



Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls

Deanna Greenstein, Brian Weisinger, James D. Malley, Liv Clasen and Nitin Gogtay

Standard Text:	This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.
Journal Name:	Frontiers in Psychiatry
ISSN:	1664-0640
Article type:	Original Research Article
Received on:	27 Feb 2012
Frontiers website link:	www.frontiersin.org

**Using multivariate machine learning methods and structural
MRI to classify childhood onset schizophrenia and healthy
controls**

Deanna Greenstein^{*1}, James D. Malley², Brian Weisinger¹, Liv Clasen¹, Nitin Gogtay¹

1 Child Psychiatry Branch, NIH/NIMH Bethesda, MD

2 CIT/NIH, Bethesda, MD

Correspondence:

Deanna Greenstein
10 Center Drive
Room 3N202
Bethesda, Maryland 20892
Phone: (301) 435-5553
Email: greenstd@mail.nih.gov

Abstract

Introduction: Multivariate machine learning methods can be used to classify groups of schizophrenia patients and controls using structural magnetic resonance imaging (MRI). However, machine learning methods to date have not been extended beyond classification and contemporaneously applied in a meaningful way to clinical measures. We hypothesized that brain measures would classify groups, and that increased likelihood of being classified as a patient using regional brain measures would be positively related to illness severity, developmental delays and genetic risk. Methods: Using 74 anatomic brain MRI sub regions and Random Forest, a machine learning method, we classified 98 childhood onset schizophrenia (COS) patients and 99 age, sex, and ethnicity-matched healthy controls. We also used Random Forest to estimate the probability of being classified as a schizophrenia patient based on MRI measures. We then explored relationships between brain-based probability of illness and symptoms, premorbid development, and presence of copy number variation associated with schizophrenia. Results: Brain regions jointly classified COS and control groups with 73.7% accuracy. Greater brain-based probability of illness was associated with worse functioning ($p=0.0004$) and fewer developmental delays ($p=0.02$). Presence of copy number variation (CNV) was associated with lower probability of being classified as schizophrenia ($p=0.001$). The regions that were most important in classifying groups included left temporal lobes, bilateral dorsolateral prefrontal regions, and left medial parietal lobes. Conclusions: Schizophrenia and control groups can be well classified using Random Forest and anatomic brain measures, and brain-based probability of illness has a positive relationship with illness severity and a negative relationship with developmental delays/problems and CNV-based risk.

Keywords: Machine Learning, MRI, cortical thickness, schizophrenia

Introduction

Structural brain magnetic resonance imaging (MRI) studies of schizophrenia indicate widespread neuroanatomic abnormalities in cortical thickness, hippocampus, subcortical structures, and total brain measures (Byne, et al. 2009; Greenstein, et al. 2006; Mattai, et al. 2011; Narr, et al. 2005; Nesvag, et al. 2008; Shenton, et al. 2001; Steen, et al. 2006; van Haren, et al. 2011). Functional MRI and diffusion tensor imaging studies of schizophrenia also support brain dysfunction in schizophrenia involving multiple brain systems, emphasizing networks and connectivity dysfunction rather than brain regions acting in isolation (Bassett, et al. 2008; Lynall, et al. 2010; Meyer-Lindenberg, et al. 2005; Repovs, et al. 2011).

If schizophrenia is indeed a disorder of connectivity, then the capacity for identifying reliable neuroanatomic signatures of the disease may be reduced if regions are not considered jointly. However, traditional statistical methods (e.g., correlation, t-tests, ANOVA, logistic regression) explore group differences effectively but only within a region or voxel at a time (Sun, et al. 2009). Also, traditional model-based methods are limited when exploring how regions/voxels interact as these models quickly become overburdened when trying to combine predictors and all of their interactions from high dimensional MRI data sets (e.g., 6 predictors have over 60 effects when including all main effects and interactions). These statistical methods may also miss a signal from brain measures interacting in nonlinear, non-multiplicative ways.

In contrast, multivariate machine learning methods can utilize available information simultaneously to understand how variables jointly distinguish between groups. These methods have had previous success classifying schizophrenia and healthy controls using structural brain MRI data with classification accuracies ranging from 81% to 93% (Davatzikos, et al. 2005; Kawasaki, et al. 2007; Koutsouleris, et al. 2009; Sun, et al. 2009; Yoon, et al. 2007). However, no structural MRI study using multivariate machine learning methods has attempted to link multivariate brain-based classifier results with clinical measures in samples of patients with schizophrenia. This is important in that behavioral correlates can provide a clinical context for classifier results.

Here we use Random Forest (Breiman 2001) to contemporaneously classify groups using anatomic brain measures and correlate clinical and genetic information with classification scheme results. We selected Random Forest as it has error rates comparable to other approaches (Malley, et al. 2011b) while being able to determine the probability of illness based on the feature set of brain regions (Malley, et al. 2011a), (henceforth referred to as brain-based probability of illness). Of note, these probabilities can be used as a continuous measure containing more information than dichotomous classification to explore relationships with clinical correlates and risk factors for childhood onset schizophrenia (COS). Accordingly, we hypothesized that brain-based probability of illness would be positively associated with clinical measures of illness severity. To explore the idea that brain-based probability of illness would covary with other risks, we hypothesized positive associations between brain-based probability of illness and presence of copy number variations (CNVs) associated with the risk of

schizophrenia. Additionally, we hypothesized that measures of developmental delays which are considered risk factors under the neurodevelopmental model for schizophrenia (Rapoport, et al. 2005; Weinberger 1987) would also be positively associated with brain-based probability of illness.

Methods

Participants

All probands were subjects in an ongoing study of COS at the National Institute of Mental Health and met DSM IIIR/IV criteria for schizophrenia with the onset of psychosis before their 13th birthday. Exclusion criteria were a history of significant medical problems, substance abuse, or a premorbid IQ below 70. We obtained informed consent from parents of minors and participants over 18, and informed assent was obtained when possible. Further details of patient selection are described elsewhere (Kumra, et al. 1996; McKenna, et al. 1994).

