Cover Page

Title: Randomized trial of 1 Hz versus 10 Hz right prefrontal repetitive Transcranial Magnetic Stimulation (rTMS) for the treatment of Posttraumatic Stress Disorder (PTSD)

Title in Clinical Trials: Study Testing if Fast or Slow rTMS is Better for the Treatment of Posttraumatic Stress Disorder (PTSD)

NCT02158663

Date of last Amendment to IRB for personnel changes 18 January 2019 – no change to treatment or assessment protocol for duration of study

1. METHODS

2.1 Participants

Veteran outpatients 18 to 50 years of age suffering from PTSD with and without depressive symptoms were recruited from the James A. Haley Veterans' Administration Hospital outpatient clinics and surrounding community. Participants were phone screened for safety and appropriateness of trial. In addition, the clinical information in the Veterans' Administration electronic medical records was reviewed. Veterans eligible to enroll and interested in participating in the study were scheduled for the initial screening/baseline visit and sent a copy of the informed consent. Those found to not be eligible or not interested in the study were referred to their primary mental health clinician or offered contact information for clinical care, if not already established with a provider.

Participants were required to meet all of the inclusion criteria in order to participate in this study, which included being male or female between the ages of 18-50 years. They had to meet DSM-5 criteria for PTSD as determined by clinical interview using CAPS for DSM-5 and have a PTSD checklist for DSM-5 (PCL-5) score greater than or equal to 45. Participants were also required to be on a stable medication regimen and psychotherapy for at least one month as well as be able to maintain this regimen for the duration of the treatment portion of the trial. The stability of the treatment regimen was assessed as part of the baseline history by an experienced psychiatrist. All participants had capacity to provide written informed consent. Candidates meeting any of the exclusion criteria at baseline were excluded from study participation. During the study, veterans could not be enrolled in an acute treatment of PTSD using evidence-based psychotherapy including Prolonged Exposure Therapy (PE), Cognitive Processing Therapy (CPT), or Eye Movement Desensitization and Reprocessing (EMDR). For safety reasons, the

standard exclusions for rTMS studies were applied and included: any metal or device implants that would increase risk of rTMS; history of epilepsy or seizure disorder, mass brain lesions, cerebrovascular accident, metal in the skull, a history of major head trauma defined as greater than mild TBI; any neurologic or medical condition likely to increase risk of rTMS; taking any medication that significantly lowered the seizure threshold (e.g., stimulants, tricyclic antidepressants (TCA), theophylline, first generation antipsychotics, etc.); or inability to determine the motor threshold in the subject. Many psychiatric comorbidities were allowed with the exceptions of: suicidal risk that precludes safe participation defined as clinical impression that the subject is at significant risk for suicide; lifetime history of schizophrenia, schizoaffective, other psychotic disorder, bipolar disorder type I or II, dementia, or dissociative disorders; personality disorder so severe that participant would be unlikely to be able to complete study protocol requirements; and history of problematic Substance Use Disorder in the last 3 months except for nicotine and caffeine. Female participants could not be pregnant, breast feeding, or not using medically accepted form of contraception when engaged in sexual intercourse. Finally, veterans could not have received prior Vagus Nerve Stimulation, Electroconvulsive Therapy, or be enrolled in another PTSD treatment study.

2.2 Procedures

2.2.1 Overview: The screening/baseline visit began by acquiring written informed consent. Subsequently, evaluations to determine safety and appropriateness, as well as clinical ratings and laboratory testing (Urine Drug Screen (UDS) and urine pregnancy testing) were performed. Those meeting eligibility criteria, including ability to obtain motor threshold, were randomized to 1 Hz versus 10 Hz stratified by significant depression (Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) > 19). Participants were treated for 5

days a week for 6 weeks with a 3 week taper (3 treatments per week for 1 week, 2 treatments per week for 1 week, then 1 treatment per week for 1 week). Participants underwent clinical evaluation weekly for clinical effect during the treatment, at the end of the taper and at 1 and 3 months post treatment. In addition, safety was assessed prior to each treatment and/or evaluation.

