**Supplementary Material - Table of Contents**

**Supplementary Methods**

1. Neuroimaging Study Inclusion Criteria
2. Clinical Assessment, Neuropsychological and Demographic Data
3. Image Processing Details
4. Final Sample & Quality control procedures
5. Choosing the best model for vertex-wise WGC and CT data.

**Supplementary Results**

1. Cross-sectional assessment of verbal memory and negative symptoms.
2. Relationship between change in verbal memory and change in negative symptoms.

**Supplementary Figures**

**Supplementary Figure 1.** Effect of cumulative chlorpromazine equivalents on WGC and CT rate of change (t-statistic maps).

**Supplementary Figure 2.** Spaghetti plots of negative symptoms over time.

**Supplementary Figure 3.** Results from main manuscript without covarying for mean cortical thickness or mean WGC.

**Supplementary Figure 3a.** Change in Negative Symptoms

**Supplementary Figure 3b.** Baseline Verbal Memory \* Change in Negative Symptoms

**Supplementary Figure 4**. Covarying for General Cognitive Ability in analysis of Baseline Verbal Memory \* Change in Negative Symptoms.

**Supplementary Tables**

**Supplementary Table 1**. Demographic/Cognitive/Clinical Characteristics comparing included and excluded patients.

**Supplementary Table 2**. Comparing negative symptoms and verbal memory performance between diagnoses.

**Supplementary Table 3**. Baseline verbal memory performance characteristics by neuropsychological test battery and by group.

**Supplementary Table 4.** Demographic/Cognitive/Clinical Characteristics of full cross-sectional sample with verbal memory data.

**Supplementary References**

**Supplementary Methods**

**i. Neuroimaging Study Inclusion Criteria**

For the neuroimaging portion of the study, additional inclusion criteria for first episode of psychosis (FEP) patients comprised: minimum age of 18, clinically stable status and no major medical disorders. Exclusion criteria included a history of neurological illnesses and head trauma resulting in loss of consciousness that could affect cognition, presence of neurological disorder determined by medical record examination, lifetime diagnosis of substance dependence, and/or any potential contraindication for the MR scan. In addition to exclusion criteria listed for FEP patients, controls were excluded if they had any current/past history of Axis I disorders, and/or a first-degree family member suffering from schizophrenia or related schizophrenia spectrum psychosis.

**ii. Clinical Assessment, Neuropsychological and Demographic Data***.*

Diagnosis for each patient was made on the basis of a structured clinical interview for DSM-IV (First *et al.* 1998), performed by a trained interviewer, and confirmed by at least one senior psychiatrist (RJ or AM). Depression and anxiety symptoms were assessed with the Calgary Depression Scale for Schizophrenia (CDSS) (Addington *et al.* 1990) and Hamilton Anxiety Rating Scale (HARS) (Hamilton 1959), respectively. Positive and negative symptoms were assessed with the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen 1984a) and Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen 1984b). Duration of untreated psychosis (DUP) was also assessed, referring to the period of time in weeks between onset of psychotic symptoms to adequate treatment with antipsychotics, as described elsewhere (Bodnar *et al.* 2016). Antipsychotic medication dosages were converted to chlorpromazine equivalents according to the literature (Leucht *et al.* 2014), and multiplied by percent medication adherence (Cassidy *et al.* 2010). For both controls and patients, parental socioeconomic status (SES) was estimated using the Hollingshead SES Rating Scale (Hollingshead 1965), and handedness determined with the Edinburgh Handedness Inventory (Oldfield 1971). Due to a change in the neuropsychological test battery used mid-way through the study, Full Scale IQ was assessed with the Weschler Adult Intelligence Scale (WAIS-III) for a proportion of subjects, and the Weschler Abbreviated Scale of Intelligence (WASI) for the remaining sample (Weschler 1997, 1999).

**iii. Image Processing Details**

**Additional details on White-Gray Contrast extraction.** As described in the manuscript, T1-weighted images were submitted to the CIVET pipeline (Version 2.1.0: <http://www.bic.mni.mcgill.ca/ServicesSoftware/CIVET)> (June *et al.* 2005) for extraction of gray and white matter surfaces. WGC was estimated using a recently published method from our group (Lewis *et al.* 2018). In detail, a distance map relative to the white surface provided by CIVET was created at 0.25mm3 resolution, smoothed with a 0.5mm FWHM Gaussian kernel, and used to create a gradient vector field of the distance map. A copy of the white surface was moved 1mm inward along this gradient vector field to produce a sub-white surface, and 1mm outward to produce a supra-white surface. The intensity values on the T1-weighted image (without non-uniformity correction or normalization) were sampled at each vertex of both the supra-white surface and the sub-white surface, and the resulting surface maps were smoothed with a 20mm FWHM Gaussian kernel in standardized space in order to diminish the impact of noise. The WGC measures were then formed by dividing the value at each vertex of the sub-white surface by the value at the corresponding vertex of the supra-white surface.