We obtained MRI scans during each proband's initial inpatient stay and at subsequent two-year follow-up visits. For the purposes of this study, each patient's first good quality MRI scan (e.g., absence of visible motion artifacts) was selected to minimize length of illness and medication history for a total of 98 scans. The study was approved by the National Institutes of Health (NIH) institutional review board. Typically developing control participants were volunteers in a prospective study of normal brain development (see (Giedd 1999) for further details) also approved by the NIH institutional review board. The current control sample of unrelated 99 participants was selected to match the COS group on age, sex and ethnicity. Scans with moderate or severe motion artifacts and scans from participants with dental braces were excluded. See Table 1 below for demographic information.

Clinical measures

We used age-appropriate versions of the Global Assessment of Functioning Scale (GAS) (APA 1994; Shaffer, et al. 1983), the Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen 1984), and the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen 1983) to assess clinical symptoms in COS probands (intraclass correlation coefficients for all measures > 0.80). We restricted ratings to NIH inpatient medication-free assessments to approximate comparable rater, treatment and environmental effects across participants.

To assess developmental delays and problems, we used the 40-item Autism Screening questionnaire (ASQ) (Berument, et al. 1999). We also conducted a chart review of previous medical records for pre-illness and pre-prodrome academic, language, motor, and social developmental problems and delays. The chart review consists of 15 items (academic (2 items); social (3 items); language (6 items) motor (4 items)) scored 1 or 0 depending on presence or absence of delay/problem and is included in Appendix A. Reliability among 3 chart reviewers was adequate (intraclass correlation coefficients > 0.70).

187 Table 1: Sample demographics and clinical measures

	COS (n=98) mean(SD) or count	Controls (n=99) mean(SD) or count	Statistic	P value
Age at scan	14.46(3.40)	14.45(4.43)	t(df=195)=0.02	0.99
Vocabulary	6.37(3.49)	11.97(2.73)	t(df=180)=12.13	<0.001
Intracranial Volume	1474032(166557)	1476771(160495)	t(df=195)=0.12	0.91
Female Male	43 55	41 58	X ² (df=1)=0.12	0.73
Race				
Asian	5	4		
African American	29	23		
Hispanic	8	8	X ² (df=4)=1.4	0.84
Other	7	7		
White	49	57		
Inpatient Medication Free Rating Scales				
Scale for the Assessment of Positive Symptoms	48.44(22.03)			
Scale for the Assessment of Negative Symptoms	61.23(28.39)			
Global Assessment of Functioning	24.57(13.22)	-	-	-
Autism Screening Questionnaire	14.36(9.63)			
Developmental Chart Review (range=0-15)	3.88(2.9)			
Years ill at time of Scan	4.5 (2.96)			

188

189

190 Copy Number Variation

191 All subjects in our COS study were genotyped using Illumina 1M SNP chip
192 (www.illumina.com), and CNV detection was performed by using three algorithms: (1)
193 PennCNV Revision 220, (2) QuantiSNP v1.1, and (3) GNOSIS. Analysis and merging of
194 CNV predictions was performed with CNVision (www.CNVision.org). Twelve subjects
195 in the current sample have at least one CNV that has been independently associated with
196 risk of schizophrenia (1q21 (n=1), 2p16 (NRXN1) (n=1), 2p25(MYT1L) (n=2),
197 3p25(SRGAP) (n=2), 7q11 (n=1), 7q35 (CNTNAP2) (n=1), 15q11 (n=1), 16p13 (n=2),
198 22q11 (n=4)), (Bassett, et al. 2010; Ingason, et al. 2011; International Schizophrenia
199 Consortium 2008; Irmansyah, et al. 2008; Kirov, et al. 2009; Levinson, et al. 2011; Li, et
200 al. 2011; Moreno-De-Luca, et al. 2010; Stefansson, et al. 2008; Stone, et al. 2008). These
201 data were not collected for controls.

202

203 MRI acquisition and analysis

204 We obtained brain MRIs using a GE Signa 1.5T MR system (General Electric
205 Medical Systems, Milwaukee, Wisconsin). T1-weighted structural brain images were
206 collected using a 3D spoiled gradient recall (SPGR) sequence. Brain volumes consisted

of 124 1.5 millimeter axial slices with a .9375 millimeter in-plane resolution. Scanning parameters were TR= 24 milliseconds, TE=5 milliseconds, and a flip angle of 45 degrees.

The brains were processed using the FreeSurfer recon-all pipeline with default settings except for the number of non-uniformity correction iterations that were increased to six. We also used the default parcellation which uses the Desikan atlas (<https://surfer.nmr.mgh.harvard.edu/ftp/articles/desikan06-parcellation.pdf>). Cortical and subcortical volumes were measured automatically with FreeSurfer (version 5.1) (<http://surfer.nmr.mgh.harvard.edu>). This method has been described in detail elsewhere (Fischl, et al. 2002; Fischl, et al. 2004) and will only be briefly described here. Processing included motion correction and removal of non brain tissue using a hybrid watershed/surface deformation procedure (Segonne, et al. 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including the hippocampus and ventricles) (Fischl, et al. 2002; Fischl, et al. 2004), intensity normalization, tessellation of the gray-white matter boundary, automated topology correction (Fischl, et al. 2001; Segonne, et al. 2007), and surface deformation following intensity gradients to optimally place the gray-white matter and gray matter/CSF borders at the location where the greatest shift in intensity defines the transition to the other tissue class. Anatomic segmentation is based on the probability of the local spatial configuration of labels given the tissue class. This technique has previously been shown to be comparable in accuracy to manual labeling (Fischl, et al. 2002) and has been demonstrated to show good test-retest reliability across scanner manufacturers and field strengths (Han, et al. 2006).

The above procedure generated average cortical thickness for 68 frontal, temporal, parietal and occipital lobe regions, and bilateral lateral ventricle, thalamus and hippocampus volumes to yield the 74 variables we used as features in the machine learning analysis (below). Before the variables were used to classify, they were each residualized using a general linear model with sex, age, age-squared, and intracranial volume as independent variables.

Statistical Analysis

1. Classification: Random Forest

We used Random Forest (RF) (Breiman 2001) (http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm) as our multivariate machine learning method to predict group membership (COS or controls) with the 74 residualized brain measures (above) as features. RF's basic unit is a classification tree. RF works by selecting a random bootstrap subset of approximately 66% of the sample per tree and randomly selecting a subset of all features (or cortical regions) at each node of the tree. At each node, RF selects the variable that best splits data into two daughter nodes. This process allows for the cortical regions to work in concert while predicting the outcome region. RF determines prediction error using the out of bag sample (i.e., roughly 33% of participants not randomly selected to build a given tree) that is sent down a tree after it is grown. It is through this process of selecting bootstrap samples to build the tree and then using the out of bag sample to determine error and variable importance that RF minimizes overfitting and contains an internal

validation step. This internal validation component built into RF is similar to leave-one-out schemes and other cross-validation procedures.