2.2.2 Screening/Baseline Visit: Participants first were informed about the study and given an opportunity to have all questions answered. When the veteran understood the protocol and was comfortable with participating, written consent was obtained. After consent, the Veteran underwent the following evaluations for safety and to confirm the appropriateness of entry into the trial: basic demographics; Transcranial Magnetic Stimulation and Safety Screen (TASS) (Keel et al., 2001); a history and physical by a TMS credentialed physician; and a UDS for all participants and a urine pregnancy test for all women of child-bearing potential. Clinical evaluations consisted of measures of PTSD, function, depression, traumatic brain injury (TBI), pain, and handedness with the Annette Handedness Scale (Annett, 1970). Clinician Administered PTSD Scale (CAPS-5) (Blake et al., 1995; Weathers et al., 2013) was used for baseline/screening as well as for post treatment evaluation. The CAPS-5 was developed by the National Center for PTSD and has become the "gold standard" for assessing PTSD in individuals over age 15. The evaluation provided a measure of symptom severity and sufficient criteria to determine whether a current diagnosis of PTSD was valid. The PCL-5 was used at the Screening/Baseline visit and during each Clinical Evaluation to assess PTSD symptoms (Blevins et al., 2015; Wortmann et al., 2016). This measure provided a minimum score for PTSD symptom severity as well as a secondary measure of PTSD symptoms. The functional status of veterans was measured using the Inventory of Psychosocial Functioning (IPF) (Rodriguez et al., 2012). The IPF was an 80

question self-report scale that assessed function in the areas of family, work, friendships and socializing, parenting, education, self-care, and romantic relationships with spouse or partner.

The Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) and Quick Inventory of Depressive Symptomatology-Self Report version (QIDS-SR) (Rush et al., 2005; Rush et al., 2003) assessed the severity of depressive symptoms. The MADRS score was also used to stratify the randomization (MADRS > 19 for depressed group) as well as a secondary measure of depressive symptoms. The QIDS-SR assessed change in depressive symptoms during the trial. The QIDS-SR was administered during initial screening and during the Clinical Evaluation visits to track any change in depressive symptomatology. In addition, a physical pain score was obtained at baseline/screening and at every Clinical Evaluation. The participant was asked to rate overall physical pain in the last 24 hours on a scale of 0 to 10, with 0 corresponding to no pain and 10 corresponding to the most severe pain ever experienced. The Neurobehavioral Symptom Inventory (NSI) (King et al., 2012) assessed possible post-concussive symptoms associated with mild TBI. The scale is a 22 item self-report measure.

2.2.3 Evaluations: A safety assessment was performed prior to each treatment or at any time clinically indicated (e.g., significant side effect during or between treatments) consisting of asking about: 1) side effects, 2) changes in medication, 3) changes in medical conditions, and 4) changes in psychiatric symptoms (e.g., suicidality, plans to harm others). The Clinical Evaluations were performed after every 5 treatments for first 30 treatments, at the end of treatment taper, and at 1 and 3 month post-treatment follow-ups. The Clinical Evaluations consisted of PCL, QIDS-SR, and pain score on every evaluation visit. The IPF was obtained at baseline, after 15 treatments, after 30 treatments, at the end of treatment, and at the post

treatment 1 and 3 month follow-ups. CAPS, MADRS, and NSI were obtained at baseline/screening and after the 30th treatment.

2.2.4 Repetitive Transcranial Magnetic Stimulation: The veterans were placed in the NeuroStar chair (Neuronetics, Malvern, PA) with hearing protection in place, and the motor threshold (MT) was determined. The MT was defined as the stimulus intensity of stimulation that induces visually perceptible movement of the contralateral (in this case, the left) abductor pollicus brevis (APB) 50% of the time. (Pridmore et al., 1998). Once the spot of maximal contraction was determined, a PEST algorithm (Mishoury et al., 2004) was used to determine the MT four times. The mean of these four MT served as the MT used in the study. After motor threshold determination, a stratified randomization was performed based on significant depressive symptoms, which were defined as a MADRS score > 19. Using a computer randomization schedule, an investigator who was independent of any treatments generated two random lists of active and sham index cards (i.e., with and without significant depression) that were placed in envelopes prior to the trial beginning. When the participant was ready to begin the first treatment, the treater opened the next envelope in line for the appropriate group to determine the randomization assignment. The participant knew the assignment as well as the treater but the investigators assessing all the clinician rating scales were masked to assignment.

