Distance Covariance Statistic

Distance covariance (dCov) is a recently developed multivariate technique that tests the statistical independence of two vectors with arbitrary dimensions {Székely, 2007 #112}:

$$dCov^2(X,Y) = \left\|\varphi_{X,Y}(x,y) - \varphi_X(x)\varphi_Y(y)\right\|^2$$

where $X \in \mathbb{R}^p, Y \in \mathbb{R}^q$ (i.e., X is a vector of p real numbers and Y is a vector of q real numbers) are random variables of arbitrary dimension with characteristic functions φ_X and φ_Y, respectively, and ||.|| denotes the Euclidean norm (the straight line, or, “as-crow-flies” distance between two points). The dCov statistic is zero if and only if the random vectors are independently distributed, and increasingly positive otherwise. Importantly for investigating fMRI connectivity, and in contrast to classical statistics such as MANOVA, dCov is applicable even when the dimension exceeds the sample size, or when observations are interrelated.

As an example application, using vectors that summarize fMRI connectivity and cognitive measures, let *X* be a vector of cognitive measurements, and let *Y* be a vector of correlations between a voxel and each ICN. In this case, dCov will measure association between the two vectors at the population level. In the applications below, X will either be the scalar MCCB domain score, or a scalar for the individual MCCB working memory subdomains. Y will either be the vector of correlations to ICNs for a single voxel, or an individual scalar correlation for a voxel.

Distance covariance can be calculated from a sample of subjects in a straightforward manner. This involves first constructing two distance matrices, for the subjects in X and Y. Note that, since the same subjects are present in both, a row and column in each matrix corresponds to the same subject. Next, both matrices are centered by subtracting their row and column means. Finally, using the correspondence between rows and columns, the two matrices are multiplied together element-wise, summed, and scaled to the product of the original distance matrix means.

Formally, from the joint distribution (X,Y) = {(X_k, Y_k): k = 1,…,n}, define

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}$$

where

$$a_{kl} = \|X_k - X_l\|$$

$$\bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{kl}$$

$$\bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}$$

$$\bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$$

and n is the number of subjects. Similarly, define $b_{kl} = \|Y_k - Y_l\|$ and $B_{kl} = b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot}$, for k, l = 1,…,n. The sample dCov statistic is then

$$\widehat{dCov} = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}\, B_{kl}$$

Székely et al.{Székely, 2007 #112} demonstrated that $\widehat{dCov} \underset{a.s.}{\longrightarrow} dCov$, as well as that the test statistic $n * \widehat{dCov}^2/(\bar{a}_{..} * \bar{b}_{..}) \underset{D}{\rightarrow} Q$, where *Q* is a positive-semidefinite quadratic form of centered Gaussian random variables with *E(Q)* = 1.

Hypothesis testing on the sample dCov is carried out using permutation test.

REFERENCE

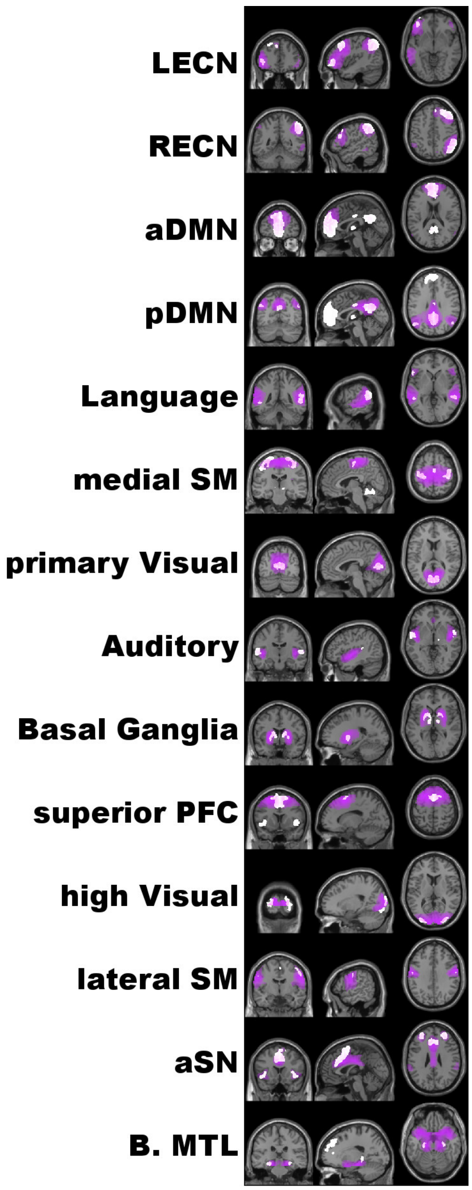
Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex 22, 158-165.