RF provides three basic outputs: classification error, importance scores, and proximities. Classification error is the percent of times a participant (when out of bag) is incorrectly classified; subtracted from one, it is the percent of times a participant (when out of bag) is correctly classified. An importance score is the difference between out of bag error when a variable is randomly permuted and when the variable is not randomly permuted. So, if a variable's values are randomly permuted and the error rates do not go up, it is not a useful predictor, since it is no better than random noise. Importance scores can be transformed to z scores ((score-mean)/standard deviation) to ease interpretability. A proximity score is a measure of the frequency at which two out of bag participants are classified in the same terminal node. These proximities are used to form an $n \times n$ matrix where n is the number of subjects. This matrix can then be transformed into a distance matrix that can be visualized with multidimensional scaling.

Because the random components in RF (out-of-bag sampling, node-level permutation testing) can make the importance scores, proximity scores and error rates vary, we ran each of the above steps 1000 times and took the average values. We used the R package randomForest (Liaw and Wiener 2002) for all analyses and set the number of trees per forest at 300 as the plotted error rate was observed to stabilize before 300 trees. We set our terminal node size to 10 and number of variables randomly selected per node (aka mtry) to 10.

Finally, we utilized recent work (Malley, et al. 2011a) which transforms RF into a probability machine and allows RF to determine the probability of belonging to the COS group based solely on the 74 residualized brain measures. Briefly, we accomplish this by running RF in regression mode where we assign a value of 1 to COS participants and a value of 0 to controls. Exactly as in coin-tossing problems, the estimated average of these scores for each subject is the estimated probability for that subject. These probability estimates are known to be consistent, as opposed to the standard RF probability estimates (e.g., those available in the standard output of the RandomForest package) which have no known optimality (Biau in press; Malley, et al. 2011a). We ran this analysis 1000 times and took the average probability of being classified as COS per participant to correlate with clinical measures.

2. Classification: Logistic Regression

We computed 74 logistic regressions to determine univariate classification accuracy for each region. For each regression, we used regional cortical thickness as the independent measure (after residualizing regional thickness using age, age squared, sex and intracranial volume) and diagnostic group as the dependent measure. Statistical significance for logistic regression coefficients was determined after false discovery rate correction (Genovese, et al. 2002) ($q=0.05$).

3. Relationships between brain-based probability of illness, clinical correlates and schizophrenia risk factors

We used linear regression to assess the relationship between brain-based probability of illness and medication free clinical measures (GAS, SAPS total, SANS total) and developmental measures. We used a t-test to assess the group difference in mean probability of illness between COS participants who have a CNV independently associated with risk of schizophrenia and those who do not. For these analyses, we checked assumptions of linearity, normality, and homoscedasticity, and visually explored data for outliers and unrealistic data points.

Results

1. Machine Learning Multivariate Classifier

Classification accuracy.

The average classification error of the 1000 RF runs was 26.3% (SD=1%), yielding an average classification accuracy of 73.7%. When we randomly permuted group membership 1000 times and ran RF for each permutation, the null distribution and the non-permuted distribution did not overlap, indicating that the 73.7% classification accuracy is far better than chance (see Figure 1).

[Insert Figure 1 about here]

Importance measures.

The entire list of 74 importance z scores is reported in supplementary table 1. The 15 regions with an importance scores at least 0.5 SD above the mean are visually represented in Figure 2. As seen in Figure 2, bilateral frontal, left precuneus, and left temporal regions had the highest importance scores.

[Insert Figure 2 about here]

Multidimensional scaling of proximity matrix and probability machine results.

The multidimensional scaling (MDS) plot (Figure 3A) for the proximity matrix is a visual representation of the accuracy of the classifier; Geometric distances between people correspond to how often they are classified in the same group (closer points correspond to being classified in the same group frequently). The groups appeared well separated, corresponding to 73.7% classification accuracy. In addition, we have provided a color overlay which represents each participant's probability of being classified as COS based on RF run as a probability machine (Figure 3B).

[Insert Figure 3 about here]

2. Univariate logistic regressions

74 univariate logistic regressions yielded 55 significant odds ratios ($p \leq 0.03$), all of which survived false discovery correction. Of these, only right caudal middle frontal thickness ($p < 0.001$, classification accuracy=73.6%) was able to classify as well as RF, although five regions individually classified subjects with at least 70% accuracy (left caudal middle frontal, left rostral middle frontal, left parstriangularis, left precuneus). These regions were also among the 15 regions with the top RF importance scores, revealing overlap between univariate and multivariate classification.

Figure 4 illustrates the curvilinear relationship between univariate results and RF important scores (also see supplemental table 1 for all regions, their importance scores and univariate classification accuracies).

[Insert Figure 4 about here]

Of note were the regions with relatively weaker univariate effects (e.g., not among the top 20 univariate classifiers) and importance scores greater than 0.50 SDs above the mean. Such predictors included right parsopercularis, left bank of the superior temporal sulcus, left fusiform gyrus (importance Z scores= 1.34, 0.56, 0.94, respectively; univariate accuracy rate= 65%, 65%, 64.5%, respectively).

3. Clinical Correlates

1. Inpatient Medication-Free Ratings. Greater brain-based probability of being classified as COS was significantly associated with worse overall functioning during inpatient medication free baseline (GAS score= 0.0004) (see Figure 5). Positive relationships between probability of being classified as COS and negative and positive symptoms during inpatient medication-free baseline (greater probability associated with more symptoms) were statistical trends (SAPS p=0.07, SANS=0.09, respectively).

2. Schizophrenia Risk Factors

1. Developmental Measures. Greater brain-based probability of being classified as COS was significantly associated with fewer documented pre-illness academic, language, motor and social difficulties and delays (p=0.02) (see Figure 5). There was no relationship between probability of being classified as COS and scores on the Autism Screening Questionnaire (p=0.22).

