PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Repetitive transcranial magnetic stimulation (rTMS) in autism spectrum disorder: protocol for a multicentre randomised controlled clinical trial
AUTHORS	Enticott, Peter; Barlow, Karen; Guastella, Adam; Licari, Melissa; Rogasch, Nigel; Middeldorp, Christel; Clark, Scott; Vallence, Ann-Maree; Boulton, Kelsie; Hickie, Ian; Whitehouse, Andrew; Galletly, Cherrie; Alvares, Gail; Fujiyama, Hakuei; Heussler, Helen; Craig, Jeffrey; Kirkovski, Melissa; Mills, Natalie; Rinehart, Nicole; Donaldson, Peter; Ford, Talitha; Caeyenberghs, Karen; Albein-Urios, Natalia; Bekkali, Soukayna; Fitzgerald, Paul

VERSION 1 – REVIEW

REVIEWER	Xiao-Li Li Beijing Normal University
REVIEW RETURNED	16-Jan-2021

GENERAL COMMENTS	On the whole, the rTMS Protocol and Outcome Measures used in this paper are appropriate, and the ethical issues and safety issues have been carefully considered in this study.But there is a mistake on line 56 of page 8: the time point ‘1-month after completion of rTMS' should have been set at T2 instead of T1,please check and revise it.

REVIEWER	Stephanie Ameis University of Toronto
REVIEW RETURNED	03-Feb-2021

GENERAL COMMENTS	The submitted report was very clearly written and provides a clear protocol to ensure transparency of the current RCT study goals/hypotheses and planned primary analyses for the research community. I am delighted to see a multi-site group undertaking a rigorous trial approach to contribute to rTMS intervention development in autism, particularly in an age group where there are so few evidence-based options, and led by an established research team with lots of experience with rTMS research. The comments raised are generally focused on improving clarity of choices made in the current protocol for the reader's understanding and for ensuring opportunity for replication. 1. Sample choices: The participant age range is quite wide. Why was such a large range chosen? How might effects of rTMS (and stimulation parameters chosen) differentially affect participants across the age range of the sample? Will data be collected to clarify age of first ASD diagnosis as this may have an impact on heterogeneity within the sample and on performance of social

	cognitive tasks and SRS score (ie those with older age of diagnosis may have lower SRS scores and less impaired social cognitive performance contributing to later diagnosis compared to those diagnosed at an early age)? I wonder about the inclusion of individuals with VIQ on the WASI of >=55. Is this the VCI composite score for the WASI? If so, this means that verbal ability among the sample could range very widely in the current sample? Why was this cut off chosen? What are the implications of this inclusion on the participant sample, ability to consent and complete language based assessments? 2. NIBS intervention: More of a justification for why the current protocol was chosen - why iTBS (excitatory), 70%RMT, 4 week duration, 600 pulses, unilateral as opposed to bilateral? 3. Outcome measures: can the authors clarify the choice of the SRS total score for their primary outcome measure? The factor structure of this measure suggests that there are at least two components related to social communication and restrictive repetitive behaviours. Among these two sub-components, subsets of questions seem to represent different components of social communication (see Frazier 2014). Therefore, the total score of SRS would not exclusively represent social communication. Why the choice of an informant measure as opposed to a lab-based social cognitive measure? what would be a meaningful change on the SRS following intervention? The social cognitive measures seem to tap emotion and face recognition - why were these social cognitive domains chosen as opposed to higher order (e.g., theory of mind) tasks? What is the rationale that rTPJ targeting is most likely to change face recognition and emotion recognition as opposed to biological motion or another social cognitive target? 4. Sample size rationale: there did not seem to be a power analysis included to justify the current sample size. Can a power analysis be conducted to indicate the power of the proposed sample (150) for detecting various effect sizes for differences bw active and sham groups? What is the expected rate of attrition? Can the authors also include some discussion of how the mid-term analysis for non-futility affect power in the current study? 5. Analysis : -is an ITT framework being used for the primary analysis? Will a method to impute for missing data be used ? -page 14 - line 13 - there seems to be some info missing from this line - did the authors mean that additional independent variables will be examined as moderators of group-by-time effects on outcome? 6. biomarker measurement: are there any a priori hypotheses that will be explored using EEG/ERP or epigenetic measures that will be collected as part of this study? will these markers be examined in an exploratory manner to better understand potential mechanisms of change with rTMS? can a brief rationale for inclusion of these pre/post markers be included? 7. It is a potential benefit for participants and an incentive for recruitment to include the opportunity for active intervention following completion of the study, if participants find out they were in the sham