Supplementary Table 1: MATRICS Consensus Cognitive Battery test scores. Behavioral results for all cognitive tests included in the MCCB, with associated cognitive domains in parenthesis. All test scores corrected with MCCB software for age, gender, and centered and scaled relative a population of healthy control subjects (mean=50, sd=10). Trails A: Trail Making Test-part A, BACS-SC: Brief Assessment of Cognition in Schizophrenia-Symbol Coding subtest, HVLT-R: Hopkins Verbal Learning Test-Revised, WMS-III SS: Wechsler Memory Scale-III Spatial Span, LNS: Letter-Number Span, BVMT-R: Brief Visuospatial Memory Test-Revised, NAB mazes: Neuropsychological Assessment Battery mazes subtest, MSCEIT: Mayer-Salovey-Caruso Emotional Intelligence Test, CPT-IP: Continuous Performance Test-Identical Pairs.

Supplementary Table 1: MATRICS Consensus Cognitive Battery test scores

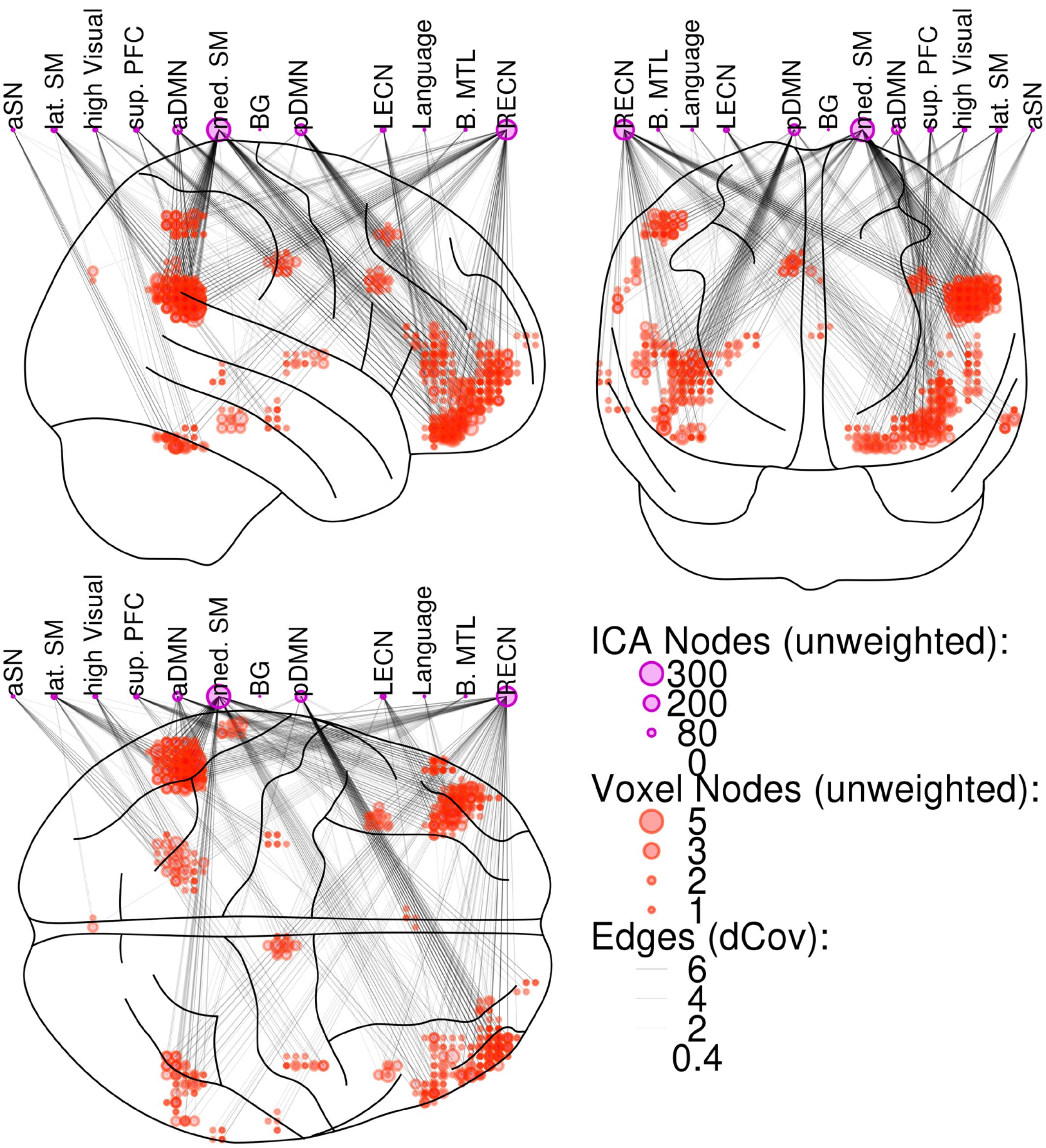
	Mean	SD
Test (Cognitive Domain):		
Trails A (Speed of Processing)	39.89	12.33
BACS-SC (Speed of Processing)	41.39	16.65
HVLT-R (Verbal Learning)	40.29	10.75
WMS-III SS (Working Memory)	48.21	10.08
LNS (Working Memory)	41.07	12.01
NAB mazes (Reasoning and Problem Solving)	46.29	9.15
BVMT-R (Visual Learning)	45.64	11.67
Category Fluency (Speed of Processing)	42.86	9.43
MSCEIT (Social Cognition)	42.75	12.86
CPT-IP (Attention/Vigilance)	36.82	12.84
Cognitive Domain:		
Speed of Processing	38.61	12.46
Attention/Vigilance	36.82	12.84
Working Memory	43.43	11.56
Verbal Learning	40.29	10.75
Visual Learning	45.64	11.67
Reasoning/Problem Solving	46.29	9.15
Social Cognition	42.75	12.86
Overall Composite	36.82	13.14

Supplementary Figure 1: Intrinsic Connectivity Networks extracted by ICA. All thresholded ICA spatial maps (in purple), displayed on top of ICN template matches (in white) from (Shirer et al., 2012). LECN: Left Executive Control Network, RECN: Right Executive Control Network, aDMN: anterior Default Mode Network, pDMN: posterior Default Mode Network, SM: Sensorimotor Network, PFC: prefrontal cortex, aSN: anterior Salience Network, B. MTL: bilateral Medial Temporal Lobe network.