[Insert Figure 5 about here]

2. Copy number variations. The 12 COS subjects who have a CNV that has been independently associated with risk of schizophrenia had a lower mean probability of illness (0.44(SD=0.23)) than patients who did not (n=86, mean = 0.64(SD=0.18) (t=3.398(df=96) p =0.001).

Discussion

Using a multivariate machine learning approach and measures of regional cortical thickness, bilateral hippocampus, thalamus and lateral ventricle volumes, we achieved good classification between COS patients and controls. We were also able to use all brain measures jointly to predict group membership, which is consistent with a current emphasis on brain systems and networks rather than regions in isolation. The regions that were most important in our multivariate classifier included temporal, dorsolateral prefrontal regions and medial parietal lobe: this is consistent with current univariate results and previous reports of gray

matter reductions and brain network abnormalities in these regions (Ellison-Wright and Bullmore 2009; Meyer-Lindenberg 2010; Shenton, et al. 2001; van den Heuvel, et al. 2010).

To our knowledge, we provide initial evidence that multivariate machine learning approaches can link probability of illness with clinical measures in a meaningful way. Specifically, here we link medication-free illness severity ratings, CNVs and developmental risk factors with nuanced, continuous information generated by machine learning at an individual level: i.e., what is the probability a person is affected given the features, rather than dichotomous affected/not affected output. For example, 52% and 85% chance of an event or diagnosis both declare for the event but clearly, there is more information available in the continuous percentage.

Consistent with our hypothesis we found a positive relationship between probability of illness base solely on brain measures and illness severity. Counter to our hypothesis, however, fewer premorbid academic, language, motor, and social developmental problems and having a CNV associated with schizophrenia were associated with a lower brain-based probability of being classified as schizophrenic. This suggests that there may be a relationship between schizophrenia patients who sustained a large genetic mutation on a pathway of unusual strength, reflected in more frequent early difficulties but with less neuroanatomic disturbance. However, caution is warranted when while interpreting the CNV group difference in probability of illness, as the group of CNVs is diverse and may not represent a single homogenous population.

We hypothesized that linear and/or non-linear relationships among brain regions would make the multivariate classifier superior to univariate classifiers. However, the current multivariate approach did not out-perform several univariate logistic regressions on a pure classification task, and our hypothesis was thus not confirmed. Specifically, right caudal middle frontal thickness alone performed as well as the multivariate classifier, and several other frontal and temporal regions had classification accuracies greater than 70%. At the same time, the strong curvilinear relationship between RF importance scores and univariate classification accuracy indicates that both approaches detect strong effects, and RF does so without incurring costs for correcting for multiple tests with unknown joint correlation structure or assumptions of normality and homoscedasticity. Also, some univariate effects that are not particularly accurate classifiers have relatively strong importance scores. This outcome suggests that the combination of univariate and multivariate methods can be used detect regions of relative importance when interacting with other regions but that do not classify particularly well when acting alone (e.g., right parsopercularis, left bank of the superior temporal sulcus, left fusiform gyrus). This is particularly important in an illness like schizophrenia, which is can be considered a disorder of dysconnectivity, as individual brain regions are unlikely to be affected in an isolated manner.

Our multivariate classification error rate of 73.7%, although good, is not high enough to warrant the use of MRI measures as a stand-alone diagnostic tool. While previous multivariate classification studies report upwards of 80% accuracy (Davatzikos, et al. 2005; Kawasaki, et al. 2007; Koutsouleris, et al. 2009; Sun, et al. 2009; Yoon, et al.

2007), clinical interview conducted by a skilled clinician still remains the most efficient, cost-effective diagnostic tool between healthy and psychotic patients. However, structural brain-based classifiers do appear to be relevant when the goal is to understand the most important neuroanatomic factors distinguishing diagnostic groups without encumbrances inherent in multiple tests and parametric test assumptions. Also, we recommend future studies using features from MEG, DTI and fMRI scans, as MEG, DTI, and fMRI data is collected specifically to detect active brain networks and connectivity. We believe this kind of study may be better suited than structural MRI to fully harness the power of multivariate methods' ability to capitalize on linear and nonlinear interactions. Also, when brain imaging features can classify cases and controls, researches can use methods like the ones currently employed to detect relationships between phenotypes and continuous probabilities from machines with brain-based (or fMRI, DTI, EEG, etc...) features. These relationships might otherwise be missed if the machine output is restricted to dichotomous classification.

Limitations of the current study include the lack of a validation sample, although COS is a very rare disorder and the current sample required several decades to acquire. Also, our assessment of developmental issues has two drawbacks: 1) retrospective chart reviews may miss relevant information that was never documented and 2) the ASQ assesses current functioning as well as premorbid development. Also, here we have chosen to use regional brain measures that provide less noise albeit lower resolution compared to the higher resolution voxel-wise measures. Despite these limitations, Random Forest appears to provide a means of distinguishing groups that has the advantage of linking clinical information and risk factors and classification using multiple brain regions jointly.

Author Disclosures

1. Contributors

Deanna Greenstein wrote the manuscript and performed the statistical analyses
James Malley provided consultation and technical report on statistical and machine
learning methods
Liv Clasen selected the sample
Brian Weisinger assisted with data management and graph preparation
Nitin Gogtay served as senior author and assisted with manuscript edits

2. The authors report no conflicts of interests

References

- Andreasen NC. (1983): Scale for the Assessment of Negative Symptoms (SANS). Iowa City: University of Iowa.
- Andreasen NC. (1984): Scale for the Assessment of Positive Symptoms (SAPS). Iowa City: University of Iowa.
- APA. (1994): Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th edition. . Washington DC: American Psychiatric Association.
- Bassett AS, Scherer SW, Brzustowicz LM. (2010): Copy Number Variations in Schizophrenia: Critical Review and New Perspectives on Concepts of Genetics and Disease. *American Journal of Psychiatry* 167(8):899-914.
- Bassett DS, Bullmore E, Verchinski BA, Mattay VS, Weinberger DR, Meyer-Lindenberg A. (2008): Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia. *The Journal of Neuroscience* 28(37):9239-9248.
- Berument SK, Rutter M, Lord C, Pickles A, Bailey A. (1999): Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry* 175(5):444-51.
- Biau G. (in press): Analysis of a random forests model *Journal of Machine Learning Research*.
- Breiman L. (2001): Random Forests. *Machine Learning* 45(1):5-32.
- Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y, Hughett P, Turetsky BI, Gur RE. (2005): Whole-Brain Morphometric Study of Schizophrenia Revealing a Spatially Complex Set of Focal Abnormalities. *Arch Gen Psychiatry* 62(11):1218-1227.
- Fischl B, Liu A, Dale AM. (2001): Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *Medical Imaging, IEEE Transactions on* 20(1):70-80.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. (2002): Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* 33(3):341-355.
- Genovese CR, Lazar NA, Nichols T. (2002): Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage* 15(4):870-878.
- Giedd JN, Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L. (1999): Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience* 2:861–863.
- Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B. (2006): Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage* 32(1):180-194.
- Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietilainen OP, Buizer-Voskamp JE, Strengman E, Francks C, Muglia P, Gylfason A, Gustafsson O, Olason PI, Steinberg S, Hansen T, Jakobsen KD, Rasmussen HB, Giegling I, Moller HJ, Hartmann A, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Bramon E, Kiemenev LA, Franke B, Murray R, Vassos E, Touloupoulou T, Muhleisen