	group following T4. I did wonder whether offering participants who have received sham intervention the opportunity to undergo active treatment while the trial is still ongoing presents the risk for unblinding of study staff - given TMS facilities tend to be small and study staff may be working in the same location and see participants coming in for treatment? Can the authors discuss safeguards to ensure that the integrity of the blind is upheld despite offering participants to come back for active treatment right after they finish the 6-month study? Are there any plans for assessing integrity of the blind in the current study? Other more minor points for considerations/clarification: -is the staff measuring RMT going to be arms length - ie is it the technician administering rTMS that measures RMT at baseline? -it would be ideal to include additional timepoints for outcome measurement to be able to clarify when change occurs and stabilizes for future design refinement (ie is 4 weeks reasonable or does change happen and stabilize at 2 weeks). -can you clarify choice for side effect measurement - might information be missed if it is being assessed on a weekly basis as opposed to after each rTMS session?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Xiao-Li Li, Beijing Normal University
Comments to the Author:
On the whole, the rTMS Protocol and Outcome Measures used in this paper are appropriate, and the ethical issues and safety issues have been carefully considered in this study.But there is a mistake on line 56 of page 8: the time point '1-month after completion of rTMS' should have been set at T2 instead of T1,please check and revise it.

Thank you, Dr Li, for your kind review and for pointing out this error. We have fixed this in the revised manuscript.

Reviewer: 2
Dr. Stephanie Ameis, University of Toronto
Comments to the Author:
The submitted report was very clearly written and provides a clear protocol to ensure transparency of the current RCT study goals/hypotheses and planned primary analyses for the research community. I am delighted to see a multi-site group undertaking a rigorous trial approach to contribute to rTMS intervention development in autism, particularly in an age group where there are so few evidence-based options, and led by an established research team with lots of experience with rTMS research.

Dr Ameis, thank you for your detailed review of our manuscript, and for your positive comments on our paper and trial.

The comments raised are generally focused on improving clarity of choices made in the current protocol for the reader's understanding and for ensuring opportunity for replication.

1. Sample choices: The participant age range is quite wide. Why was such a large range chosen? How might effects of rTMS (and stimulation parameters chosen) differentially affect participants across the age range of the sample? Will data be collected to clarify age of first ASD diagnosis as this may have an impact on heterogeneity within the sample and on performance of social cognitive tasks and SRS score (ie those with older age of diagnosis may have lower SRS scores and less

impaired social cognitive performance contributing to later diagnosis compared to those diagnosed at an early age)?

While we acknowledge that this is a broad age range, this age range was chosen to ensure the feasibility of the trial (i.e., to enable sufficient recruitment) and to target age groups (i.e., adolescents and young adults) where interventions for ASD are lacking. This age range was also selected as it is identical to our pilot study of iTBS to rTPJ in ASD, which provided pilot data that supported the funding application for this trial.

This has been clarified in the revised manuscript (p. 6):

> **"While broad, this age range was selected to ensure the feasibility of participant recruitment and to target age groups (i.e., adolescents and young adults) where interventions for ASD are lacking."**

Yes, we will collect information about age of diagnosis. We agree that age might differentially affect the response to rTMS; given the sample size, we intend to investigate whether there is a relationship between participant age/age at diagnosis and response to rTMS.

I wonder about the inclusion of individuals with VIQ on the WASI of >=55.
Is this the VCI composite score for the WASI? If so, this means that verbal ability among the sample could range very widely in the current sample? Why was this cut off chosen? What are the implications of this inclusion on the participant sample, ability to consent and complete language based assessments?