**iv. Final Sample & Quality Control Procedures**

**Neuroimaging Sample Details.** The neuroimaging study began in 2003 and spanned over a decade, comprising three scheduled visits: baseline, one-year follow-up (FUP1), and two-year follow-up (FUP2). All scans were acquired on the same 1.5T scanner. In total 150 patients and 95 controls were recruited for the study. From this sample, 142 patients and 94 controls completed a baseline scan. At this point, 7 patients and 4 controls were excluded from analysis based on the exclusion criteria for the study (e.g., incidental findings, substance-induced psychosis, low IQ). Forty-two patients and forty-five controls dropped out of the study after baseline, leaving 100 patients and 47 healthy controls with longitudinal data. In the FEP group, baseline scans were performed on average 4.1 (SD=1.9) months from entry into PEPP. For the entire group, including controls, inter-scan intervals were approximately 13.1 (SD=1.3) months between baseline and FUP1, and 12.5 (SD=1.7) months between FUP1 and FUP2. Nine participants (6 FEP, 3 controls) were not scanned at FUP1, and had an average interscan interval of 27.0 (SD=3.2) months between FUP2 and baseline. Sex ratios did not differ between the included cross-sectional sample (70% Male in FEP group, 65% Male in HC group) and the longitudinal sample (71% Male in FEP group, 65% Male in HC group), thus participant attrition did not affect the proportion of males:females in this study.

**Quality Control Procedures.** One rater (C.M.) manually rated all of the CIVET outputs, using a quality control module built in-house, where snapshots of the mask and surfaces overlaid on the T1-weighted image, gray and white matter surface renderings, surface-surface intersections, and surface extraction convergence plots were created. These features were examined to ensure that a) the brain was extracted accurately, b) the surfaces generally followed the anatomy, c) no visible abnormalities (i.e. visibly jagged surfaces, or surfaces jutting out inaccurately) in gray and white matter surfaces could be detected, and d) complete convergence of the surface extraction process. For any questionable areas or outputs, the surfaces were overlaid on the T1 image and viewed slice by slice using Display software (<https://mcin-cnim.ca/technology/visualization/display/)>. A three-point rating system was adopted: 2 for good quality surfaces (minimal error), 1 for satisfactory surfaces (mild to moderate error), and 0 for failed outputs. For the “questionable” outputs that were rated as a “1”, a second rater independently rated the outputs, and a consensus was reached as to the inclusion/exclusion of the scan.

Three control scans (all follow-up scans) and 7 FEP scans (5 baseline and 2 follow-up), all belonging to unique subjects, failed the CIVET processing pipeline. From that point, 3 control scans (belonging to two participants) and 25 FEP scans (belonging to 17 patients) did not pass QC after inspection of outputs. 116 FEP patients remained after this step. From there, FEP patients with cross-sectional data only were removed for the analysis. The final sample with usable longitudinal data comprised 88 FEP patients, and a sample of 80 controls that had verbal memory data available, in addition to passing QC.

**v. Choosing the best model for vertex-wise CT and WGC data.**

To determine which covariates should be used for analyses, the Akaike Information Criterion (AIC) (Akaike 1998) was used to determine the best model for the data at hand. The method uses a likelihood function to reward goodness of fit, while including a penalty that is an increasing function of the number of estimated parameters. The baseline model was a general linear model, which included parameters for group, centered age, and sex, as follows:

Y=intercept + β1(Centered Age) +β2(Sex) + ε

where Y represents vertex-wise *rate of change* in WGC or CT, β1-2 represent regression coefficients and ε is residual error. From here, the model was tested with the addition of a third covariate, β3, from the following variables: handedness, years of education, total brain volume (TBV), and mean WGC (for WGC data) and mean CT (for CT data). The best fit for the WGC data (with the smallest AIC value) included mean WGC as a covariate, in addition to centered age and sex. A similar model was selected for the CT data, including mean CT as a covariate. It should be noted that TBV also modestly improved the AIC fit, but did not reduce the AIC value to the same extent and as consistently as mean CT across all models. For consistency, only results covarying for mean WGC and mean CT and not TBV are presented in both manuscripts. Results remain largely the same when removing mean WGC and mean CT as a covariate. See **Supplementary Figures 2a and 2b** below.

Given that both gray and white matter have been found to be affected by cumulative dosage of antipsychotic medication in patients with psychosis (Bartzokis *et al.* 2009; Roiz-Santiáñez *et al.* 2012; Vita *et al.* 2015), a separate analysis was included to observe potential effects of antipsychotic medication on the calculated rate of change of WGC and CT. Antipsychotic medication was defined as cumulative chlorpromazine equivalents, multiplied by adherence, as measured at the end of the neuroimaging study (i.e. this measure was taken from each patients’ last scanning timepoint, and is a cumulative measure of antipsychotic medication exposure measured monthly at PEPP-Montreal). Antipsychotic medication was regressed against Y (i.e. change in CT or WGC), controlling for age, sex, and Mean WGC or Mean CT (denoted as Mean(σ) in the main manuscript). No significant effects of antipsychotic medication on rate of change in WGC or CT were observed (see **Supplementary Figure 1**), and this variable did not lower the AIC score when added to the above-described models. Thus we felt justified in leaving out medication as a covariate.