581 TW, Tosato S, Ruggeri M, Djurovic S, Andreassen OA, Zhang Z, Werge T,
582 Ophoff RA, Rietschel M, Nothen MM, Petursson H, Stefansson H, Peltonen
583 L, Collier D, Stefansson K, St Clair DM. (2011): Copy number variations of
584 chromosome 16p13.1 region associated with schizophrenia. *Mol*
585 *Psychiatry* 16(1):17-25.

586 InternationalSchizophreniaConsortium. (2008): Rare chromosomal deletions and
587 duplications increase risk of schizophrenia. *Nature* 455(7210):237-241.

588 Irmansyah, Schwab SG, Heriani, Handoko HY, Kusumawardhani A, Widyawati I, Amir
589 N, Nasrun MWS, Holmans P, Knapp M, Wildenauer DB. (2008): Genome-wide
590 scan in 124 Indonesian sib-pair families with schizophrenia reveals genome-wide
591 significant linkage to a locus on chromosome 3p26-21. *American Journal of*
592 *Medical Genetics Part B: Neuropsychiatric Genetics* 147B(7):1245-1252.

593 Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou S-Y, Nakamura K, Matsui M,
594 Sumiyoshi T, Seto H, Kurachi M. (2007): Multivariate voxel-based morphometry
595 successfully differentiates schizophrenia patients from healthy controls.
596 *NeuroImage* 34(1):235-242.

597 Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. (2009):
598 Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr Bull* 35(5):851-4.

599 Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J,
600 Schmitt G, Zetzsche T, Decker P, Reiser M, Moller H-J, Gaser C. (2009): Use of
601 Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental
602 States of Psychosis and Predict Disease Transition. *Arch Gen Psychiatry*
603 66(7):700-712.

604 Kumra, S., Frazier JA, Jacobsen LK, McKenna K, Gordon CT, Lenane MC, Hamburger
605 SD, Smith AK, Albus KE, Alaghband-Rad J, Rapoport JL. (1996): Childhood-
606 onset schizophrenia. A double-blind clozapine-haloperidol comparison. *Arch Gen*
607 *Psychiatry* 53(12):1090-1097.

608 Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, Zhang N, Mowry BJ, Olincy A,
609 Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW,
610 Kendler KS, Freedman R, Dudbridge F, Pe'er I, Hakonarson H, Bergen SE,
611 Fanous AH, Holmans PA, Gejman PV. (2011): Copy number variants in
612 schizophrenia: confirmation of five previous findings and new evidence for 3q29
613 microdeletions and VIPR2 duplications. *Am J Psychiatry* 168(3):302-16.

614 Li W, Wang X, Zhao J, Lin J, Song XQ, Yang Y, Jiang C, Xiao B, Yang G, Zhang HX,
615 Lv LX. (2011): Association study of myelin transcription factor 1-like
616 polymorphisms with schizophrenia in Han Chinese population. *Genes, Brain and*
617 *Behavior* 11(1):87-93.

618 Liaw A, Wiener M. (2002): Classification and Regression by randomForest. *R News*
619 2(3):18-22.

620 Lynall M-E, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, Bullmore
621 E. (2010): Functional Connectivity and Brain Networks in Schizophrenia.
622 *The Journal of Neuroscience* 30(28):9477-9487.

623 Malley J, Kruppa J, Dasgupta A, Malley K, Ziegler A. (2011a): Probability Machines.
624 Consistent Probability Estimation Using Nonparametric Learning Machines.
625 *Methods of information in medicine* 50(5):74-81.

626 Malley J, Malley K, Pajeviv S. (2011b): Statistical Learning for Biomedical Data New
627 York: Cambridge University Press.

- McKenna K, Gordon CT, Lenane M, Kaysen D, Fahey K, Rapoport JL. (1994): Looking for childhood-onset schizophrenia: the first 71 cases screened. *Journal of the American Academy of Child and Adolescent Psychiatry* 33(5):636-44.
- Meyer-Lindenberg AS, Olsen RK, Kohn PD, Brown T, Egan MF, Weinberger DR, Berman KF. (2005): Regionally Specific Disturbance of Dorsolateral Prefrontal-Hippocampal Functional Connectivity in Schizophrenia. *Arch Gen Psychiatry* 62(4):379-386.
- Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP, Pakula AT, Eisenhauer NJ, Uhas K, Weik L, Guy L, Care ME, Morel CF, Boni C, Salbert BA, Chandraredy A, Demmer LA, Chow EW, Surti U, Aradhya S, Pickering DL, Golden DM, Sanger WG, Aston E, Brothman AR, Gliem TJ, Thorland EC, Ackley T, Iyer R, Huang S, Barber JC, Crolla JA, Warren ST, Martin CL, Ledbetter DH. (2010): Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* 87(5):618-30.
- Rapoport JL, Addington AM, Frangou S. (2005): The neurodevelopmental model of schizophrenia: update 2005. *10(5):434-449.*
- Repovs G, Csernansky JG, Barch DM. (2011): Brain Network Connectivity in Individuals with Schizophrenia and Their Siblings. *Biological Psychiatry* 69(10):967-973.
- Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. (2004): A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22(3):1060-1075.
- Segonne F, Pacheco J, Fischl B. (2007): Geometrically Accurate Topology-Correction of Cortical Surfaces Using Nonseparating Loops. *Medical Imaging, IEEE Transactions on* 26(4):518-529.
- Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S. (1983): A Children's Global Assessment Scale (CGAS). *Arch Gen Psychiatry* 40(11):1228-1231.
- Stefansson H, Rujescu D, Cichon S, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Touloupoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeny LA, Franke B, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. (2008): Large recurrent microdeletions associated with schizophrenia. *Nature* 455(7210):232-6.
- Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DH, Corvin A, Craddock NJ, Gill M, Hultman CM, Lichtenstein P, McQuillin A, Pato CN, Ruderfer DM, Owen MJ, St Clair D, Sullivan PF, Sklar P, Purcell Leader SM, Korn J, Macgregor S, Morris DW, O'Dushlaine CT, Daly MJ, Visscher PM, Holmans PA, Purcell SM, Scolnick EM, Sklar Leader P, Williams NM, Georgieva L, Nikolov I, Norton N, Williams H, Toncheva D, Milanova V, Thelander EF, Sullivan P,

