Acupuncture for PTSD in Combat Veterans
A VA Merit Review Study
Principal Investigator: Michael Hollifield, M.D.

Data Analysis Plan

General Approach. General linear mixed models (GLMM) is capable of handling multiple underlying distribution and model structures through link functions, such as repeated measures random effects models of continuous outcomes (identity link), repeated measures logistic models (logit link), and Poisson and negative binomial models (log link). In addition to modeling global fixed effects across subjects, GLMM can also model individual subject random effects. Survival analysis allows for the assessment of time to event modeling using probability estimates on survivor functions and estimates of hazard ratios. Time to event modeling using Cox proportional hazard modeling allows for all subjects to contribute information to the model up until occurrence of event, drop out, or end of study. Cox Proportional Hazards modeling will be used to investigate for effect-size of the treatment by allowing subjects to contribute their information to the model while they are being observed and censored once they are no longer being observed (loss to follow-up or end of study period). Further, using the Cox modeling we will investigate for time dependencies and informative loss to follow-up. In the case where dropouts may be associated with the treatment assignments, we will leverage intention to treat methodology and construct a piecewise random effects model with both "on-" and "off-treatment" slopes. GLMM will be used to evaluate the primary clinical hypothesized effects of treatment (ACU) on the clinical outcome of PTSD symptom severity (CAPS) over time (mid- and end-treatment, and 1-month follow-up), controlling for baseline severity of symptoms and demographic characteristics (e.g., age, gender) in comparison with placebo control group (MIN) with assumption of intent to treat. Cohen's d within and between subjects will be calculated. Interaction terms will be included in the models to evaluate treatment fidelity and treatment expectancy as potential moderators.

GLMM will also be used to evaluate the secondary biological hypothesized effects of treatment on pre- to post-treatment in PPR (decreased EMG eyeblink). These same statistical procedures (GLMM, survival analysis, and Cohen's *d* with intent to treat) will be applied to evaluate exploratory outcomes: clinical symptoms comorbid with PSTD, PPR (HR, HRV, SCR), and PTSD diagnosis.

Statistical Analyses, Primary Hypothesis. Prior to the application of any statistical modeling, underlying assumptions and conditions will be examined. Univariate analyses, including tests of distribution assumptions, t-tests, and chi-square tests will be conducted to determine possible significant covariates to be included in further multivariable modeling as well as to investigate efficiency of randomization techniques. All models will be assessed for goodness of fit as well as other diagnostics including assessment for collinearity performed on all covariates chosen for subsequent analyses. Repeated measures analysis of variance will be used to formally test the null hypothesis of no differences in baseline PTSD symptom mean levels and subsequent retest of PTSD symptom mean levels while simultaneously adjusting for any significant covariates. All models will be run accounting for any potential confounders. Analyses will be performed assuming intent to treat and informative loss-to-follow up will be investigated. Additional analyses will be conducted to determine interactions and mediating effects while simultaneously adjusting for other covariates in the model.

Assumptions for Power Calculations. Assumptions are based on interpolated data since primary outcomes in our first ACU study were assessed with the PSS-SR scale and there are no PTSD studies using MIN as the sham. Furthermore, to be consistent with major PTSD research, we will use the CAPS-5 (DSM-5 version) as the primary outcome, and there are as of yet no completed clinical trials using CAPS-5. In our first study, mean pre-treatment PTSD severity score was about 62% of maximum (31.55 of 51). ACU mean decreased 50% - 31.33 (10.1) to 15.65 (13.95) and wait-list mean decreased 9.3% - 30.79 (9.54) to 27.92 (12.33) at post-treatment. We thus assume that subjects will have a baseline mean CAPS-5 score of 50 (10) and a 50% reduction of symptoms and a 38% increased variance (CAPS-5 = 25; SD = 13.8) with ACU. There are sparse data about effects of

MIN, noting a 33% improvement compared to pharmacological treatment only, and a 33% less effect than ACU in the Shen study. A 2010 Cochrane systematic review of over 200 trials investigating 60 clinical conditions about general placebo effect found placebos to not have important clinical effects but may influence patient-reported outcomes in some situations (e.g. pain and nausea). The pooled relative risk calculated for placebo was 0.93 (effect of only 7%) but significant. Confidence intervals are generally wide in the placebo arm. Several clinical and methodological factors were associated with higher effects of placebo. Since our study includes some of these factors, and since we expect MIN to have some physical effect, MIN could provide an effect as much as 33% with a wider variance. This would predict mean CAPS-5 scores (MIN) at end-treatment of 33.5 (SD 15). The conservative prediction is mean CAPS-5 reductions of 25 points with ACU and 16.5 points with MIN, an 8.5 point difference with a pooled SD of 14.4. Given our experience and data about placebo effect, we can modestly expect a 12 point between group CAPS-5 difference and a pooled SD of 15.