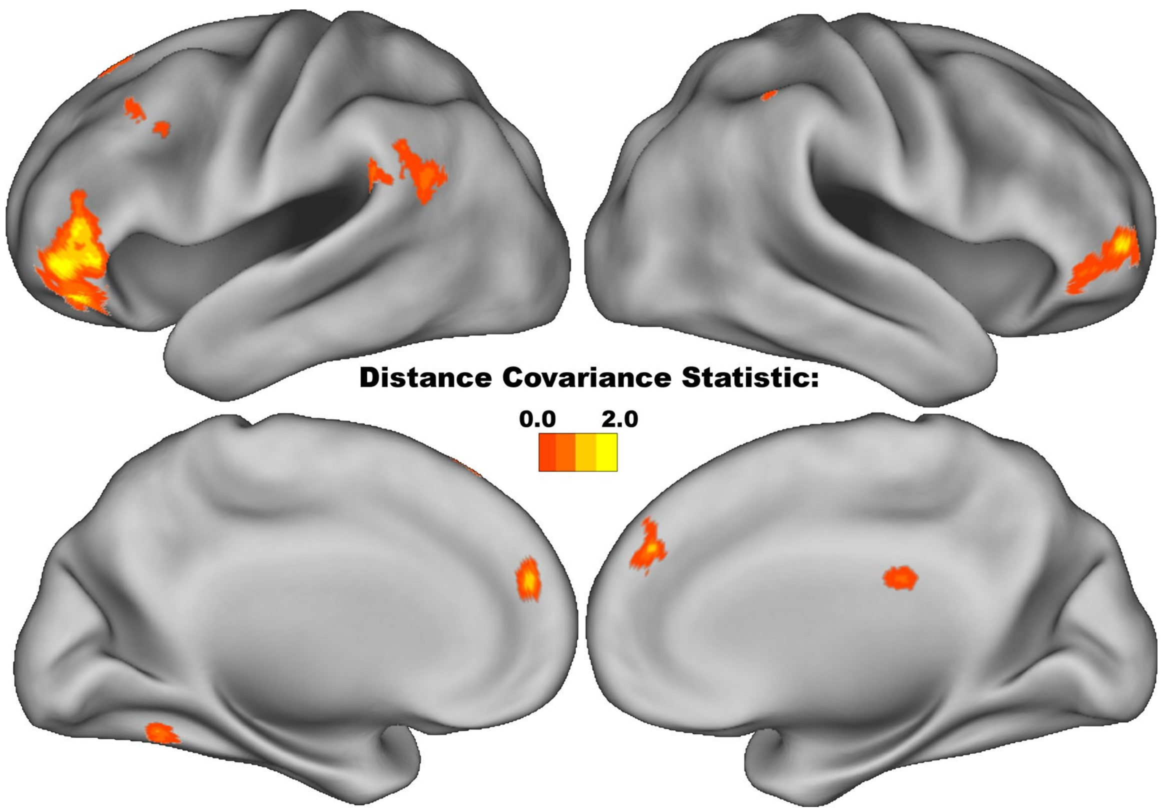


Data Supplement for Wylie et al., Association of Working Memory With Distributed Executive Control Networks in Schizophrenia.
J Neuropsychiatry Clin Neurosci (doi: 10.1176/appi.neuropsych.18060131)