- Kenny E, Waddington JL, Choudhury K, Datta S, Pimm J, Thirumalai S, Puri V, Krasucki R, Lawrence J, Quedsted D, Bass N, Curtis D, Crombie C, Fraser G, Leh Kwan S, Walker N, Muir WJ, McGhee KA, Pickard B, Malloy P, Maclean AW, Van Beck M, Pato MT, Medeiros H, Middleton F, Carvalho C, Morley C, Fanous A, Conti D, Knowles JA, Paz Ferreira C, Macedo A, Helena Azevedo M, McCarroll SA, Daly M, Chambert K, Gates C, Gabriel SB, Mahon S, Ardlie K. (2008): Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*.
- Sun D, van Erp TGM, Thompson PM, Bearden CE, Daley M, Kushan L, Hardt ME, Nuechterlein KH, Toga AW, Cannon TD. (2009): Elucidating a Magnetic Resonance Imaging-Based Neuroanatomic Biomarker for Psychosis: Classification Analysis Using Probabilistic Brain Atlas and Machine Learning Algorithms. *Biological Psychiatry* 66(11):1055-1060.
- Weinberger D. (1987): Implications of Normal Brain Development for the Pathogenesis of Schizophrenia. *Arch Gen Psychiatry* 44(7):660-669.
- Yoon U, Lee J-M, Im K, Shin Y-W, Cho BH, Kim IY, Kwon JS, Kim SI. (2007): Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage* 34(4):1405-1415.

722 Appendix A: 15 item chart review for developmental issues with inclusion examples.
 723 Each item is scored 1 or 0.

724
 725

		Inclusion Examples
	ACADEMIC	
1	Delay	Skills ≥ 1 grade level behind; repeating a grade; learning disabilities
2	Special Education	Special needs school; resource room help
	SOCIAL	
3	Abnormal Peer Relations	Difficulty making or keeping friends; difficulty with reciprocal interaction
4	Withdrawal	Keeps to self; loner
5	Disinhibition	Aggression (physical and verbal); impulsivity
	SPEECH/LANGUAGE	
6	Rhythm	Speech/language evaluation results ≤ 1 standard deviation below the mean
7	Articulation	Difficulties pronouncing "R's" at age 7.
8	Comprehension	Speech/language evaluation results ≤ 1 standard deviation below the mean
9	Production	Speech/language evaluation results ≤ 1 standard deviation below the mean
10	Mutism	Total or selective
11	Delay	First words spoken after 18 months
	MOTOR	
12	Tics	Vocal and motor tics
13	Repetition	Rocking; flapping
14	Clumsiness	Poor coordination; difficulties skipping
15	Delay	Not crawling by 10 months

726

Figure 1.JPEG

Figure 1 Classification error histograms for (A) 1000 Random Forest runs using 74 cortical and subcortical regions to predict group membership for COS and control groups; (B) 1000 Random Forest runs using 74 cortical and subcortical regions to predict group membership after group membership was randomly permuted each run