After the randomization, the stimulator coil was positioned over the dorsolateral prefrontal cortex - DLPFC (Brodmann Area 9/46). The right dorsolateral prefrontal cortex was targeted using head measurements and a computerized program that provides coordinates for the approximate F4 electrode site under the 10/20 electrode convention (Beam et al., 2009). The dose of rTMS over the DLPFC was 110% of MT. The intensity of 110% MT was chosen as our sample was limited to 18-50 years of age. The 110% MT was thought to be a reasonable option

to balance adequate dose to overcome distance from coil to cortex (Kozel et al., 2000; Nahas et al., 2004) and tolerability. For those randomized to 1 Hz frequency, the 1 Hz rTMS was continuous for 40 minutes for a total of 2400 pulses/session. For those randomized to 10 Hz, rTMS was 4 seconds on and 36 seconds off for 40 minutes for a total of 2400 pulses/session. For both groups, the total number of treatments per participant was 36, and the total number of pulses per participant was 86,400. The interstimulus interval of the 10 Hz group was extended past the standard 26 seconds to 36 seconds in order that the time of treatments was equal for both groups. Both pulse parameters were within the safety profile of rTMS (Rossi et al., 2009; Wassermann, 1998). If subjects were initially intolerant of 110% MT, then the coil could be rotated to find a more tolerable position. If that did not make the rTMS tolerable, then the rTMS output was reduced to 100% of MT to enable the veteran to become used to the stimulation. All participants were quickly raised to 110% MT.

For return TMS visits, participants were interviewed for side effects, medication changes, medical and psychiatric changes. A TMS credentialed physician reviewed any changes prior to treatment. After hearing protection was in place, the TMS coil was positioned based on previous measurement and the patient underwent treatment. Participants were allowed to continue current medications as long as they did not increase the risk of rTMS and were held constant during the six weeks of the trial. Participants were compensated for their time in a prorated manner up to $350 if all visits were completed.

2.3 Data Analysis

Data were double entered into an Access database and cross checked for discrepancies prior to analysis. Discrepancies were resolved by returning to primary sources. Descriptive

statistics were reported for the demographic and clinical characteristics of the obtained sample, including age, sex, education, employment, baseline function (IPF), baseline severity of PTSD (CAPS and PCL-5), baseline depression (QIDS-SR, MADRS), baseline pain (Pain Score), and baseline post concussive symptoms (NSI). Categorical variables were expressed as frequencies with percentages, and continuous variables as means with standard deviations. To assess how well the samples were randomized, the two active treatment groups were compared on baseline variables including demographics and the outcomes. No missing data imputation was performed. Chi-square tests and Fisher's Exact Test (whenever cell count is less than 5) were used to compare categorical variables*; t-tests were* used to compare continuous variables within and between groups.

For the primary hypothesis, a Welch two-sample t-test that assumes unequal variances was used to investigate whether 1 Hz rTMS versus 10 Hz rTMS provides a significantly greater improvement in PTSD symptoms (CAPS score) and/or function (IPF score) by 30th treatment. Improvement in an outcome was primarily defined as change in scores from pre-treatment baseline to post 30 treatments. The presence of a significant improvement from baseline to post 30 within each treatment group was also tested by performing a paired t-test.

Secondarily, improvement was also evaluated as the proportion of responders and remitters by post 30 treatments. Response was defined as score reduction of 50% (QIDS-SR and MADRS) or 30% (PCL-5 and CAPS-5). Remission was defined using the following score thresholds: 5 or less for QIDS-SR; 10 or less for MADRS; 33 or less for PCL-5; and no longer meeting criteria of PTSD on the CAPS-5. Other secondary analyses were similarly performed on QIDS-SR, pain, NSI, MADRS and PCL-5 scores.

Additionally, sustainability of these treatment effects was investigated using loess (local least square regression), a non-parametric approach suitable for smoothing our numerical vector, of repeated observations. Loess fits least squares regressions to localized subsets of outcome data to produce graphical description of nonlinear empirical relationships (Jacoby, 2000). The graphical displays included smooth curves plus 95% confidence bounds which allowed us to evaluate a statistically significant difference between the mean trajectory over time versus its starting baseline score. We preferred the loess method to the simple pointwise plot of estimates of the mean for two reasons: addition of pointwise 95% confidence intervals will result in less efficient use of data compared to the loess estimates that are based on regressions in local neighborhood of each time point; and an alternative use of standard error of the mean around each estimate is not appropriate for inference as it underestimates the variability between individuals within the study sample (Diong et al., 2018; Nagele, 2003).

The proportions of side effects and dropouts per group were captured. Given the low proportions, a nonparametric binomial test was used to test for significant group difference in their proportions. Two-sided p-values were calculated and used to assess statistical significance defined as $p<0.05$. All statistical analysis was carried out using the R statistical application.