Yes, this refers to the VCI composite score on the WASI-2. Our previous rTMS trials have been expanded to include those with a "mild" intellectual disability, as literature indicates that a mild intellectual disability does not necessarily preclude a capacity to provide informed consent (e.g., Horner-Johnson & Bailey, 2013). Accordingly, we were determined not to exclude these individuals from the opportunity to participate in our trial, and for the study results to be applicable to a broad range of individuals on the autism spectrum. Within our ethical/IRB approval we have a stringent procedure for determining a participant's capacity to provide signed informed consent, and this procedure will be implemented at enrolment and at various points during the participant's involvement in the trial. An adult participant will not be enrolled in the study if they do not demonstrate a capacity to provide informed consent, and a subsequent failure to demonstrate this capacity will result in the individual being withdrawn from the study. As with age, we will conduct exploratory analyses determine whether responses to rTMS are associated with VIQ. With respect to language-based assessments, we have carefully selected measures that do not require an advanced reading/comprehension level, and we have successfully used these and similar measures in our previous studies of ASD.

2. NIBS intervention: More of a justification for why the current protocol was chosen - why iTBS (excitatory), 70%RMT, 4 week duration, 600 pulses, unilateral as opposed to bilateral?

This excitatory iTBS paradigm was primarily chosen based on neuroimaging data that shows consistent reductions in TPJ activation in ASD, and a meta-analysis that demonstrates reduced rTPJ network connectivity in ASD (Wang et al., 2018). We also selected iTBS (rather than conventional high-frequency rTMS) because it can be administered very quickly and at a low intensity, which are important considerations in this clinical population. (Of note, a very recent trial involving stimulation of a similar region in ASD also employed iTBS [Ni et al., 2021].) 600 pulses was selected as it is the "standard" TBS paradigm, and one for which there is extensive safety and efficacy data across various clinical and non-clinical samples.

Unilateral stimulation was selected due to the well-demonstrated right-sided bias for TPJ activation during social cognitive tasks and the right-sided effects seen in ASD (cited in revised manuscript section below). From a mechanistic perspective, faciliatory TBS protocols, such as iTBS, produce a long-term potentiation (LTP)-like effect, and we felt that this was a suitable paradigm when targeting a region that shows reduced activation. Importantly, iTBS also demonstrates changes in regional network connectivity (Alkhasli et al., 2019).

A four-week paradigm is again similar to what we have done in past trials of rTMS in ASD, which was based on established treatment regiments in other conditions (e.g., depression) and was considered to achieve a good balance between feasibility and avoiding underdosing.

This has been clarified in the revised manuscript (p. 7):

> **"Participants will receive standard intermittent theta burst stimulation (iTBS) to the rTPJ each consecutive weekday for a four-week period (20 sessions). iTBS was chosen as it is an "excitatory" paradigm that has the potential to target the reduced activation and connectivity commonly seen in rTPJ in ASD [16 20 21 23]. It can also be administered quickly and at a low intensity, which are important considerations in this clinical population. Participants will undergo either active iTBS or sham iTBS, where a "sham coil" is used to mimic the appearance, sound, and sensation of rTMS, but without delivering electromagnetic stimulation."**

3. Outcome measures: can the authors clarify the choice of the SRS total score for their primary outcome measure? The factor structure of this measure suggests that there are at least two components related to social communication and restrictive repetitive behaviours. Among these two sub-components, subsets of questions seem to represent different components of social communication (see Frazier 2014). Therefore, the total score of SRS would not exclusively represent social communication. Why the choice of an informant measure as opposed to a lab-based social cognitive measure? what would be a meaningful change on the SRS following intervention?