**Supplementary Results**

**i. Cross-sectional assessment of verbal memory and negative symptoms.**

**Verbal Memory at Baseline.** As reported in **Table 1** of the main manuscript, FEP patients had significantly lower VM performance (immediate and delayed recall) compared to healthy controls, when controlling for neuropsychological test battery version (F1,163>16.77, p<0.0001). This held true for the full cross-sectional sample as well, included in **Supplementary Table 4**. Descriptive statistics of VM performance by group and version, and other relevant statistics, are included in **Supplementary Table 3**. No significant group by version interaction term or main effect of version was found on VM performance. Also, no significant difference was observed within the FEP group comparing pen and paper vs. computerized test battery data, confirming that the data from the two different test batteries are comparable, and thus could be pooled in our analysis.

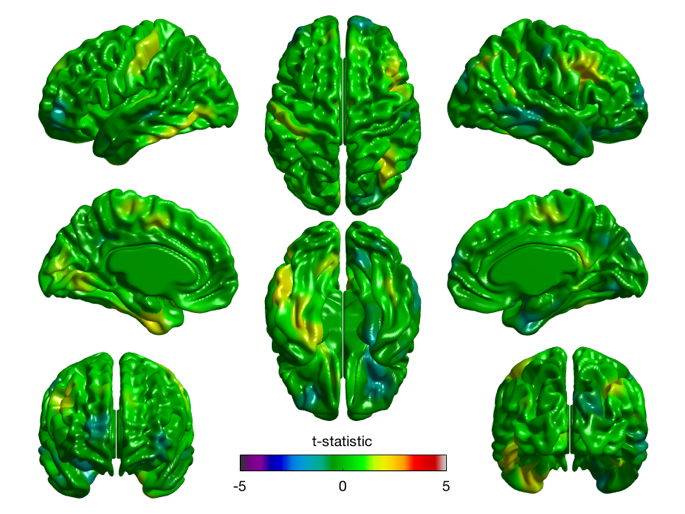
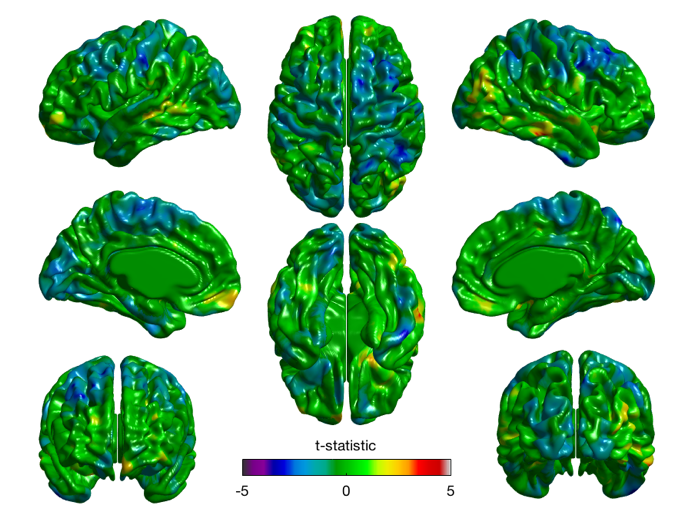
**Verbal Memory in Relation to Negative Symptom Domains at Baseline***.* The relationship between VM (i.e. z-scores of immediate and delayed recall) and negative symptom domains (i.e. log-transformed scores for Amotivation and Expressivity) at baseline were assessed in 115 FEP patients with Pearson r-correlations, adjusting verbal memory scores for age, sex, and test battery version. Expressivity was found to be significantly negatively associated with both immediate (r=-0.35, q<0.001) and delayed recall (r=-0.28, q=0.036). Only a trend-like significant negative association was noted between Amotivation and immediate recall (r=-0.17, p-uncorrected=0.067), and no relationship was found between Amotivation and delayed recall (r=-0.12, q=0.215).

