SI1: supplement to 4.3 Statistical Meta-Analyses

The meta-analyses pooled each sample’s Cohen’s $d$ group effect sizes, for each ROI, using an inverse variance-weighted random-effects model as implemented in the R package *metafor* (version 1.9-1)\(^1\). Random effects models, compared to fixed effect models, do not make the assumption of a true (the same) effect size for each study. They estimate the mean of a distribution of effect sizes, allowing effect sizes to vary across studies due to study-specific differences (e.g., mean age). Random effects models therefore weigh within-study as well as between-study variance in the pooled effect size estimates and protect against dominating effects of the largest samples in the meta-analysis\(^2\). The random-effects models were fit using the restricted maximum likelihood method\(^3\); for additional model details see\(^4\). For the regions with a priori directional hypotheses based on the literature (amygdala, hippocampus, accumbens, thalamus, intracranial, lateral ventricle, and pallidum) the significance threshold for effect sizes of group contrasts was set at $p<0.05$ (one-tailed). We also indicate which of the findings pass the more conservative Bonferroni multiple comparison corrected threshold of $p<0.0056$ (two-tailed); the latter is based on $p=0.05/9$ (multiple comparison correction for 8 subcortical regions + ICV). Cohen’s $d$ effect sizes of 0, 0.2, 0.5, or 0.8, are considered to reflect no, small, moderate, or large effects, respectively\(^5\).

In addition to Cohen’s $d$ effect size estimates, *metafor* computes other measures, including standard errors, p-values, confidence intervals (CIs), and measures of heterogeneity (e.g., $I^2$). $I^2$ ($100\% \times (\text{Cochran’s } Q - \text{df}) / \text{Cochran’s } Q$) describes the percent variance across studies that is associated with heterogeneity rather than chance and has the principal advantage over other methods of heterogeneity estimation in that it is independent of a) the size of the meta-analysis,
b) the types of studies included in the meta-analysis, and c) the outcome data used in the meta-analysis and hence can readily be compared across meta-analyses studies. $I^2$ values of 0%, 25%, 50%, and 75% are considered reflective of no, low, moderate, and high variability/heterogeneity in effect size estimates across studies.