1
2 1. eMethods - Assessing Intraclass Correlation Coefficients (Unadjusted and Adjusted)

3 2. eTable 1. Intraclass Correlation Coefficients (ICCs) for the secondary ROIs.

4 3. eTable 2. Predicted Sample Size Estimates for the Left and Right hippocampal ROIs.

5 4. eFigure 1. Stability of whole-brain SPM maps

6

eMethods - Assessing Intraclass Correlation Coefficients (Unadjusted and Adjusted)

Unadjusted Intraclass Correlation Coefficients. We based computation of the intraclass correlation coefficient (ICC), in its unadjusted form, on Shrout and Fleiss (1979). Although other analysis designs are possible, we discuss these issues with respect to our situation in which we assessed test-retest reliability over time for a sample of subjects. Specifically, the formula we used is the one referred to by Shrout et al. as "ICC(2,1)"

$$ICC(2,1) = (BMS - EMS) / (BMS + (k - 1)EMS + k(TMS - EMS)/n)$$

where, using common analysis of variance terms,

BMS = the mean square (sum of squares/degrees of freedom) for the subjects,
TMS = the mean square for the Time factor,
EMS = the mean square for error, which in this case is the mean square for the interaction of Subjects X Time,
k = number of time points of assessment, and
n = number of subjects.

For this formula and its interpretation, the time points of assessment are considered randomly selected from a universe of time points, although the same points of time are used for all the subjects, and inference is intended to be made to the larger universe.

The only other type of ICC discussed by Shrout et al. relevant here is one they term "ICC(3,1)", in which time points are considered fixed and results are only relevant to those specific time points used in the study.

One way in which an ICC can be viewed is as a ratio of legitimate variance of interest our instrument is intended to measure relative to total variance of scores that includes this legitimate variance plus "error variance".

The numerator of the ICC formula, is often a function of differences between subjects in their mean levels across time of the variable being assessed, something considered real individual differences that we want the instrument to reliably assess. If baseline demographics or clinical variables (e.g., CDRsb) are responsible for some of this inter-subject level variance, they are producing part of the real inter-subject variance to be measured, yet it may not be imperative that such components be separated out for purposes of computing the ICC.

The denominator of the ICC formula, on the other hand, often includes a number of components of different kinds of "error" variance, depending on the type of ICC being computed. For the type of ICC we used, any aggregated across-subjects mean which changes from time assessment to time assessment would be considered error variance if there is not expected to be any test-retest change in the variable of interest, in the given situation for purposes of the reliability assessment. In addition, any subject X time interaction (essentially the inverse of correlation) is considered an additional separate component of error variance, in that a reliable instrument should not give values that change in different directions across time for different subjects. The denominator of the kind of ICC we are using is basically the sum of the variance due to mean change across time and the subject X time interaction variance (in addition to the inter-subject mean variance which is in the numerator also).

An important implication of our use of Shrout et al.'s ICC(2,1), rather than ICC(3,1), is that in the latter case, change in the mean from time point to time point is not included in the denominator, and only the subject X time interaction is included (and inter-subject variance). This means our ICC assesses more than the mere "consistency" or correlation across time indexed by ICC(3,1), but rather the actual "agreement" in values from time point to time point, which is negatively affected not only by lack of correlation (subject X time interaction) but also time to time mean level variation.

The number "1" in ICC(2,1) and ICC(3,1) refers to the reliability being relevant to a single score measurement at a time point, whereas a higher number k>1 in place of the "1" indicates that the ICC is indexing the reliability of a mean score across the k time points. The latter ICC is almost always larger in value than that for a single time point score.

Adjusted Intraclass Correlations. The ICC Adjusted for Covariate X Time Instability. Any significant interaction between a baseline covariate (e.g., demographic or clinical variable like the CDRsb) and time is assumed to

constitute real variance that is being reliably assessed by the instrument (e.g., BOLD-fMRI signal). However, unlike level effects of baseline values of a covariate, this interaction variance masquerades as part of the subject X time interaction error variance and therefore inaccurately inflates the denominator of the ICC formula unless it is estimated and removed (in our case by using regression analysis and separation of residuals). Our removal of this otherwise spurious deflation of the ICC presupposes that in the analysis of the study proper, the same interaction of baseline covariate X time will be assessed and separated out if found significant because it will likewise contribute to error variance in that case also and may even confound differences in group change over time if the groups differ in levels of the covariate and the covariate interacts with time. (Removal of significant baseline covariate level variance in the study proper is also assumed although this will otherwise only obscure and confound group mean level differences, and not impact change over time). In computing "adjusted ICCs", we chose the more conservative approach of not adding into the numerator of the ICC, the covariate X time interaction variance we removed from its denominator.

eTable 1. Intraclass Correlation Coefficients (ICCs) for the secondary ROIs.

	Individual ICC[b] For baseline to week 12 scans (mean ICC)[c]	
ROI Measure[a] (NvF)	**Raw ICC**	**Adjusted ICC[d]**
L Precuneus EXT	0.40 (0.57)	0.41 (0.58)
L Precuneus MAG	0.57 (0.73)	0.60 (0.75)
R Precuneus EXT	0.33 (0.49)	0.33 (0.49)
R Precuneus MAG	0.47 (0.64)	0.48 (0.65)
L Post Cing. EXT	0.42 (0.59)	0.44 (0.61)
L Post Cing. MAG	0.54 (0.70)	0.54 (0.70)
R Post Cing. EXT	0.32 (0.48)	0.32 (0.48)
R Post Cing. MAG	0.46 (0.63)	0.47 (0.64)

Abbreviations: NvF, Novel versus Fixation face-name contrast; EXT, Extent; MAG, Magnitude;
[a]ROI Measure (1st column) lists ROI and activation measure (EXT and MAG);
[b]Individual ICCs (2nd column) show raw individual-ICCs (2nd column, 1st sub-column) for T_1 (baseline = week 0) to T_3, week 0-12, fMRIs and adjusted individual ICCs (2nd column, 2nd sub-column).
[c]Mean ICCs for the "average" (mean) score across the time-points are shown in parenthesis. [d]Adjusted ICCs account for CDR-sb by time interactions.

eTable 2. Predicted Sample Size Estimates for the Left and Right hippocampal ROIs.

Predicted Sample Size To detect 12-week change from baseline				
ROI Measure (NvR)	**L Hipp. EXT**	**L Hipp. MAG**	**R Hipp. EXT**	**R Hipp. MAG**
Sample sizes required to estimate 25% Change				
70% power	40	385	37	40
80% power	51	489	46	50
90 % power	67	653	61	66
Sample sizes required to estimate 50% Change				
70% power	12	98	11	12
80% power	15	124	14	14
90 % power	19	165	17	19
Sample sizes required to estimate 75% Change				
70% power	7	45	6	7
80% power	8	56	8	8
90 % power	10	75	9	10

Predicted sample size estimates required to detect 25%, 50% and 75% 12-week changes from baseline fMRI measure with 70% , 80% and 90% power assume a 2-sided- α of 0.05 (i.e. bidirectional change from baseline) and utilize adjusted ICCs values.
EXT, Extent; MAG, Magnitude;
The contrast is NvR (Novel versus Repeated) contrast.

eFigure 1. Stability of whole-brain SPM maps. N>R contrast maps (p<0.001, 5-voxel extent) for the same template coordinate (-24, -24, -9) at baseline (T_1, week 0) for ALL subjects (n=24).