**ii. Relationship between Change in Verbal Memory and Change in Negative Symptoms**.

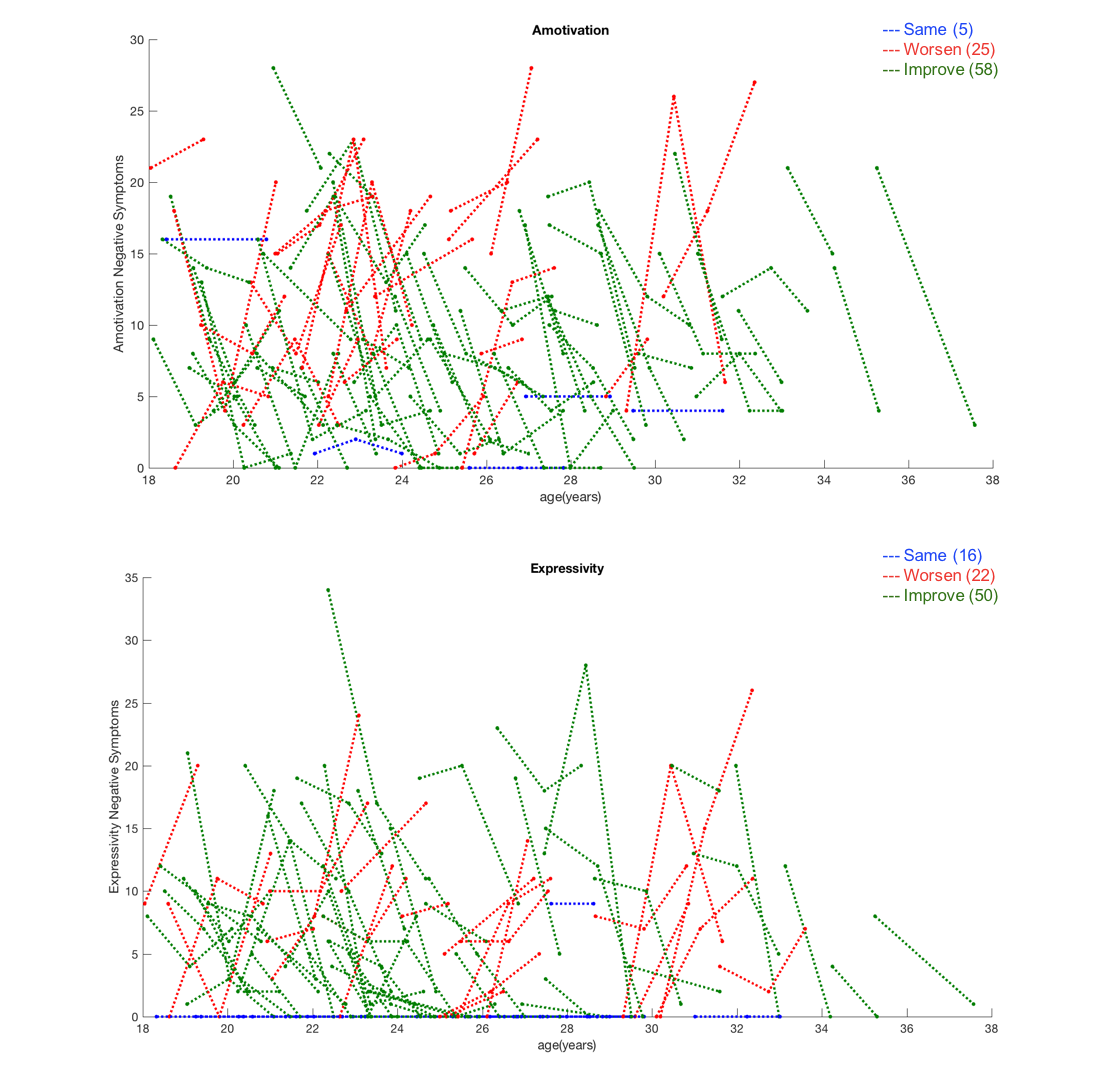
A subset of patients included in the study (N=49 from 88 patients) had verbal memory assessed at one-year follow-up. Thus, we conducted an exploratory analysis to see if there was a relationship between change in verbal memory and change in negative symptoms over this one-year period, using Pearson r correlations, adjusting verbal memory scores by age, sex, and test battery version. For the included 49 FEP patients, we found no relationship between  immediate recall and amotivation (r=-0.096, p=0.51), nor immediate recall and expressivity (r=-0.085, p=0.55). Similarly for  delayed recall, we found no relationship with amotivation (r=-0.19, p=0.18) or expressivity (r=-0.13, p=0.39). Note, a negative r value denotes that a worsening of negative symptoms is associated with a worsening of verbal memory performance (given that a negative change value reflects improvement in negative symptoms but worsening of verbal memory).