Treatment Effects, Power and Sample Size, Primary Hypothesis. "The efficacy of verum acupuncture (ACU) for PTSD symptom severity will be large (pre- to post-treatment Cohen's $d \geq 0.8$), and significantly better than sham acupuncture (MIN) (between group Cohen's $d \geq 0.30$, with 80% probability of detecting a true group difference at $p<0.05$ (2-sided)."

Effects. The conservative assumptions noted above will result in pre- to post-treatment Cohen's $d = 2.07$ within group (ACU) and between group (post-treatment difference) Cohen's $d = 0.59$, which will prove the null hypothesis false and show a large treatment effect for ACU and a moderate between-group effect size.

Sample size needed and power. The conservative assumptions would require a total of 90 patients in a two-treatment parallel-design to provide a probability of 80 percent to detect a treatment difference at a two-sided 0.05 significance level. Modest assumptions of a 12 point difference with a pooled SD of 15 requires a total of 50 patients with power of 0.80 at alpha <0.05.

This was determined using a formula for a random effects modeling (Montgomery, 1991), which may determine the necessary sample size to detect a statistically significant difference in treatment/control group means at different levels of projected effect differences.

$$n = \frac{2 * a * M^2 * \Phi^{*2}}{\Delta^2}$$

Where:
M = measure of noncentrality parameter to estimate the standard normal probability of making a type II error
n = number of replicates necessary for each treatment
a = number of treatments (a=2)
Φ^* = estimated standard deviation
Δ = difference in baseline and post treatment

Table. Examples of sample sizes necessary for significance testing at different mean and SD's (between group: treatment (ACU) and control (MIN)) for PTSD severity (primary outcome) at power = 0.80.

Mean difference between ACU and MIN PTSD severity scores post-treatment	Φ	Total N Using Formula Above
8.5	12	62
	14.4	90
	15	96
	20	162
12	12	32
	15	50
	20	86

Recruitment goals are N=90, 45 per group. With a post-randomization attrition rate of 10% we will have at least 40 individuals in each group. Considering attrition, the above table indicates that for an average difference in means of 12 at end-treatment with a standard deviation of 15, a total sample size of 50 (25 in each group) would

be more than adequate to report a statistical difference with 80% power. If differences in treatment group means were found to be larger than 12, we would need even fewer subjects in our treatment groups to achieve statistical significance.

PPR Data Recording and Reduction. A description of the procedure/data is in Appendix 4. Data reduction and analyses will follow current established protocols in our lab (Norrholm et al, 2011a). Data will be collected with the BioSemi recording software (BioSemi B.V, Amsterdam, Netherlands) and the resulting data will be exported to Mindware software (Mindware Technologies LTD, Gahanna, OH) for data reduction and generation of analyzable variables.

Startle data reduction for analyses. The raw EMG signal will be recorded at a rate of 1000 Hz throughout the experimental session using a 28 Hz high pass and 500 Hz low pass filter (as recommended by guidelines for human eyeblink startle in Blumenthal et al., 2005; Psychophysiology, 42:1-15). Raw signals will be stored and exported for analysis in microvolt (μV) values.

Skin Conductance Response. Skin conductance responses are scored as the largest amplitude responses beginning in a window of 1 to 3 seconds following stimulus onset. A response is defined as having amplitude greater than 0.01 μS relative to the pre-stimulus baseline (Boucsein et al, 2012).

Heart rate. Average heart rate (HR), the standard deviation of the HR, power of high frequency (HF), low frequency (LF), very low frequency (VLF) components, % power of LF (%LF [of VLF+LF+HF]) and of HF (%) and the ratio of the LF over the HF (LF/HF ratio is used as an indirect autonomic balance index) of HRV are calculated as cardiac activity measures. Artifact-corrected 3-min long recording epochs are analyzed with FFT to assess HRV. Inter-beat Intervals (IBIs) will be scored as the time difference between successive R waves in the ECG signal. IBIs will be used as the dependent variable analyzed instead of heart rate because of a lowered susceptibility to artifact due to differences in baseline values (Stern et al., 2001). A window of 3 seconds pre-stimulus onset to 6 following stimulus onset will be scored. Instantaneous IBIs will be recorded at half-second intervals during the pre- and post-stimulus time windows. A difference score between the average pre-stimulus IBI for each trial and each post-stimulus IBI value will be computed for each trial. IBIs will later be converted heart rate (beats per minute) for more readily interpretable post-analytic write-ups.

Respiration. Respiration will be scored in a similar fashion to the ECG data, and reported in interbreath intervals. Peak detection of each positive deflecting curve in the breathing cycle will be manually reviewed visually in order to assess accuracy. Interbreath intervals will later be converted to respiration rate (breaths per minute) for ease of interpretation.