**A Guide to the ENGAGE Data in its current state**

Sarah Chang

1. **Comprehensive Data Dictionary** 
   1. All the ENGAGE variables, where they come from, what circuits they are hypothesized to elicit, etc
2. **PAMF**
   1. These are the PAMF variables (ie. Not accounted for in the Comprehensive Data Dicionary, since that’s ENGAGE variables only) There are A LOT of variables in here! I’ve included a data dictionary which will hopefully clarify what you might work with.
3. **Neuroimaging**
   1. The reduced\_biotype files are raw, meaning they have not gone through quality control in this state. Depending on the aims of your project, I think it would be worth asking what kind of quality control would fit your analyses/or if we want to be consistent across everything, I can walk you through how we’re currently doing QC.
   2. No imaging data has been proceeded for 24MO in the reduced csv format.
4. **Medication**
   1. In long format. Each medication that a participant was prescribed is described relative to their baseline scan for ENGAGE. -5 in the START\_DATE\_Ndays\_ENGAGE column would mean that the participant started the medication 5 days prior to their first scan. Likewise, 5 would mean they started the medication 5 days after their first scan.
5. **Webneuro**
   1. You’ll notice that the subject numbers in the DISC100288…csv spreadsheet are different than the subNum variables, the mapping between the two can be found in ENGAGE\_subNum\_wnlogin\_mapping.csv.
   2. Also suffix matches to the order in which the participant did Webneuro. For example, SOBC-00002 completed webneuro three times, and you see this as 1,2,3 in the suffix column. You cannot know for sure for what visits these were completed – so he or she could have completed 1 at baseline, 2 at 6 months, and 3 at 24 months. The best way to know would be to compare the RecDate to their scan date, as recorded in Redcap.
   3. The \_norm variables do not match the directionality of the regular variables, so be sure to check before you run analyses if that’s important!
6. **Self-Report**
   1. To be completed today!
7. **Virtual Reality**
   1. Stored on Box, due to the large files generated from screen grab videos.
   2. N1/N2 = negative scene, P1/P2 = positive scene, R = self-reflective scene
8. **win-unified-2018-08-30\_share.csv**
   1. This is Carlos’ clean data set – ie. You have the winsorized biotypes (win\_biotype\_type\_score) variables as well as some of the important variables for previous analyses here. The webneuro values are aligned properly to the timepoint.