**Supplementary Figures**

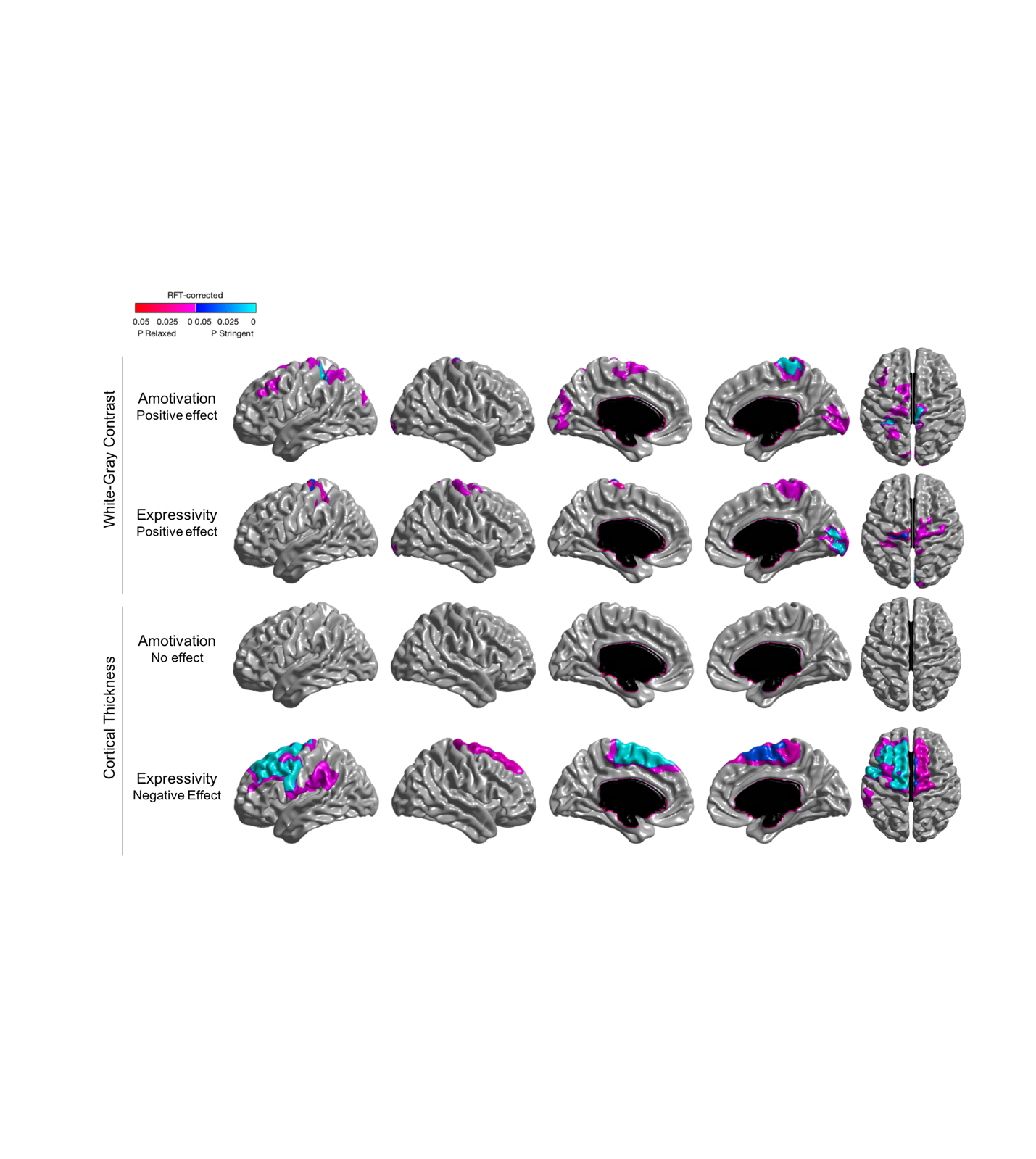
**i. White-Gray Contrast ii. Cortical Thickness**



**Supplementary Figure 1**. T-statistic maps (df=83) showing main effect of antipsychotic medication on change in i) WGC and ii) CT, controlling for centered age, sex, and mean WGC for i), and mean CT for ii). Antipsychotic medication is defined as cumulative chlorpromazine equivalents, as measured at each patients’ last scanning timepoint, multiplied by adherence. A positive t-statistic (warm colours) reflects a positive effect of antipsychotic medication on brain structure, and a negative t-statistic (cool colours) reflects a negative effect of antipsychotic medication. Nothing was found to be significant after RFT correction. It should be noted that there was also no significant relationships between mean baseline σ and chlorpromazine equivalents at baseline (p’s>0.30), or between mean Y and cumulative chlorpromazine equivalents (p’s>0.16).

