**B-SNIP ADEPT Phase I Report**

**Biotype Classification Based on Clinical Features Only**

Brett Clementz

University of Georgia

Ishanu Chattopadhyay

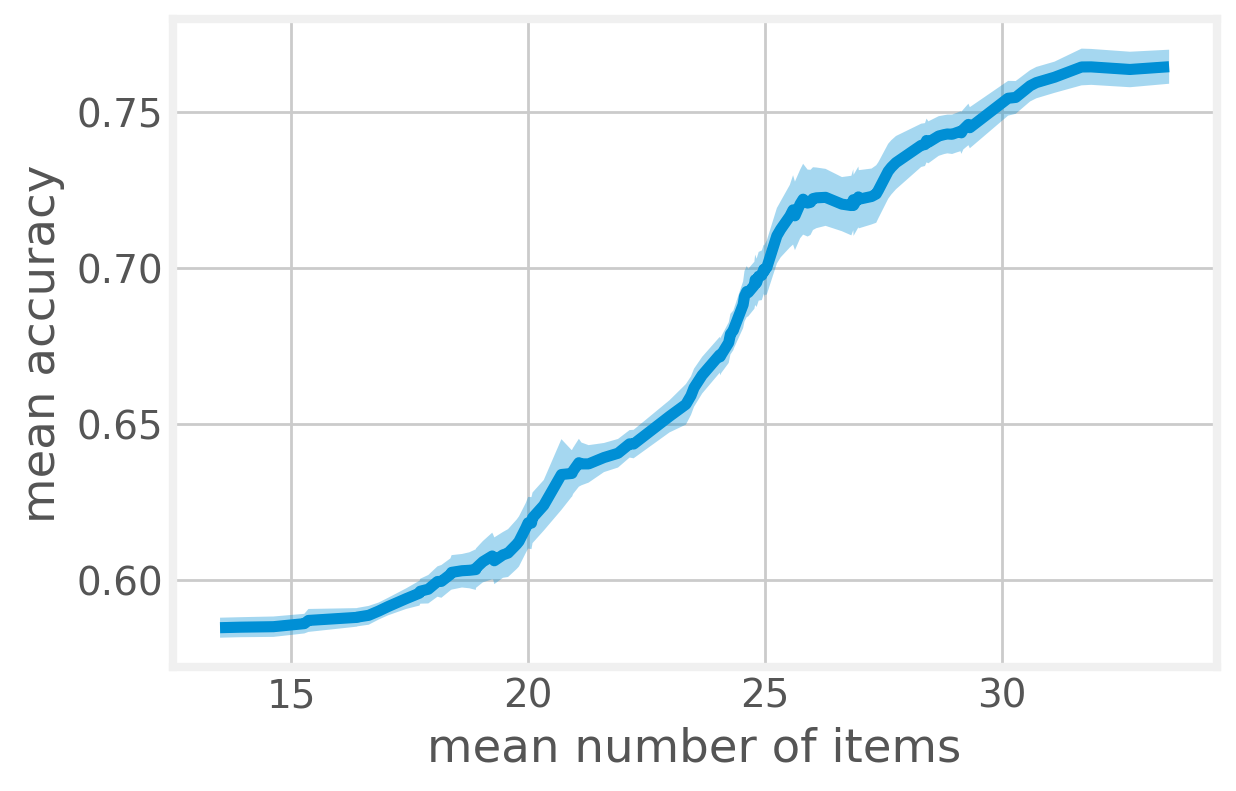
Robert Gibbons

University of Chicago

November 29, 2021

The BSNIP dataset presents a multi-class classification problem with three biotype classes, which we will refer to as B1, B2 and B3. The goal of Phase I is to use clinical symptoms only to produce a low-burden adaptive classifier that can be used by non-B-SNIP sites to inexpensively derive a biotype classification. The expectation is that we can achieve approximately 60% overall classification accuracy using clinical symptoms alone. Phase II involves the use of cognitive measures and Phase II EEG measures that are expected to ultimately provide classification accuracy of 90% or more, using a reduced set of features (measures) that were used in deriving the biotypes. In Phase I, we began with 58 clinical features characterizing each patient. We consider the following approach to inferring our classifier:

1. We compute a one-vs-all extremely randomized ensemble of decision trees for each of the three classes (i.e., biotypes). An ensemble classifier using the extremely randomized tree algorithm differs from a standard random forest. As in random forests, a random subset of candidate features is used for computing splits, but instead of looking for the most discriminative thresholds, they are drawn at random for each candidate feature and the best of these randomly generated thresholds is selected as the splitting rule. This usually reduces the variance of the model, at the expense of a slightly greater increase in bias. The AUC obtained in out-of-sample runs is shown in Figures 1-3. While the out-sample AUCs obtained for each biotype are rather high (>0.85), the end-to-end performance for the three-class problem is expectedly lower. The average number of features used (out of 58 items) for the overall classification of a patient into one of the three biotypes as a function of the mean accuracy is shown in the following figure (shown with 99% confidence intervals)



1. With these base classifiers we proceed to construct an end-to-end pipeline for the overall problem. In our preliminary studies, we used an empirical maximum likelihood approach to decide on the class of test samples. In our out-of-sample tests, we obtain a median end-to-end accuracy of 62.5%, when the total number of items used is restricted to 20 and rises to ~70% if 25 items are allowed. One example of a normalized confusion matrix is shown below.

Estimated

B1 B2 B3

B1 0.79 0.11 0.10

True B2 0.13 0.62 0.25

B3 0.30 0.11 0.59

The decision algorithm using the base classifiers needs to choose 2 out of the 3 classifiers (each individually aiming to classify one of the biotypes from the rest). Denoting each of these one-vs-all classifiers as **Ci, i=1,2,3**, we can enumerate the overall the classification algorithm as:

1. Use C2
2. If decision == B2:

terminate

Else:

* 1. Use C3
  2. If decision == B3:

