Nicholas Montalbano

Stat 998 Report: Prabhakaran Lab (Ms. Sinha, Dr. Nair, and Dr. Prabhakaran)

**Summary**

Stroke patient data is examined on EEG-based brain-computer interface (BCI) intervention for stroke rehabilitation from the Prabhakaran Lab. The data consists of 28 stroke patients with persistent mild to severe upper-extremity motor impairment. The patients are randomly split into an early treatment and late treatment group. This analysis focuses on whether there was an improvement in behavioral outcomes and brain measures after BCI therapy interventions. A model is built for each behavioral and brain measure that includes the following variables: treatment group, visit number, the interaction between treatment group and visit number, and a patient random effect to account for the fact that the same patient is measured across a time interval. In addition, other variables that the researchers are interested in were included in the model: stroke severity, handedness, time since stroke, and gender.

The analysis finds that there is weak evidence for a BCI therapy effect on behavioral measures and no evidence for a BCI therapy effect on brain measures. For behavioral outcomes, the results are weak despite consistently higher outcome measures for patients in their visits after receiving the therapy. This can potentially be due to the relatively small sample size of patients. For the brain measures, it is possible that no change was found due to the potential for positive brain connectivity to increase in positivity and negative brain connectivity to increase in negativity, which results in the netting out of any detectable effect. Interestingly, handedness is found to have a particularly important relationship with the level of connectivity of regions of the brain to the right thalamus.

**Background and Data Demographics**

Before digging into the details of the models and analysis, it is worth describing general characteristics of the data. The treatments groups between which the data were split—early and late treatment—are relatively even across relevant covariates, including age, gender, time since stroke, and stroke severity. However, we note that the early treatment group does not have anybody who is left-handed. This will limit our ability to determine the interaction of handedness with treatment group.

The early treatment group is assessed four times: a baseline assessment, a mid-therapy assessment, a post-therapy assessment, and a one-month post-therapy assessment. The late treatment group receives these same assessments, but also goes through a control phase in which they do not receive any intervention before the therapy phase. In total, the early treatment group has 4 visits and the late treatment group has 7 visits. When appropriate, the late treatment group will be separated into two groups: visit numbers 1-4 (late control) and visit numbers 4-7 (late treatment). We note that some measures could not be taken for certain patients on a few visits.

For the analysis of the brain measures, we look at the functional connectivity between 8 cortical regions. This gives us 28 unique pairwise connections of Fisher’s r-to-z transformed scores for each seed region of interest. We note that some of the correlations should be reinvestigated for quality. For instance, one of the patients (E003) has a near perfect correlation between unique seed regions. For the behavioral measures, we look at the Stroke Impact Scale (SIS), Action Research Arm Test (ARAT), and Nine-Hole Peg Test (9-HPT). For the SIS, we focus on the transformed scores for Activities of Daily Living (ADL), Hand Function (HF), Mobility, and Physical Strength.

In our analysis of the ARAT and SIS strength, mobility, HF, and ADL, there is the problem of the “ceiling effect,” or the inability to score above a certain amount due to the upper threshold of a test. To deal with this, patients who scored the maximum on the tests on their first relevant visit are removed from the sample for that test. However, patients who exhibit the “floor effect,” or fail to score the minimum on a test, are kept in. The exception to this is the 9-HPT in which we remove patients who could not score better than the upper time threshold of 2 minutes.

**Key Graphs**

*Behavioral Measures*

The analysis focuses on the changes that occur from the pre-therapy baseline all the way through to one-month post-therapy. To get a picture of how patients in the different groups and phases changed across the behavioral measures, median score change is plotted for each outcome measure separated by early treatment, late control, and late treatment. Median was chosen as the summary statistic due to the presence of outliers in the data. The plot of the means looks very similar. Note that a decrease in score for the 9-HPT is associated with an improvement in outcome (a quicker time).



The plot suggests a couple of preliminary findings. First, the treatment groups do show a median score increase from the pre-therapy to the one-month post-therapy phase. It is important to note that the data shows wide variation within each group phase, so that all patients in the treatment groups do not show improvement. Second, the median value in the treatment groups perform better than the median value in the control group. This is true in all the tests except for the ARAT score, in which the early treatment ties the control group. Third, the median value of the control group does not exhibit improvement on any of the behavioral measures except for SIS mobility. In all the other cases, the median value remains the same as it did during the pre-therapy phase.

*Brain Measures*

To get an understanding of how the brain measures change from the pre-therapy to one-month post-therapy phase, the change in correlation value for each pairwise connection is plotted. As an example, if patient 3 in the early treatment group had a correlation value of 3.4 between the right motor cortex and the right supplementary motor area on visit one, and then had a correlation value of 4.7 between the right motor cortex and the right supplementary motor area on visit four, then that would add to the proportion under the 1.3 area of the curve for the red density.



