Network
REVIEW RETURNED	25-Aug-2021

GENERAL COMMENTS	Thank you for addressing my comments. I look forward to reading about the results of this trial.