Author's response to reviews

Title: Cognitive correlates of repetitive transcranial magnetic stimulation (rTMS) in treatment-resistant depression- a pilot study

Authors:

Karina K Kedzior (kkedzior@graduate.uwa.edu.au)
Vikram Rajput (vikraj001@hotmail.com)
Greg Price (Greg.Price@health.wa.gov.au)
Joseph Lee (Joseph.Lee@health.wa.gov.au)
Mathew T Martin-Iverson (mathew@cyllene.uwa.edu.au)

Version: 2 Date: 24 July 2012

Author's response to reviews: see over
Bremen, July 24, 2012

Editors of BMC Psychiatry

Dear Editors,

Please find enclosed a revised version of our manuscript entitled: “Cognitive correlates of repetitive transcranial magnetic stimulation (rTMS) in treatment-resistant depression- a pilot study” to be published in the journal *BMC Psychiatry*.

On behalf of the co-authors I would like to thank the reviewers for their comments which have helped us to further improve our manuscript. The detailed responses to these comments are listed below. In addition, any changes to main text are highlighted in the manuscript for your convenience. None of the authors had any competing interests and this is now mentioned after the Author Contributions section.

Thank you for your consideration of our revised manuscript.

Yours sincerely,

Dr. Karina Kedzior on behalf of the co-authors
Responses to reviewer comments (highlighted)

Reviewer: Dr. Samir Kumar Praharaj

Methods:
It is not clear why two different protocols for administering rTMS were employed in the current study. How do the authors justify combining the results of these two different protocols? Considering the small sample it would be difficult to examine the effect of these two protocols in the present study.

Reply: The reason is already briefly explained in the manuscript (p. 6): ‘This so-called interactive technique had shown a trend towards improving the clinical efficacy compared to the standard rTMS in treatment-resistant patients with depression [15- Price et al., 2010].’

As the reviewer correctly points out it would be difficult to compare the results of the two stimulation paradigms in the current study due to very small sample sizes (and thus a low power to detect statistical significance). Further data collection with this method is certainly required to investigate if this method is superior to the standard rTMS.

Statistical analysis:
There is no mention regarding data normality. The data has been presented as Mean±SEM within the text and Median and Range in the tables. It is preferable to include the data as Mean±SD for all the measures in the tables (or Median and IQR if non-normally distributed). Also, for comparison 1, the authors conducted Wilcoxon signed rank test, a non-parametric test, though within the text mentioned it as paired t-test, a parametric test. For comparisons 2-4, though data were presented as Median and Range, parametric tests (RM ANOVA) were conducted. If data were non-normally distributed, which is probable considering small sample size, non-parametric alternatives to RM ANOVA may be better.

Reply: The reviewer correctly points out an inconsistency in terms of tests mentioned in text (parametric t-tests) and results shown in tables (non-parametric Wilcoxon Signed-Rank T-tests). In general, only some of the variables assessed in the results section were not normally distributed. For the sake of consistency the non-parametric single comparisons were conducted and reported in Tables 1 and 3. The disadvantage of the non-parametric tests is their low power and thus an increased rate of Type II errors (and resulting lack of statistical significance). In order to remove the reliance on statistical significance alone the results were expressed and discussed in terms of effect sizes.

The results in Table 2 were analysed with a parametric ANOVA. In general, the parametric ANOVA is assumed to be more robust against the violation of univariate normality (Tabachnik B, Fidell L (2006) Using multivariate statistics. Allyn & Bacon, Boston:USA) than t-tests (which are computed based on differences in means and means can be skewed if normality is violated). Furthermore, the advantage of the repeated measures ANOVA in SPSS is that it shows the results for main effects, interactions and pairwise comparisons compared to one main effect only using the nonparametric equivalent to the repeated measures ANOVA (the Friedman test). Of course, one option would be to conduct single comparisons as multiple non-parametric T-tests and correct the p-values for an inflated Type I error using e.g. Bonferroni’s correction for multiple comparisons. However, interaction effects cannot be determined with the non-parametric Friedman test in SPSS.
To remove the inconsistency in the manuscript the following statements are now included: ‘Comparison 1 was tested using the non-parametric Wilcoxon Signed-Rank T-test (equivalent to the paired-samples t-test) due to the violation of normality of scores in some of the variables included in these analyses’ (p. 9).