The first and most striking feature of this plot is that all three of the densities look extremely similar. Most of the connectivity scores remained about the same, with the mean, median, and mode existing almost right above zero. The variances also look extremely similar. Without a good interpretation of the meaning of an “increase in negative correlation”, we did not make any changes to account for the fact that an increase in positive value is treated the same as an decrease in negative value, e.g. a change from 5 to 6 is treated the same as a change from -3 to -2. This gives us an overall picture of how the correlation values changed but does not give us information about the specific correlation increases and decreases for each region of interest. The models in the next section will add rigor to determining the truth of some of the findings seen in the plots.

**Analysis**

Since each patient is recorded on the same measures across either 4 or 7 visits (repeated measures data), a linear mixed effect model is used. The random effect in this model is the patient. The general format of the model for both brain and behavioral measures is:

**Measure = Intercept + Early/Late + Visit Number + Interaction Between Early/Late and Visit Number + Patient Random Effect + Time Since Stroke + Handedness + Gender + Stroke Severity + Error**

This analysis focuses on whether or not there is a difference before or after the intervention with respect to the brain and behavior. This means that the data includes visits 1-4 for the early treatment group and visits 4-7 for the late treatment group. This is done so that it can be determined whether the treatment had any effect. It is justified by our initial findings that the control group showed little change over the course of the control phase. We include the early/late indicator variable in order to determine if there is a difference between the two groups, and potentially determine the existence of any carryover effects.

The inclusion of patient as a random effect allows the model to account for the fact that each patient is measured at various timepoints. This allows the model to reduce the unknown variation in the data and increase our power to detect treatment effects. Diagnostics on the residuals generally justify using a linear model with normal error and a normal random effect in the case of both the brain and behavioral measures. For the behavioral measures, this is true in large part because the patients exhibiting ceiling effects were removed for the test that they exhibited these effects. This reduces the sample size somewhat, but not by too large an amount.

The variable of utmost interest is visit number. If a measure increases significantly with each visit, and the increase persists through the one-month post-therapy visit, then we can confidently say that the therapy had a positive effect on the outcome or brain measure of the patients. An interaction between visit number and the early/late indicator variable is included so that it can be determined whether there is a difference in the effect of the therapy in each treatment group.

Also included in the model are other covariates of interest: time since stroke, handedness, stroke severity, and gender. The researchers are interested in whether differences in these covariates are associated with higher or lower outcome and brain measures. Handedness, stroke severity, and gender are natural indicator variables. But due to the variation in times since stroke in the sample, three groups are set: (1) less than 12 months since stroke, (2) less than 30 months since stroke, and (3) 30 or more months since stroke. These cutoffs were made after observing the distribution of the variable. In addition, time since stroke was adjusted to the start date of the therapy phase, which started about two months after the start date of the control phase for the late treatment group.

Due to the nature of the data, we ended up with a total of 34 models. Each brain and behavioral measures were modeled individually, which meant that there was a model for the six behavioral measures and a model for each pairwise connection between each region of interest. The models included in this analysis are only a small subset of the total potential models that could have been used for this data. The general process for analyzing each model is as follows: (1) run the full model, (2) examine the coefficients for effects sizes and potential significance at each level of each variable, e.g. for visit number, a coefficient is estimated for visit 2, visit 3, and visit 4 (with visit 1 being included in the intercept, or “baseline”),(3) run an F-test to determine whether or not there is a difference in the means between all levels of a variable, e.g. is the mean outcome score for visit 1 the same as visit 2 and visit 3 and visit 4, (4) use Tukey contrasts (which adjusts P-values for multiple comparisons) to determine whether or not there is a difference between each visit combination; in particular, we are interested in whether there is a difference in the outcome measure pre-therapy and one-month post-therapy. We will discuss important and notable findings from these models but will not report every estimate and P-Value.

*Behavioral Measures Models*

Six outcome measures are modeled: Action Research Arm Test, Nine-hole Peg Test, and Stroke Impact Scale Physical Strength, Mobility, Activities of Daily Living, and Hand Function.

*Behavioral Measures Model: SIS Strength*

For this model there are positive coefficients for the main effect of each visit after the pre-therapy assessment (visit 2, visit 3, and visit 4). This means there is an increase in score from baseline (visit 1) after at least one session of therapy. Both moderate and severe stroke severity have large negative coefficients, -20.75 (P-value > .10) and -31.68 (unadjusted P-value of .00016) respectively. This means patients with more severe strokes perform worse on the outcome measure. The F-tests reveal that there is significant evidence that the means for each visit number are different. This indicates that the mean outcome scores are not the same for each visit number. However, the Tukey contrasts only provide weak evidence (adjusted P-value of 0.072) that visit 4 has a different outcome mean from visit 1. This is to say that there is weak evidence that the one-month post-therapy scores are different from the pre-therapy scores.