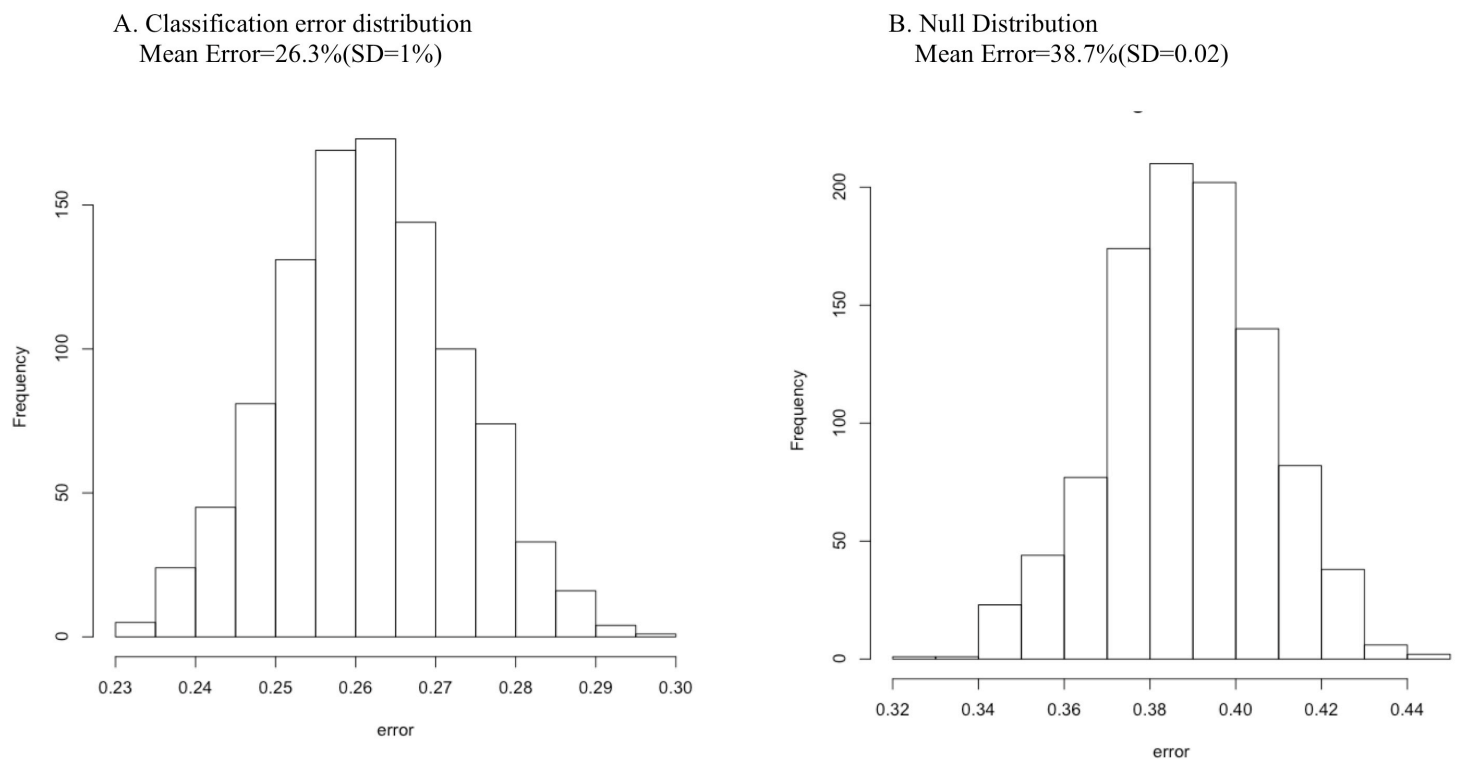
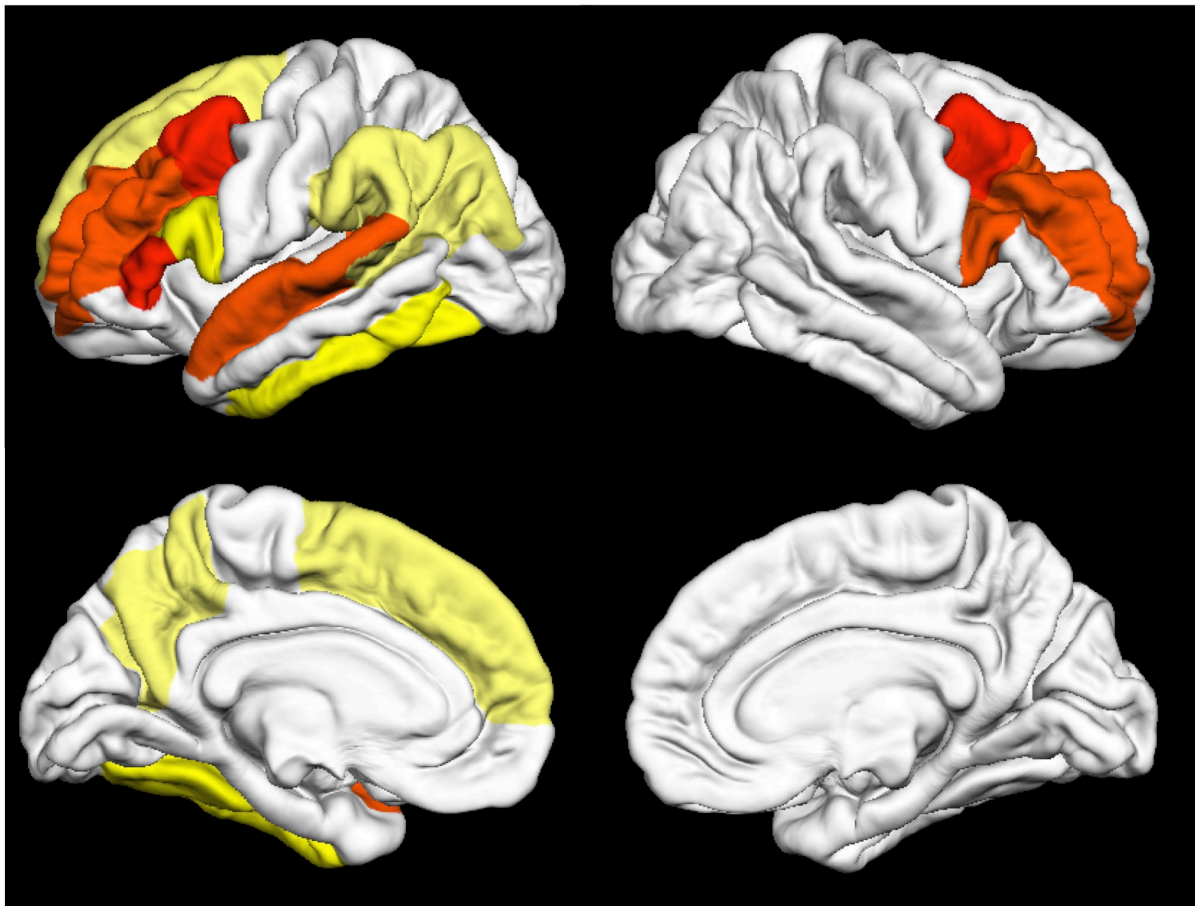


Figure 2.JPEG

Figure 2: Fifteen cortical regions with importance Z scores at least 0.5 standard deviations above the mean. Colors go from red (high Z scores) to light yellow (lower Z scores)*



* Red: $Z > 1.5$; Orange: $Z = 1 - 1.5$; Yellow: $Z = 0.8 - 1.0$; Light yellow: $Z = 0.5 - 0.75$

Figure 3.JPEG

Figure 3: Proximity values averaged over 1000 Random Forest runs for all participants (represented by the dots) visualized with two dimensional multidimensional scaling (MDS)