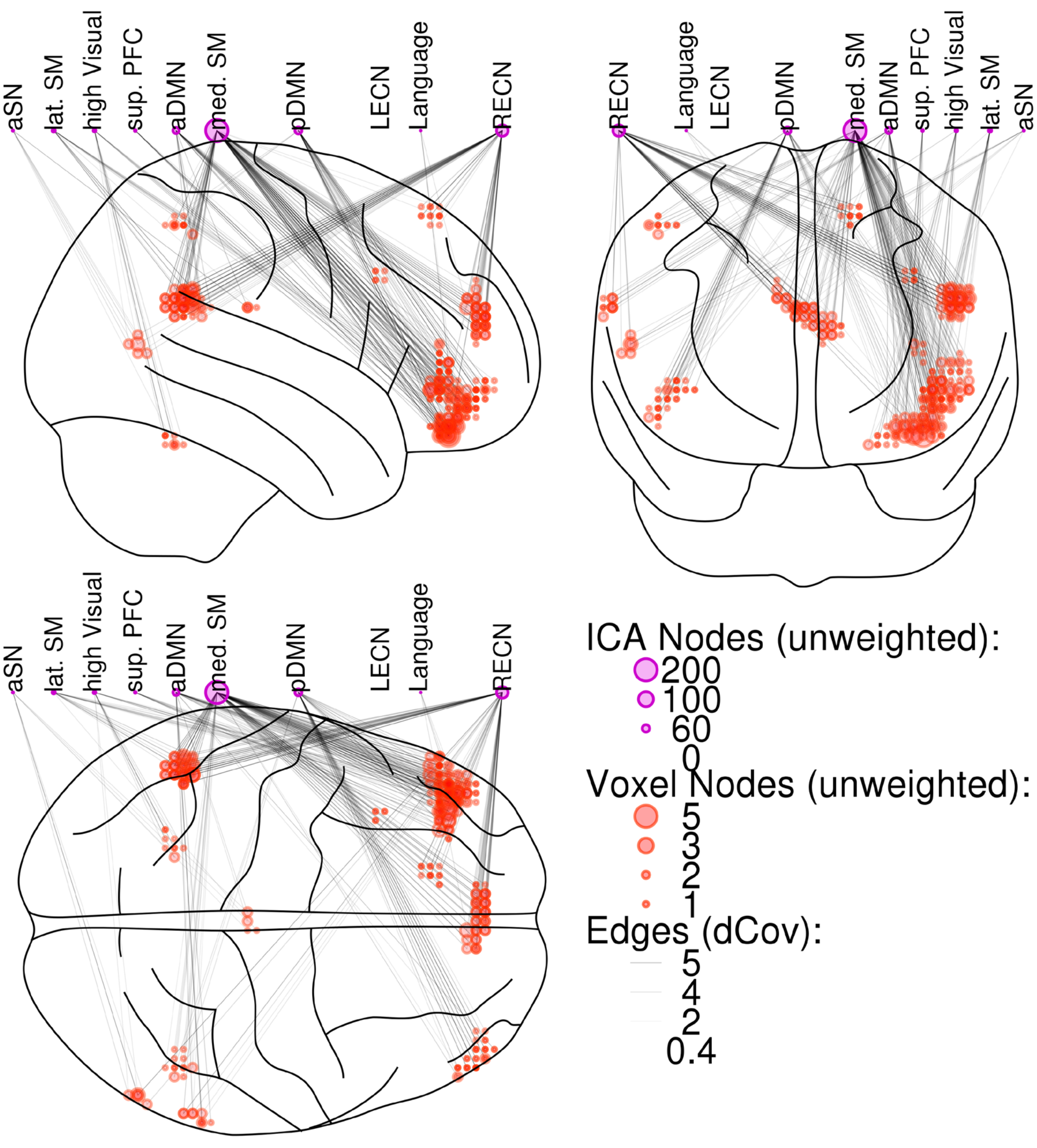
Supplementary Figure 2: Connectivity significantly associated with Working Memory, with additional ICN nodes labeled compared with Figure 7. All significant connections (Edges, grey lines), between voxels (Voxel nodes, red circles) and ICNs (ICA nodes, purple circles) associated with MCCB working memory test performance. All voxels & connections significant at p<0.05, corrected, are plotted. Node size is proportional to unweighted degree, edge size proportional to Distance Covariance statistic.