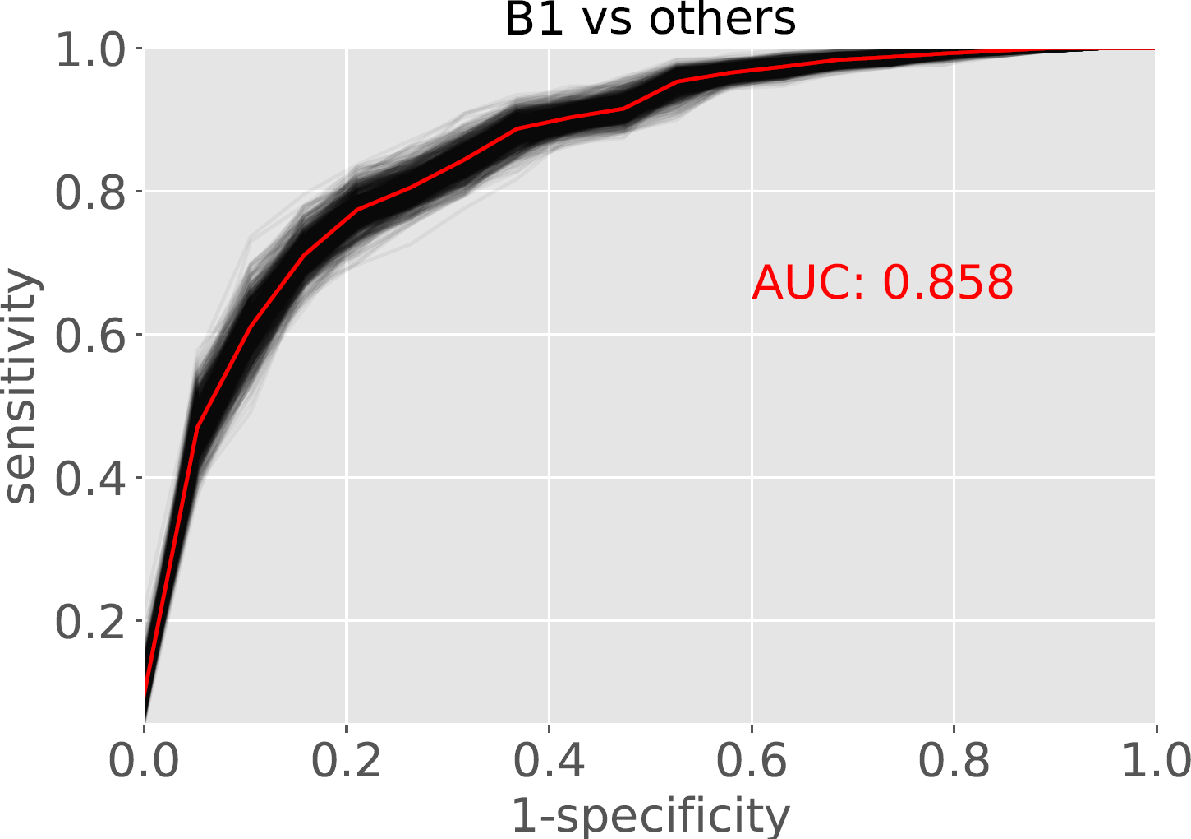
Terminate

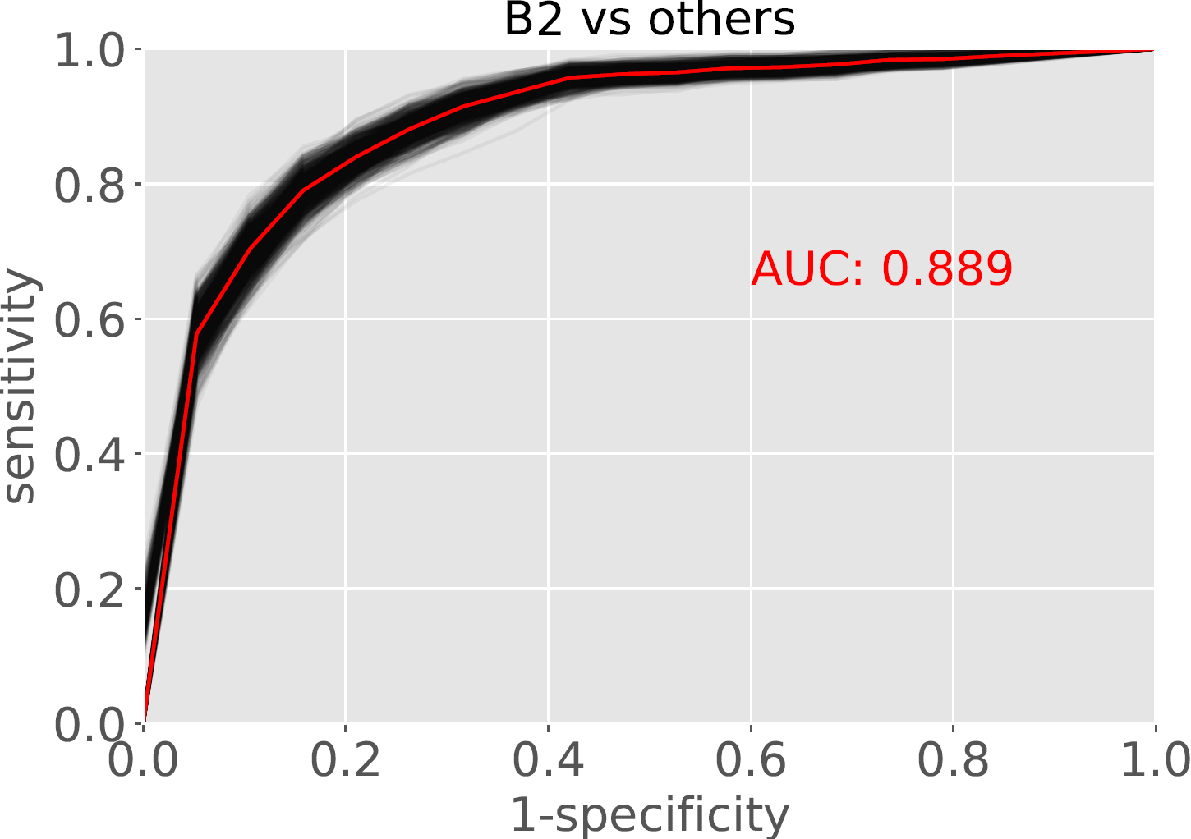
Else:

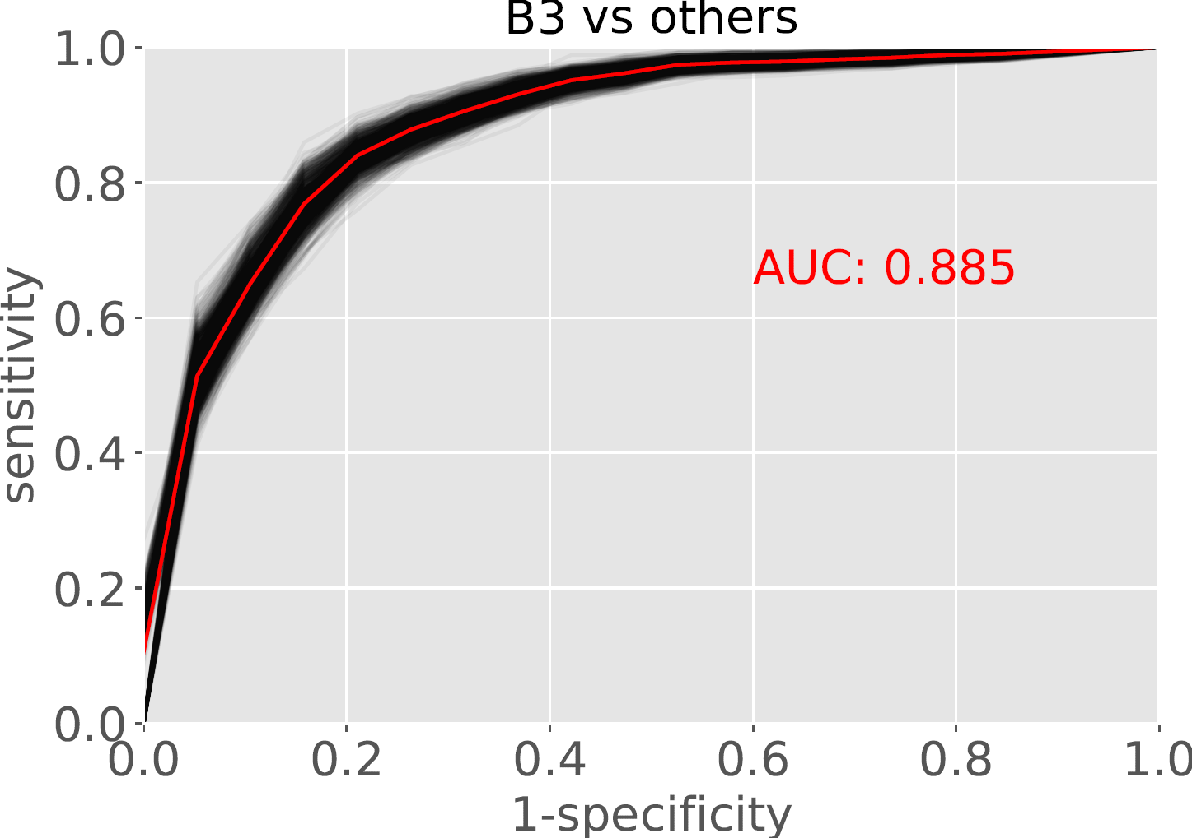
decision=B1

We achieve the best classification accuracy for B3 using clinical features alone.

