



Supplementary Figure 1. The group-wise changes in MADRS score from the first neuropsychological assessment battery (Baseline) to the third (Follow-up) in a subgroup of patients with fMRI data. The blue line indicates the mean change in MADRS scores in the verum group (N= 9), and the orange line shows that in the sham group (N=10), with the error bars representing standard error. We found a significant main effect of session sequence (baseline, second, and third) ($F_{1,17}= 23.7$, $p<0.001$), main effect of group (verum versus sham) ($F_{1,17}=5.8$, $p=0.028$), and group-by-session sequence interaction effect ($F_{1,17}= 7.3$, $p=0.008$) on the MADRS scores. In the post hoc t-tests, no difference was found in baseline ($t=0.2$, $p=0.840$), but significant group differences were found in end of treatment and follow-up ($t=2.6$, $p=0.015$ and $t=2.9$, $p=0.009$, respectively). In the verum group, we found a significant difference in MADRS scores between baseline and end of treatment ($t=4.7$, $p=0.001$) and baseline and follow-up ($t=4.4$, $p=0.002$) but no difference between end of treatment and follow-up ($t=2.2$, $p=0.063$). Similarly, in the sham group, the MADRS scores decreased between baseline and end of treatment ($t=2.5$, $p=0.036$), though no significant difference was found between baseline and follow-up ($t=2.1$, $p=0.069$) and end of treatment and follow-up ($t=0.8$, $p=0.430$). MADRS, Montgomery-Åsberg Depression Rating Scale; fMRI, functional magnetic resonance imaging.