*Behavioral Measures Model: SIS Mobility*

Similar to the findings for strength, there are positive coefficients for the main effect of each visit and large negative coefficients as stroke severity increases. For the F-test there is very strong evidence for a difference in means across the four visits. However, we again only find weak evidence for a difference in outcome score between visit 4 and visit 1 (adjusted P-value 0.053).

*Behavioral Measures Model: SIS Activities of Daily Living*

The pattern continues with the ADL outcome measure. There are positive coefficients for the main effect of each visit and large negative coefficients as stroke severity increases. For the F-test there is very strong evidence for a difference in means across the four visits. Here, the Tukey contrasts show that there is no evidence of a difference between one-month post-therapy and pre-therapy (adjusted P-value 0.440).

*Behavioral Measures Model: SIS Hand Function*

Again, there are positive coefficients for the main effect of each visit and large negative coefficients as stroke severity increases. However, for the F-test there is only weak evidence for a difference in means across the four visits. In addition, there is no evidence of a difference between one-month post-therapy and pre-therapy (adjusted P-value 0.192).

*Behavioral Measures Model: Nine-hole Peg Test*

For this test, it should be noted that all of the severe stroke patients were removed from the sample due to none of them being able to complete the task under two minutes on their first visit. Therefore, we dropped stroke severity as a covariate. There are negative coefficients for the main effect of each visit (the outcome measure is time, so a lower score is associated with an improvement in outcome). For the F-test there is very strong evidence for a difference in means across the four visits. There is no evidence of a difference between one-month post-therapy and pre-therapy (adjusted P-value 0.1087), but there is moderate evidence for a difference between immediately post-therapy (visit 3) and pre-therapy (adjusted P-value 0.0447).

*Behavioral Measures Model: Action Research Arm Test*

The variable for stroke severity was removed from this model because the ARAT test is what determines the designation of stroke severity. As was true for all of the prior tests, the coefficients for the main effect of each visit indicate an improvement in outcome measure after the therapy. The F-test provides moderate evidence that the means across all of the visits are different. However, here it is notable that there is strong evidence of a difference in outcome score between one-month post-therapy and pre-therapy (adjusted P-value 0.0035). In addition, there is moderative evidence for a difference between visit 3 and visit 1 (adjusted P-value 0.0233).

*Behavioral Measures Model: Summary*

Some clear patterns emerged from these six models. The mean scores for visit 2, 3, and 4 were consistently above the mean score for visit 1. However, the scores did not necessarily increase from visit 2 to visit 3 or visit 3 to visit 4. This suggests that the therapy sessions did not improve behavioral outcome in a straight upward direction. Additionally, for all of the SIS outcome measures there was no evidence for a difference between visit 3 and visit 1. It is also noted that there was no difference between the behavioral outcome measures of the early and late treatment groups, which could suggest a lack of carryover effect.

*Brain Measures Models*

Each unique pairwise connection is treated as a dependent variable to model the changes in the connectivity matrices of each patient’s brain. The same covariates are used as in the behavioral measures except without the interaction between early/late group and visit number. The key results and takeaways are summarized.

*Brain Measures Models: Summary*

The most striking result is that the models show no evidence that there is any difference in the mean connectivity change over any of the visits. This would indicate that the therapy effects are minimal in changing the connectivity matrices in the positive direction. It is possible that the effects of the therapy are being netted out in our data so that increases in positive connectivity are being negated by increases in negative connectivity. This is because there is no identifiability of what would constitute a “good” change in connectivity.

In addition, there is another pattern to note in these models. There tends to be a difference in the connectivity outcomes based on handedness when observing connections with the right thalamus. There is strong evidence that the left primary motor cortex, right motor cortex, and right supplementary motor area exhibit a difference in connectivity outcomes with the right thalamus. There is moderate evidence of the same phenomena for the left thalamus’ connectivity with the right thalamus and weak evidence for the right premotor cortex’s and the left supplementary motor area’s connectivity with the right thalamus.

**Conclusion**

In summary, we find weak evidence for a therapy effect on behavioral measures and no evidence for a therapy effect on brain measures. A potential reason for a lack of significant evidence for behavioral measures, despite the consistently positive coefficients for visits after the pre-therapy phase, is the relatively small sample size of patients and the need to reduce this sample size even further due to ceiling effects. Further clarity on what would constitute a “good” change in connectivity would bolster the brain measures analysis.