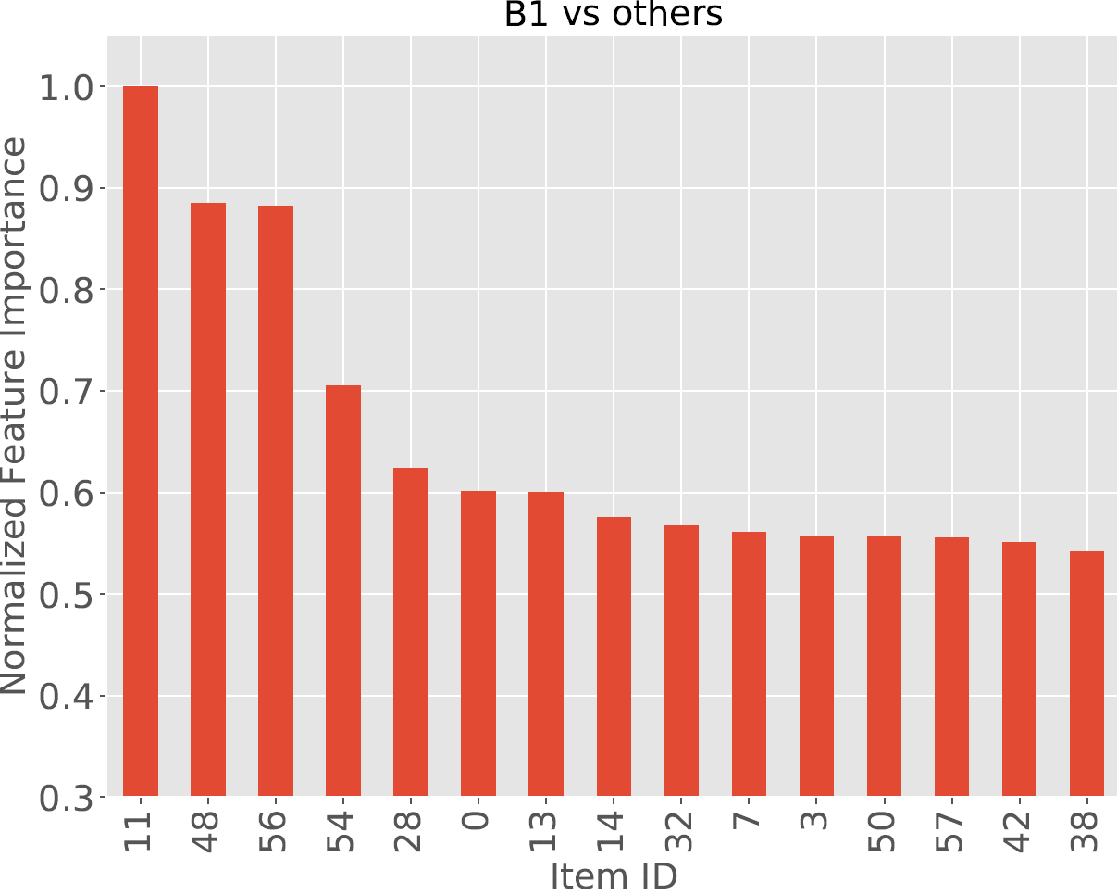
Individually, the classification of each biotype has very high sensitivity and specificity as illustrated by the following ROC curves and their associated AUCs, all greater than 0.85.