This is a good point, and one that we considered in determining our primary outcome measure, but we ultimately decided to use the SRS total score to obtain a "full" picture of the clinical response to rTMS, particularly when considering the potential interaction between social and behavioural aspects of ASD. Our pilot data were also based on the SRS total score. An informant measure, rather than a lab-based measure, was selected because it is considered to have superior clinical relevance and ecological validity, and here we are primarily concerned with determining the clinical efficacy (in addition to safety) of rTMS in ASD. In terms of a meaningful change on the SRS, this is a difficult question to answer, but we would generally consider a change of 1 standard deviation to be clinically significant; apart from those scoring at the extreme end of the "severe range," a reduction of one standard deviation would result in a shift to a lower risk "range" for those individuals scoring in the "at risk" category (i.e., 60T+).

The social cognitive measures seem to tap emotion and face recognition - why were these social cognitive domains chosen as opposed to higher order (e.g., theory of mind) tasks? What is the rationale that rTPJ targeting is most likely to change face recognition and emotion recognition as opposed to biological motion or another social cognitive target?

The specific region that we are targeting has been heavily implicated in both face processing and ToM. Indeed, we would argue that the RMET is a higher-order ToM task, albeit a measure

of affective (rather than cognitive) ToM. Unfortunately, there is also a general lack of sensitive, repeatable ToM tasks that we could have employed with the selected sample in the current trial. While our stimulation paradigm might also be expected to show effects on a biological motion processing task, the large number of assessments being conducted meant that we were unable to include all possible outcome targets. Those selected were felt to be most relevant to our social communication target.

This has been included in the revised manuscript (p. 9):

> **"The neuropsychological and neurophysiological measures were selected as they are associated with activation of the target cortical region (e.g., [38]); while there were additional paradigms that could have been used (e.g., biological motion processing), we were mindful of not overburdening our participants, and selected those that we felt most relevant to our social communication target."**

4. Sample size rationale: there did not seem to be a power analysis included to justify the current sample size. Can a power analysis be conducted to indicate the power of the proposed sample (150) for detecting various effect sizes for differences bw active and sham groups? What is the expected rate of attrition? Can the authors also include some discussion of how the mid-term analysis for non-futility affect power in the current study?

We have now included details of our power analysis in the revised manuscript (p. 10):

> **"With respect to statistical power, allowing for 10% attrition of our 150 participants, and based on the estimated effect size from our previously published RCT (which revealed a moderate effect of rTMS[34]), a sample size of n = 135 in a mixed-model (2 groups, 5 time- points) will yield power of 0.99 (f = .20, α = 0.01). While this sample size is larger than the minimum suggested by a priori power analysis (n = 64, based on f = .20, α = 0.01, Power = 0.95), this will enable exploratory analysis to determine demographic, clinical, neuroimaging, and genetic predictors of treatment response."**

With respect to the mid-term analysis, we have included the following sentence (p. 10):

> **"The above-mentioned a priori power analysis, where n = 64 is required to detect a moderate effect, suggests that we will be sufficiently powered to detect an effect of rTMS in this interim analysis."**

5. Analysis :
-is an ITT framework being used for the primary analysis? Will a method to impute for missing data be used ?
-page 14 - line 13 - there seems to be some info missing from this line - did the authors mean that additional independent variables will be examined as moderators of group-by-time effects on outcome?

Yes, we will use an ITT framework for the primary analysis. With respect to missing data, we will conduct a random effects linear mixed model, which has the capacity to account for missing data using maximum likelihood estimation. This will ensure the inclusion of participants who have missing data, including those that might withdraw from the study.

This has been included in the revised manuscript (p. 10):

"Random effects linear mixed models will be used to ensure the inclusion of participants who have missing data, including those that withdraw from the study. Specifically, this will involve a between-subjects factor (rTMS condition: active vs. placebo) and a within-subjects factor (time of assessment: pre vs. post vs. one-month vs. three-months vs. six-months), with participant and site entered as random effects. We will employ an intention to treat (ITT) framework for these analyses."

With respect to the second point, we have reviewed this section and do not feel that any information is missing. In general, we are conveying that we will indeed examine whether there is an interaction between these independent variables and the group by time effects.