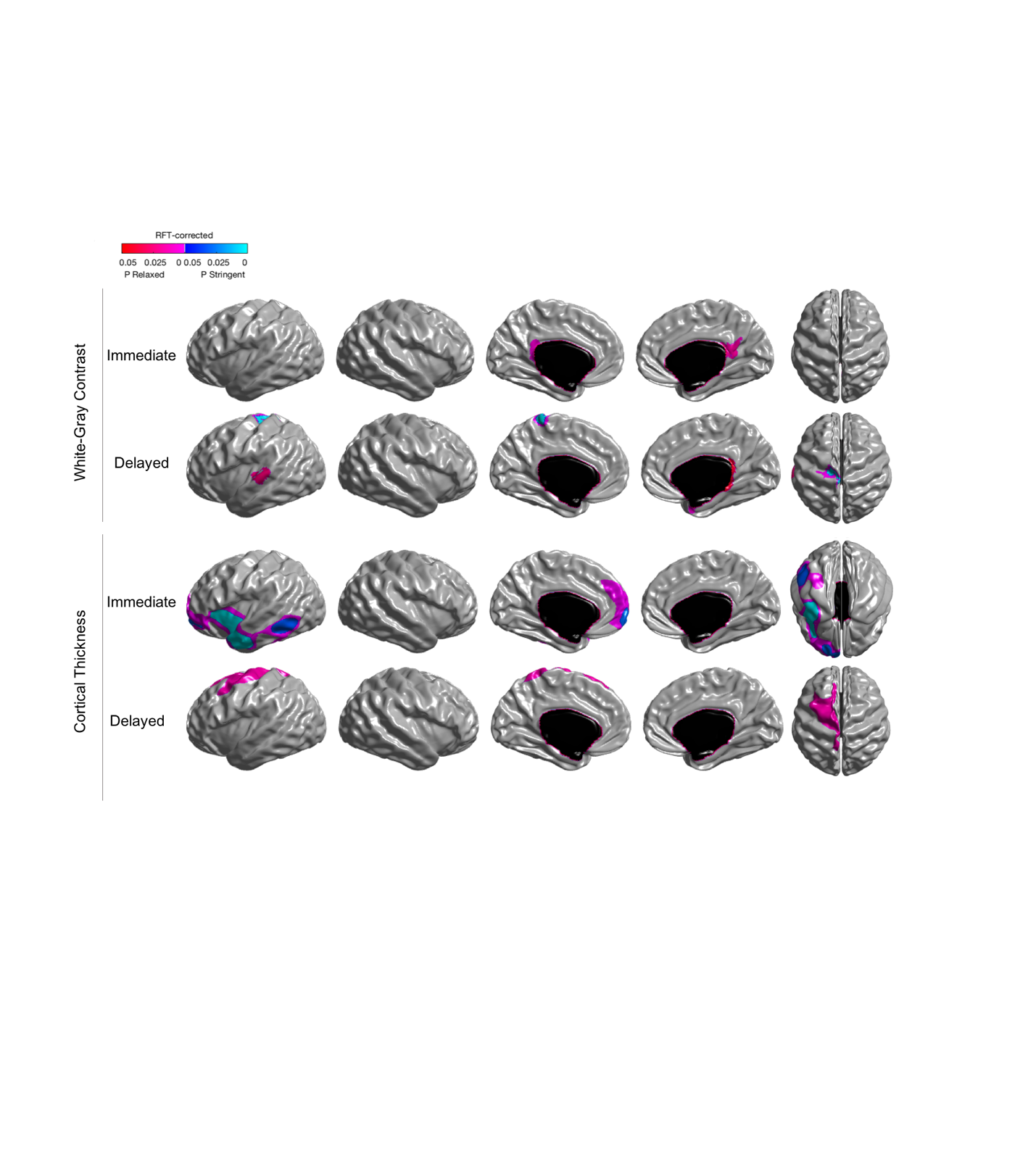


**Supplementary Figure 2. Spaghetti plots of negative symptoms.** Plots show change over time in Amotivation (top panel) and Expressivity Symptoms (bottom panel) in FEP patients. Each line joins timepoints from the same subject, and is colour-coded based on the slope of change, such that blue represents no change in symptoms (slope = 0), green represents improvement of symptoms (slope < 0) and red represents worsening of symptoms (slope > 0). The proportion of patients comprising each category are included in brackets in the legend. For amotivation symptoms, 5 patients had the same level (or absence) of symptoms over time (Mean SD=0.0086), 25 patients had worsening symptoms (Mean SD=2.91) and 58 patients improved (Mean SD. For expressivity symptoms, 16 patients had the same level (or absence) of symptoms over time (Mean SD=0), 22 patients had worsening symptoms (Mean SD=4.02), and 50 patients improved (Mean SD=4.01).