3. RESULTS

3.1 Patient population

Participant recruitment, enrollment, assignment to study conditions, and study retention are presented as a Consort diagram in Figure 1. Forty-four participants were enrolled with data

being acquired from September 2014 until February 2018. Nine veterans were not randomized with six failing to meet diagnostic criteria by CAPS-5, one was determined to have a history of seizures, one was unable to get off work, and one was unable to have motor threshold determined. Of the 35 randomized, 17 were allocated to the 1 Hz group and 18 to the 10 Hz group. Of the 17 veterans in the 1 Hz group, 14 completed the primary endpoint and all subsequent visits. The three participants who did not reach the primary endpoint included two that stopped due to work concerns and one who was lost to follow-up. Of the 18 veterans in the 10 Hz group, 13 completed the primary endpoints with one of those not completing the taper or follow-up visits. The five who did not reach the primary endpoint included two that could not tolerate the first treatment, one stopped due to headaches from another medical condition, one was too frustrated by parking, and one was lost to follow-up. A total of 27 who completed the primary end points constituted the final sample for testing the primary hypothesis, and they had no missing data on the tested outcomes.

- Figure 1 about here –

There were no significant differences between treatment groups on their baseline assessment scores, including demographic and baseline rating scales. (Table 1)

- Table 1 about here –

3.2 Effects at the end of 30 treatments

The results of the two-sample t-tests failed to reject the null hypothesis of no significant difference between the two treatment groups (Table 2) in improvement on the CAPS ($p = 0.77$) and IPF ($p = 0.69$) scores from baseline to post-30.

- Table 2 about here –

The secondary analyses, however, yielded some significant findings with respect to the primary outcomes (Table 2). For the 1 Hz group, a paired t-test indicated that PTSD symptoms as measured by CAPS scores were significantly lower by the 30th treatment (improved, p = .03). The scores on MADRS, PCL-5, and QIDS-SR also significantly improved. The change in function as measured by the IPF, however, did not quite reach significance (p=0.075) in the 1 Hz group. In contrast, the 10 Hz group demonstrated significant improvement from baseline to the 30th treatment on all outcomes listed except pain score. (Table 2) Similarly, there was failure to reject the null hypothesis of no relationship between the groups on either response or remission rate by 30th treatment (Table 3).

- Table 3 about here –

3.3 Effects at 3 months post treatment

Figures 2 and 3 present trajectories of study outcomes using loess smooth curves (with 95% confidence bounds), which track changes in repeated assessment scores from baseline to 3 months post treatment. The top graph of Figure 2 fitted one trajectory to the whole sample (combined groups). It demonstrated that the average PCL-5 score declined below the lower 95% confidence limit of the baseline estimate (lower score is better) and stayed below this limit for up to 3 months post treatment.

The bottom graph (Figure 2) displays the loess curve split by group to investigate group difference in the trends. By showing overlap of their 95% confidence bounds everywhere the curves failed to reject the null hypothesis of no difference in the effects of 1 Hz versus 10 Hz rTMS treatments on PCL-5 over time. Similar patterns with equivalent interpretations for

depression symptoms (QIDS-SR) are shown in Figure 3. The loess curves for IPF and Pain score (not shown) also did not reveal a significant change in trend or treatment group difference.

- Figure 2 & 3 about here –

3.4 Tolerability of treatment

With respect to patient safety, there were no significant events related to the treatment in this trial. There were two Serious Adverse Events (SAE) that were unrelated to the study interventions. There were no seizures and no continuing complications. The 10 Hz group had two participants (2/18 or 11%) that could not tolerate the treatment at the first visit but no such report from the 1 Hz participants (see Figure 1). Exact binomial tests (2-sided) failed to show significant differences between the two randomized groups on the proportions of Veterans who reported headache, site pain during treatment, or stopped treatment due to side effects (Table 4).

- Table 4 about here –

3.5 Post-hoc power calculations

Given that the final group sample sizes of 14 and 13 (see Figure 1) used to test our primary hypothesis fell short of our anticipated evaluable sample of 40, inadequate statistical power was suspected. Therefore, post-hoc power calculations were performed based on the statistics obtained from the results of t-tests for CAPS-5 (Table 2). Using a paired samples t-test with alpha of 0.05, there were 98% and 76% power (for 10 Hz and 1 Hz groups respectively) to reject the null hypothesis of no difference between the baseline and post-30 scores (that the mean of the paired differences is zero); which support the conclusions reached from our significant results. For the nonsignificant results of 1 Hz versus 10 Hz, however, there was low power to

reject the null hypothesis of no difference in group means using a two-sided two-sample unequal-variance t-test with alpha of 0.05. Therefore, no definitive conclusion can be derived from any failure to reject the null (i.e. non-significant findings) in this study. Although we did not find a definitive conclusion with our study, the preliminary results are that if a difference exists at the group level, it is not likely a large clinical difference.