‘In general, ANOVA is robust against any violations of univariate normality of scores. Furthermore, the Sidak’s posthoc test was chosen because it provides an optimised balance between Type I and Type II errors compared to the two other tests available under the pairwise comparisons options in SPSS - the Least Significant Difference (LSD) Test (equivalent to an uncorrected t-test) with a high Type I error chances and Bonferroni’s test with low power and high Type II error chances [Field A (2009) Discovering statistics using SPSS. SAGE Publications Ltd., London:UK]’ (p. 9)

‘The neurocognitive functioning (RBANS scores) and depression scores (BDI, HAM-D) were compared before the first rTMS vs. after the last rTMS in patients using the non-parametric Wilcoxon Signed-Rank T-test due to violation of normality of scores in some of the variables included in these analyses’. (p. 12)

**Post hoc power calculations are unnecessary and may be removed.**

Reply: We disagree that the post-hoc power values should be removed from Table 2. These values show that the Type II error chances are high in small samples and thus the results should be interpreted based on the effect sizes rather than statistical significance alone.

It will be useful to examine the percentage of patients who responded/remitted and whether the cognitive changes were different in them.

Reply: This suggestion would be indeed interesting to investigate if more patients were tested. Splitting such a small group further would not produce reliable results (at least the power would be too small to detect any differences in cognitive performance in those who responded vs. remitted).

**Discussion:**

The similarities in the effect sizes in most of the comparisons of mCST variables between patients and healthy volunteers, except one rather points towards lack of effect of rTMS on cognitive functions.

Reply: The current results indeed do not show an overall improvement in ALL measured cognitive functions and rather an improvement in selective functions - accuracy of performance on the mCST and immediate memory. We have now modified the abstract, parts of the discussion, and the conclusion to emphasise these findings. (e.g. Abstract: ‘The rTMS was associated with an improvement in selective cognitive functions that was not explained by practice effects on tasks administered repeatedly’.)

Also, the lack of correlation with changes in depression scores does not automatically suggest independent effect of rTMS on cognition.
Reply: We do not suggest independent effects. Instead, we argue that the lack of correlation between clinical scores and cognitive scores suggests that the improvement in cognitive functioning was not confounded by improvement in clinical scores (Discussion, p. 14). However, could the reviewer please explain how there can be a dependency (causal relationship) between one measure and another, without a correlation? If one measure varies and another does not vary in a similar fashion, then they must have some independence. Correlation does not imply causation, but causation or a common cause does imply correlation. If two measures both vary as a consequence of rTMS, then those measures should correlate, if measured across different levels of rTMS.

The major limitations in the study were small sample and absence of sham control. The argument by authors that use of sham stimulation is questionable based on correct guesses by patients is untenable.

Reply: The two limitations are already discussed in the manuscript (p. 15). Since the rTMS treatment is becoming more widespread the patients might indeed have certain expectations regarding the treatment and its outcome (e.g. see a YouTube video sponsored by a company NeuroStar in which rTMS procedure and effects are described non-scientifically for the general population: http://www.youtube.com/watch?v=hMIJ3DnpZDk). Therefore, administering the sham treatment is technically not as simple as offering placebo in double-blind pharmacological studies. For this reason and the limited budget of the current study the sham was omitted from our design. In some double-blind placebo or sham controlled studies, the placebo or sham can be differentiated from the treatment at a level greater than chance. For rTMS, it may depend on how the sham treatment is given. When we have asked people given some kinds of sham treatment (e.g., a tilted rTMS), they can differentiate between the directed rTMS and the sham rTMS quite reliably. Sham treatments that include the sound but not the pulse are even more easily correctly differentiated.