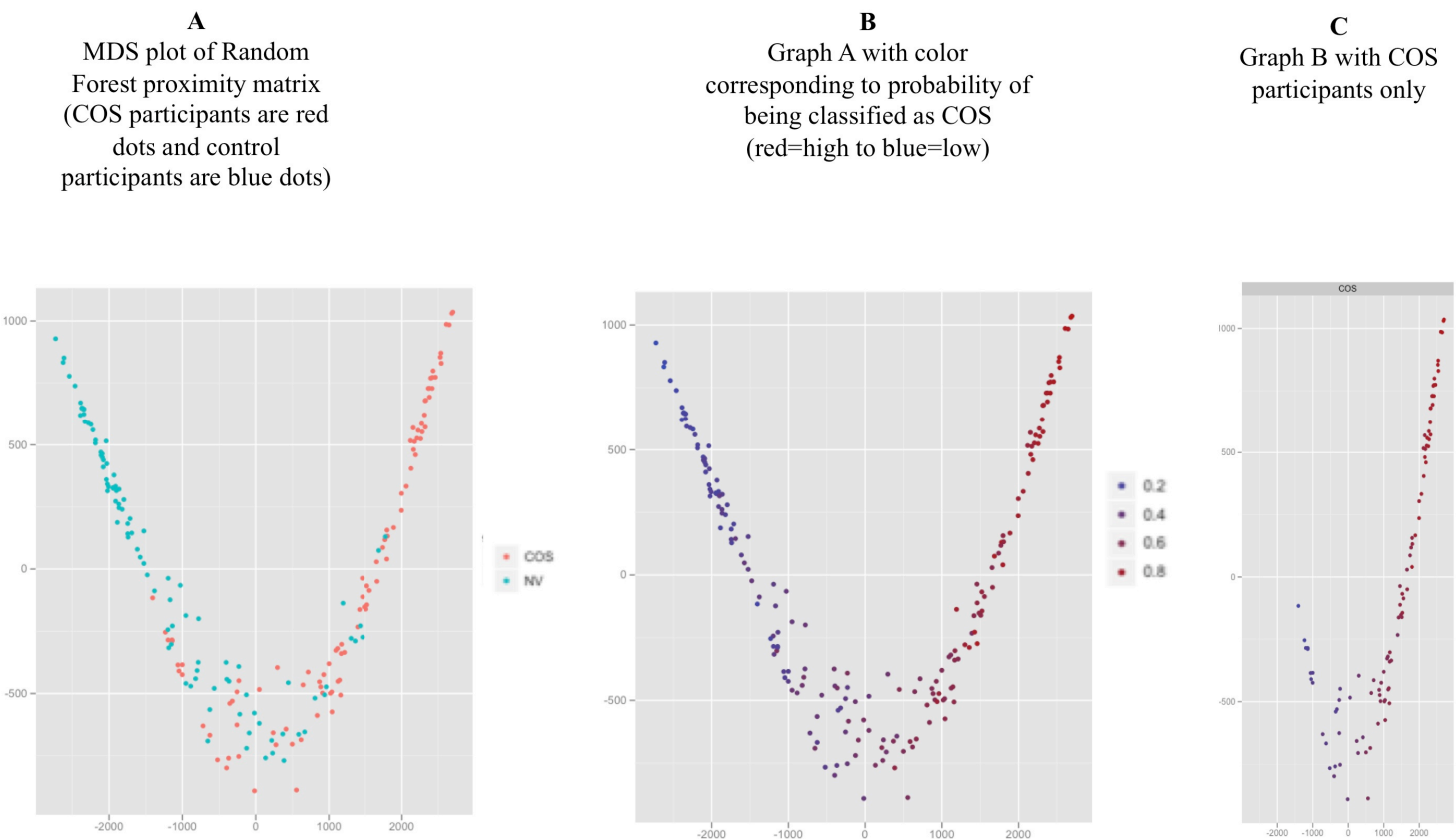


Figure 4.JPEG

Figure 4 The relationship between univariate logistic regression coefficients and Random Forest importance Z scores (importance score-mean importance score/ SD importance scores) (loess line in blue with 95% confidence interval).

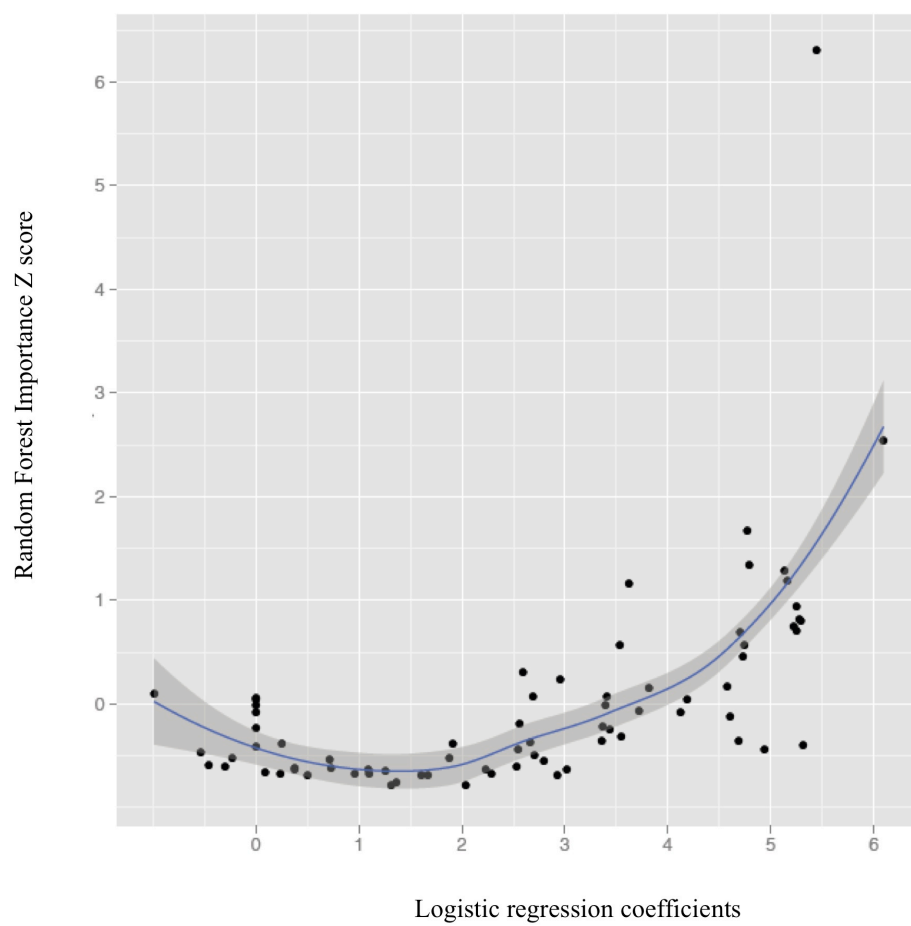


Figure 5.JPEG

Figure 5: Scatter plots for probability of being classified as COS using structural MRI-based Random Forest classifier (X axis) and A) medication-free inpatient Global Assessment of Functioning Scale score (GAS) ($p=0.0004$) ; B) number of premorbid academic, language, social and motor difficulties ($p=0.02$) with fitted regression line and 95% confidence interval.