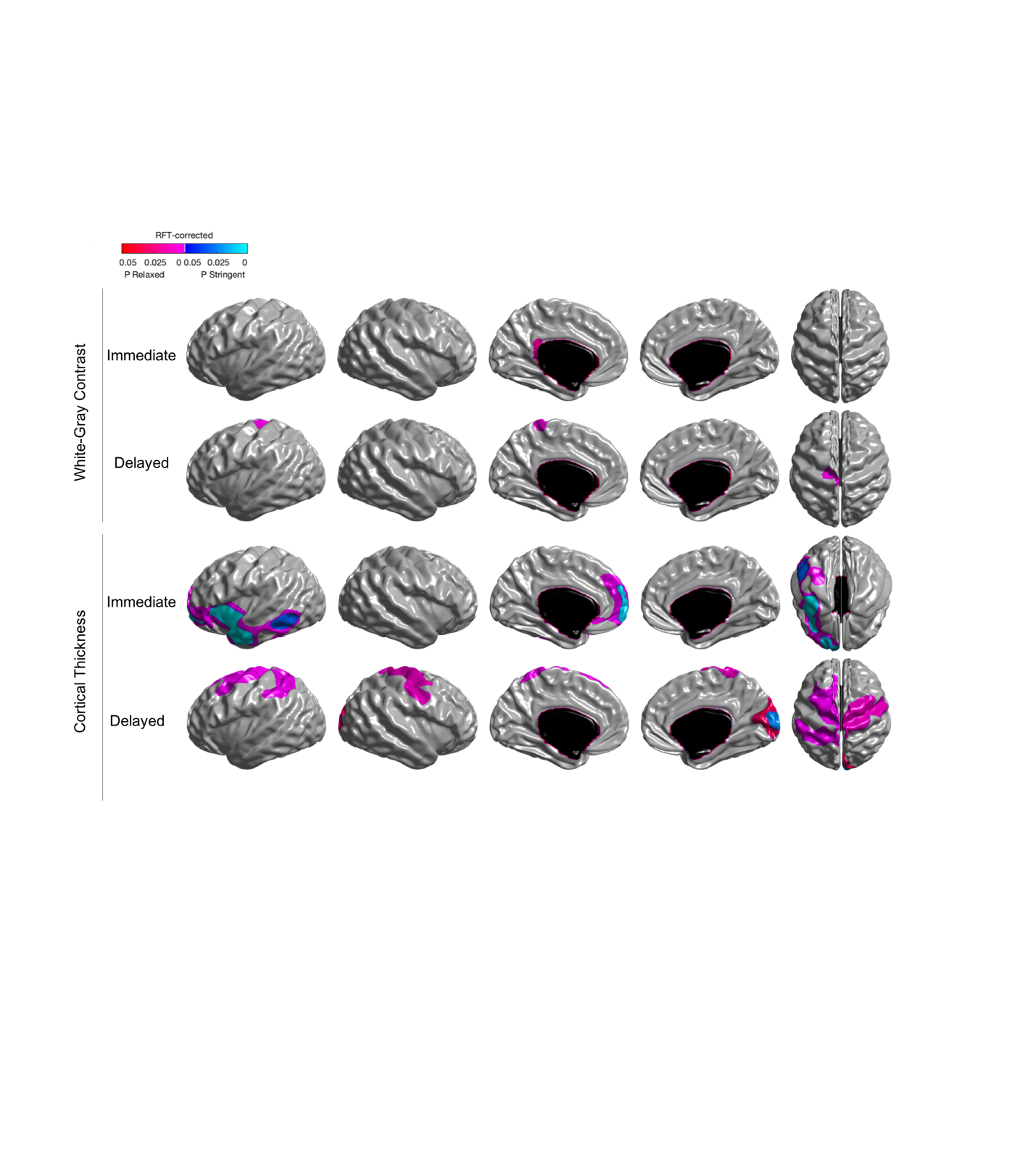
**Supplementary Figure 3a**. Change in Amotivation and Expressivity deficits associated with WGC and CT across time, ***excluding mean(σ) as a covariate***. Results are RFT-corrected, where blue colours represent significant results cluster-thresholded at a “stringent” threshold of p=0.001, whereas red/pink colour represent significant results at a “relaxed” threshold of p=0.01. Results remain relatively unchanged from those presented in the main manuscript.

Orientation: Surfaces from left to right in each row: left lateral, right lateral, left medial, right medial, dorsal.



**Supplementary Figure 3b**. Interaction between baseline verbal memory and change in Expressivity on WGC and CT, ***excluding mean(σ) as a covariate***. “Immediate” and “Delayed” labels in left-hand side panel refer to immediate and delayed recall of verbal memory domain, respectively. Inflated brain is presented to better visualize results within cortical folds. Results are RFT-corrected, where blue colours represent significant results cluster-thresholded at “original” stringent threshold of p=0.001, whereas red/pink colour represent significant results at a “relaxed” threshold of p=0.01. Results remain relatively unchanged from those presented in the main manuscript, with the largest change reflected in a weaker extent of results for the interaction between delayed recall and change in expressivity on changes in CT.

Orientation: Surfaces from left to right in each row: left lateral, right lateral, left medial, right medial, dorsal (with the exception of the third row, where the right-most surface is a ventral view).



**Supplementary Figure 4**. Interaction between baseline verbal memory and change in Expressivity on WGC and CT, ***covarying for general cognitive ability***, in addition to covariates presented in main manuscript. Z-scores of general cognitive ability comparing FEP patients against healthy controls were calculated based on five domains of cognition: speed of processing, attention, executive function, visual memory, and executive function. A composite cognitive index was calculated by averaging the scores obtained for the five cognitive domains. Similar to the verbal memory scores, cognition was assessed with two different batteries (i.e. pen and paper, and the computerized CogState battery). Additional information on the tests comprising each cognitive domain for each battery can be found in Benoit et al (2015). Within the figure, “Immediate” and “Delayed” labels in left-hand side panel refer to immediate and delayed recall of verbal memory domain, respectively. Inflated brain is presented to better visualize results within cortical folds. Results are RFT-corrected, where blue colours represent significant results cluster-thresholded at “original” stringent threshold of p=0.001, whereas red/pink colour represent significant results at a “relaxed” threshold of p=0.01. Results remain relatively unchanged from those presented in the main manuscript, suggesting that results are specific to verbal memory, and cannot be explained by general cognitive ability.