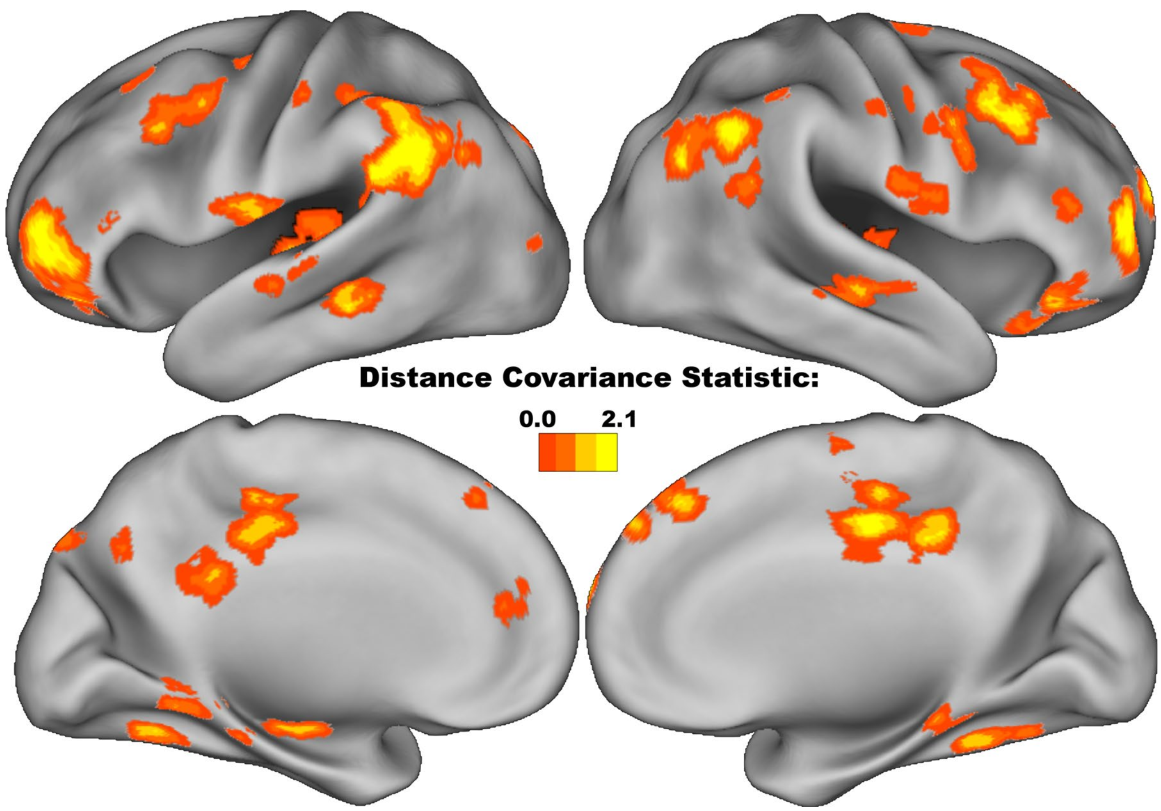
Supplementary Figure 3: Anatomical associations with the Letter-Number Span, the MCCB test of verbal working memory. All voxels, with connectivity is associated with LNS test scores using Distance Covariance statistic ($p<0.05$, corrected), displayed as cortical surface renderings. Compared to overall MCCB Working Memory domain results (Figure 5), the LNS verbal subtest was more associated with left-lateralized regions such as the left IFG and TPJ.



Supplementary Figure 4: Connectivity significantly associated with the Letter-Number Span, the MCCB test of verbal working memory. All significant connections (Edges, grey lines), between voxels (Voxel nodes, red circles) and ICNs (ICA nodes, purple circles) associated with LNS test performance. All voxels & connections significant at $p<0.05$, corrected, are plotted. Node size is proportional to unweighted degree, edge size proportional to Distance Covariance statistic.



Supplementary Figure 5: Anatomical associations with the WMS-III SS, the non-verbal test of MCCB working memory. All voxels, whose connectivity is associated with WMS-III SS test performance using Distance Covariance statistic ($p<0.05$, corrected), displayed as cortical surface renderings. Compared to the verbal LNS results (Supplementary Figure 3), WMS-III SS associations are more diffuse and bilateral.



Supplementary Figure 6: Connectivity significantly associated with the WMS-III SS, the test of MCCB non-verbal working memory. All significant connections (Edges, grey lines), between voxels (Voxel nodes, red circles) and ICNs (ICA nodes, purple circles) associated with WMS-III test performance. All voxels & connections significant at $p<0.05$, corrected, are plotted. Node size is proportional to unweighted degree, edge size proportional to Distance Covariance statistic.