Expand ICD-10-AM

Reply: ICD-10 Australian Modification (ICD-10-AM) is now expanded on p. 5.

The details of pharmacotherapy need to be given.

Reply: As this was a pilot study, the detailed information on the pharmacotherapy was not collected from the patients. This is certainly one of the limitations of the study already discussed on p. 15/16.

Reviewer: Dr. Dave Hayes

Abstract:
The inclusion of the 8 healthy volunteers and the general comparison should be included in the abstract.
The use and findings regarding the depression scores (BDI, HAMD) should be noted here.

Reply: The abstract has been modified according to the comments above (p. 2).

Background:
1 Hz< should be >1 Hz for consistency

Reply: This change is now implemented (p. 2).
How were the 8 of 54 participants chosen? It is not clear if they performed the task again (i.e. within the same timeframe as the depressed subjects) or whether their prior initial data was used. Given the latter case, are the results replicable using another random sample from the group of 54?

Reply: The 8 participants completed the mCST task concurrently with the patients for 20 days. Since the data of N=8 had too low statistical power to test for practice effects we have tested the task over consecutive 8 days on additional healthy participants in subsequent 4 studies using various designs (paper-and-pen task vs. electronic task, self-administration and scoring vs. scoring by hand). Once 54 participants were tested in total (including the original N=8) we have combined the results in a meta-analysis which has shown that accuracy of performance on the mCST was robust against practice (Kedzior K, Kochhar S, Eich H, Rajput V, Martin-Iverson M T (2011) Practice effects on the modified Concept Shifting Task (mCST): A convenient assessment for treatment effects on prefrontal cognitive function. BMC Neuroscience 12: 101). In the current study we report only the data of the original 8 participants who completed the task for 20 days together with the patients at the same location (Perth, Australia). Thus, the mCST data for the first 8 days of the first N=8 have already been published as part of a larger data set. The current paper shows the healthy volunteer data for the full 20 days. This issue is now explained on p. 4:

‘The performance on the mCST over the first 8/20 days of testing has already been published as part of a larger data set for this N=8 group [10] while the current study reports their performance on the task over the full length of experiment (20 days).’

Results:
It is noted on pg. 8 that “In case any participants missed experimental days...” It should be noted how many subjects missed treatments and on which days.

Reply: This information is now included on p. 8/9:
‘In case any participants missed experimental days 1 or 20 the performance on their individual first and last day was taken into account when computing T-tests (2/10 patients missed day 1 and one other patient missed day 20 of the experiment).’

It is not clear that the use of Sidak’s correction is appropriate given that it requires the assumption of independent sampling – unlike the Bonferroni correction. Why have the author’s not chosen to use the Newman-Keuls post hoc test?

Reply: For Sidak’s correction to provide an exact estimate of the probability of a Type I error, corrected for the number of tests (p(corrected) = real probability), the samples are assumed to be independent. If they are not independent then the Sidak correction provides a lower bound probability of a family-wise Type I, so the Sidak test is more conservative with dependent samples. The Bonferonni test is the first (linear) term of the Taylor expansion of the Sidak equation, and is therefore related to the Sidak, as its most conservative limit (the probability of a type I error is always less with a Bonferonni test than with a Sidak, and the probability of a type II error is always greater with a Bonferonni test). The only real reason to use a Bonferonni over a Sidak is if you don’t have a good computer, because it is easier to calculate, not having a fractional power term. The Sidak test is therefore more conservative with dependent samples, but as long as the number of comparisons are not too large (as in brain imaging), it works fine, with a bit more power than the Bonferonni, and using a fixed alpha
level that allows for confidence intervals (unlike the Newman-Keuls or other step-down tests). We dislike the Newman-Keuls test because it does not use a fixed 0.05 probability, but uses a sliding scale, depending on the size of the differences between means.

The comparison statement beginning on pg. 10 that states “only patients performed significantly more accurately...” is somewhat incorrect, given that (it appears) that the analyses for patients and volunteers were done separately. If compared directly, given the data, it looks as though the variability in each group would make it statistically insignificant from the other. In other words, though only one group has reached significance for accuracy, it is not appropriate to claim/imply that the two groups are significantly different without performing the necessary between-groups test.