6. biomarker measurement: are there any a priori hypotheses that will be explored using EEG/ERP or epigenetic measures that will be collected as part of this study? will these markers be examined in an exploratory manner to better understand potential mechanisms of change with rTMS? can a brief rationale for inclusion of these pre/post markers be included?

No, we have not proposed a priori hypotheses for these biomarkers; they are indeed considered exploratory (and highly exploratory in the case of epigenetics), but we do hope that they will help to understand mechanisms by which rTMS might exert an influence on social communication.

In terms of the broad rationale for these measures, they were included in recognition of their potential to provide a clearer understanding of the mechanisms underlying clinical effects of rTMS to TPJ. In the case of ERPs, they were selected given their strong relevance to the site selected.

This has been added to the revised manuscript (p. 9):

"The neuropsychological and neurophysiological measures were selected as they are associated with activation of the target cortical region (e.g., [38])"

"The various electrophysiological and genetic measures that are being collected are highly exploratory but may help us to understand mechanisms by which rTMS exerts an influence on social communication."

7. It is a potential benefit for participants and an incentive for recruitment to include the opportunity for active intervention following completion of the study, if participants find out they were in the sham group following T4. I did wonder whether offering participants who have received sham intervention the opportunity to undergo active treatment while the trial is still ongoing presents the risk for unblinding of study staff - given TMS facilities tend to be small and study staff may be working in the same location and see participants coming in for treatment? Can the authors discuss safeguards to ensure that the integrity of the blind is upheld despite offering participants to come back for active treatment right after they finish the 6-month study? Are there any plans for assessing integrity of the blind in the current study?

This is an excellent point, and one that we have considered extensively and addressed in our trial's standard operating procedures. It is a particular risk for those sites where assessments are conducted at the same premises as rTMS interventions are delivered. We have developed a range of procedures to ensure that there will be adequate separation (physical and administrative) between TMS interventions and those conducting the assessments. It should be noted, however, that the "open label" component is conducted after all assessments have

been administered, scored, and entered into REDCap. With respect to blinding integrity, we are assessing participant's perception and confidence of treatment condition.

This has been addressed in the revised manuscript (p. 11):

> **"We will assess blinding integrity by asking participants to indicate, at the end of their four-week intervention, which condition they believed they received and the confidence (on an 11-point scale) in this judgment. At the conclusion of the final assessment (T4), participants will be unblinded as to their intervention condition by a member of the research team who is not blinded. Those who were allocated to the sham rTMS intervention will be offered the opportunity to undergo the real rTMS intervention. While this will occur after all assessments have been administered, scored, and entered into REDCap, we have developed standard operating procedures to minimise the likelihood that assessors will encounter participants completing the open label component."**

Other more minor points for considerations/clarification:
-is the staff measuring RMT going to be arms length - ie is it the technician administering rTMS that measures RMT at baseline?

Yes, the TMS Clinician who administers rTMS (who is not blinded) will also conduct RMT at baseline.

This has been clarified in the revised manuscript (p. 8):

> **"All rTMS procedures, including resting motor threshold, will be administered by a TMS clinician who is not blinded to study condition."**

-it would be ideal to include additional timepoints for outcome measurement to be able to clarify when change occurs and stabilizes for future design refinement (ie is 4 weeks reasonable or does change happen and stabilize at 2 weeks).

While this would be a useful addition, we have elected to limit our timepoints to minimise participant fatigue. We also unfortunately do not have funding that would enable resources for collecting additional timepoints.

-can you clarify choice for side effect measurement - might information be missed if it is being assessed on a weekly basis as opposed to after each rTMS session?

While this is a possibility, we elected to assess formally on a weekly basis to minimise participant fatigue. Participants will, however, be regularly asked about their wellbeing during and immediately after each rTMS session. Any side effects reported in this manner will be documented in the participant's file and will also be examined at the completion of the trial.

This has been clarified in the revised manuscript (p. 8):

> **"Participants will also be regularly asked about their wellbeing during and immediately after each rTMS session. Any side effects reported in this manner will be documented in the participant's file and will be examined at the completion of the trial."**