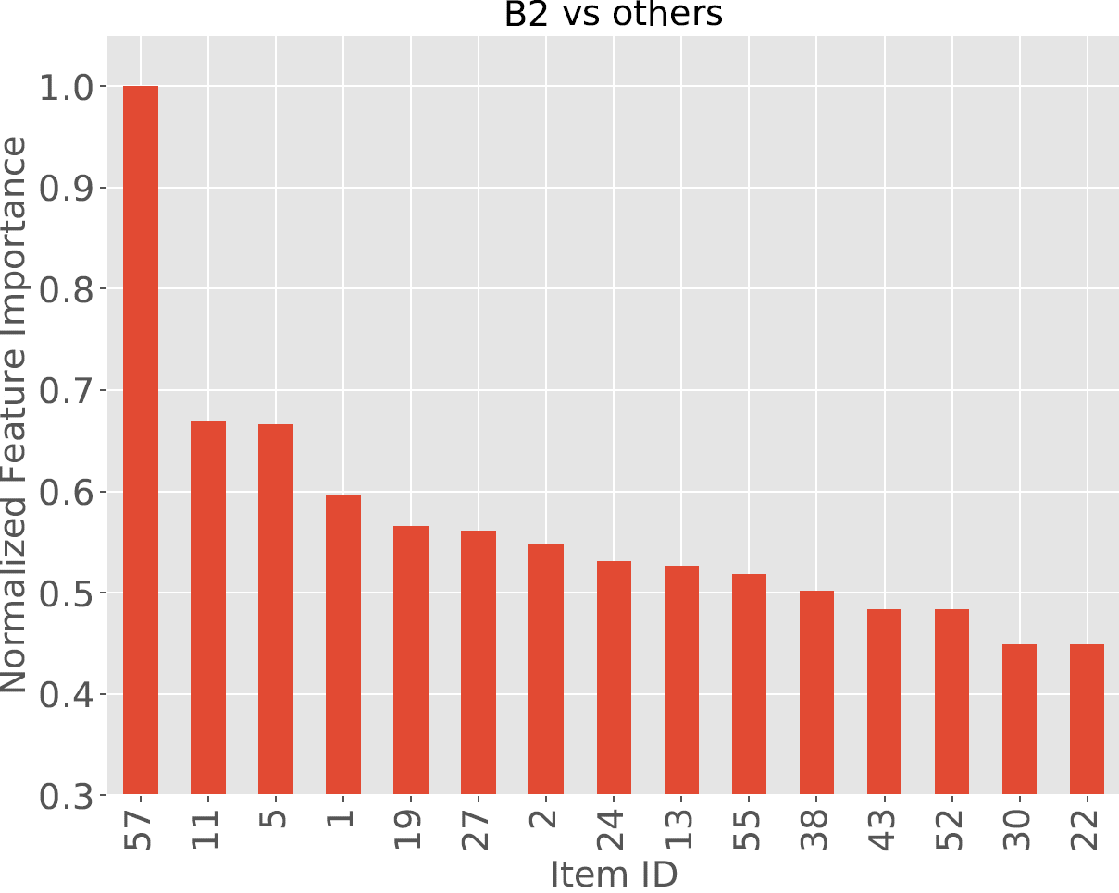


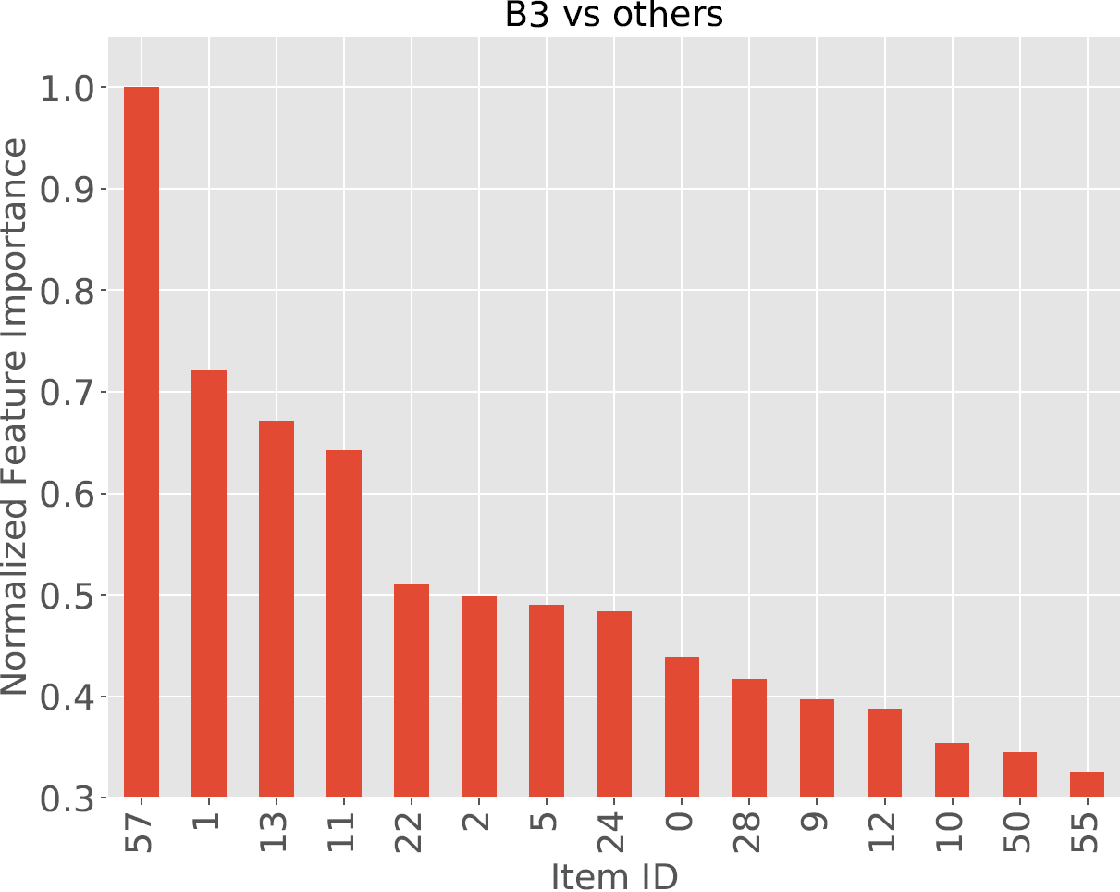




The following graphs provide individual feature importance out of the 58 clinical symptoms.







The features (58 clinical symptoms) are in the order of:

* PNASS
  + Positive 0-6
  + Negative 7-13
  + General 14-29
* MADRS 30-39
* Young 40-50
* SFS 51-57