Reply: All statements mentioning that ‘only patients performed significantly …’ are now reworded according to this comment. For example:

“Specifically, patients performed significantly more accurately ($p=.038$) on the last vs. the first mCST trial (Table 1 Accuracy). The same comparison was not statistically significant ($p=.059$) in the healthy volunteers”. (p. 11)

Discussion:
The comment that “it’s unlikely that the improvement...was due to practice...not observed in healthy volunteers” seems somewhat strong given the small sample size and what appears to be greater variability in accuracy for the volunteers in the before/block 1 condition compared to patients, as well as the appearance of better performance in general during this period for patients > volunteers. At this sample size, a single subject with highly variable performance can greatly affect the overall statistical significance. These and related issues should be raised in the limitations section. Overall, because this is pilot work in a small sample, the authors should always be careful not to overstate the importance of their findings.

Reply: Some statements in Discussion (all highlighted in the manuscript) have been reworded not to overstate the results and to remind the readers of the small N of the study. However, in general, many other prominent studies on the effects of rTMS in depression have indeed been conducted on small Ns (between 10-20 patients) and yet their results are interpreted as ‘strong evidence’ of the effectiveness of rTMS in depression (for example, see Dell’Osso B, Camurri G, Castellano F, Vecchi V, Benedetti M, Bortolussi S, Altamura A (2011) Meta-Review of Metanalytic Studies with Repetitive Transcranial Magnetic Stimulation (rTMS) for the Treatment of Major Depression. Clinical Practice and Epidemiology in Mental Health 7: 167-77).

Pg. 13 one too many “right’s” for right DLPFC.

Reply: This typo is now corrected on p. 13.

Another possible control would be having the volunteers receive the rTMS treatment. Along these lines, the authors might consider recent work by Schaller et al., 2011 looking at changes in mood following rTMS of the IDLPFC. Is it possible that improved mood in both groups (healthy and patient) would lead to improved accuracy?
Reply: Stimulating the healthy volunteers could introduce another potential confounder to the study. Similarly to depression patients, the effects of rTMS in healthy controls seem to depend on various issues such as gender, the number of stimulations, and other rTMS parameters (e.g., slow vs. fast stimulation). Thus, Mosimann et al., 2000 and Baeken et al., 2006 and 2008 have failed to show any effects of rTMS on mood in healthy volunteers: Baeken C, Leyman L, De Raedt R, Vanderhasselt M, D’haenen H (2006) Lack of impact of repetitive High Frequency Transcranial Magnetic Stimulation on mood in healthy female subjects. Journal of Affective Disorders 90 (1): 63-66

We have already mentioned the possibility that improved mood could lead to improvement in accuracy—Discussion, p. 15: ‘One possibility could be that the rTMS alleviates depression and as a consequence improves cognitive functioning but the opposite could also hold true’.

Also, (although a minor point), I wonder if, because subjects perform so well across both blocks, the authors might anticipate that a ceiling effect might prevent the detection of real differences in these groups.

Reply: Even though not shown, on average the participants were not able to reach 100% accuracy throughout the 20 days of testing and thus the ceiling effect was not reached in 20 days. It can only be speculated that the performance would reach a ceiling if the task were repeated for longer than 20 days (although reaching 100% accuracy is difficult when completing the task in descending order with letters—crossing out the letters in reverse-alphabetical order).

Reviewer: Dr. Francesco Saverio Bersani
There is only one important issue: the positive effect of TMS treatment on prefrontal cortex on cognitive symptoms of depression is a known data, already assessed by several studies.
For this reason, the results of the study do not increase the knowledge of scientific community on the topic.

Reply: The novel result of the current study is that rTMS is associated with an improvement in the concept-shifting ability using a task (mCST) that appears to be robust against the effects of practice. This statement is now included in the conclusion (p. 16/17).