Orientation: Surfaces from left to right in each row: left lateral, right lateral, left medial, right medial, dorsal (with the exception of the third row, where the right-most surface is a ventral view).

**Supplementary Tables**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Included Patients (N=88) | Excluded Patients (N=47) | Statistics Comparing Groups | | |
|  | **N (%)** | | **Statistic** | **df** | **p-value** |
| Male | 62 (70) | 35 (74) | χ2=0.2 | 1 | 0.6 |
| Right Handed | 73 (83) | 41 (87) | χ2=0.4 | 1 | 0.5 |
| *Diagnosisa* | |  |  |  |  |
| Schizophrenia/ Schizophreniform | 63 (72) | 28 (60) |  |  |  |
| Affective Disorder | 16 (18) | 12 (26) |  |  |  |
| Delusional Disorder | 3 (3) | 0 (0) |  |  |  |
| Psychosis Not Otherwise Specified | 6 (7) | 6 (13) |  |  |  |
|  | **Mean (SD)** | | **Statistic** | **df** | **p-value** |
| Age | 24.3 (4.1) | 23.7 (4.0) | U=1893.5 |  | 0.4 |
| Verbal Memory – Immediateb | -1.3 (1.4) [87] | -1.3 (1.3) | F=0.1 | 1,130 | 0.8 |
| Verbal Memory – Delayedb | -0.9 (1.1) [87] | -0.9 (1.0) | F=0.02 | 1,130 | 0.9 |
| Education in Years | 12.0 (2.5) | 11.8 (2.6) | t=0.4 | 133 | 0.67 |
| Socioeconomic Status | 3.2 (1.1) [81] | 3.2 (1.1) [41] | U=1588.5 |  | 0.7 |
| Performance IQb | 98.7 (17.1) | 94.3 (14.2) | F=1.2 | 1,131 | 0.3 |
| Verbal IQb | 99.8 (16.1) | 97.7 (14.7) | F=0.6 | 1,131 | 0.4 |
| CPZ equivalent (in mg) with adherence | 747.9 (720.8) | 818.4 (911.4) | U=2021.0 |  | 0.8 |
| **Duration Untreated Psychosis (weeks)\*** | 76.4 (142.0) | 36.9 (64.1) [44] | U=1423.5 |  | 0.05 |
| **Duration Untreated Illness (years)\*** | 7.3 (6.5) | 4.4 (4.1) [44] | U=1346.0 |  | 0.012 |
| Amotivation | 11.3 (6.1) | 11.4 (7.8) | t=-0.09 | 133 | 0.9 |
| Expressivity | 7.1 (7.2) | 7.4 (7.5) | U=2023.0 |  | 0.8 |

**Supplementary Table 1.** Comparison of demographics for patients Included in the main manuscript (N=88) and patients excluded from the main manuscript (N=47), due to presence of cross-sectional data only, and exclusion on the basis of failed image quality control (QC). All data represented as Mean (SD), unless otherwise specified. Square brackets [] include adjusted sample size included in statistical analysis due to missing datapoints. A Mann Whitney U non-parametric test was used for variables that were not normally distributed.

\***Bolded variables with asterisk** reflect significant differences between Included and excluded patients. Specifically, included patients had longer duration of untreated psychosis and untreated illness compared to excluded patients. The two patient groups did not differ on any other variable. Age, socioeconomic status, chlorpromazine (CPZ) equivalents, duration of untreated psychosis/illness, and expressivity data were non-normally distributed, thus non-parametric tests were used to assess group differences.

aDiagnosis missing for one excluded patient.

bCovaried by test version. Specifically, IQ was collected with WAIS-III (Weschler 1997) and WASI (Weschler 1999) and verbal memory was collected with a Pen and Paper neuropsychological test battery, and CogState Research Battery (Pietrzak *et al.* 2009).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SSZ (N=64) | AFF (N=15) | Statistics Comparing Groups | | |
|  | **Mean (SD)** | | **Statistic** | **df** | **p-value** |
| Verbal Memory - Immediatea,b | -1.5 (1.3) [63] | -0.9 (1.3) | F=2.6 | 1, 74 | 0.1 |
| Verbal Memory - Delayeda,b | -1.0 (0.9) [63] | -0.5 (1.2) | F=3.2 | 1, 74 | 0.08 |
| Baseline Amotivation | 11.8 (6.1) | 9.5 (5.6) | t=1.3 | 77 | 0.2 |
| Baseline Expressivityc | 8.8 (8.2) | 4.4 (4.5) | U=335 |  | 0.07 |
| Δ Amotivation | -1.6 (4.6) | -2.3 (3.2) | t=0.6 | 77 | 0.6 |
| Δ Expressivity | -1.4 (5.9) | -1.0 (2.5) | t=-0.3 | 77 | 0.78 |

**Supplementary Table 2**. Comparing negative symptoms and verbal memory performance between FEP patients with a schizophrenia spectrum (SSZ) diagnosis (e.g. schizophrenia, schizoaffective) and patients with an affective (AFF) diagnosis (e.g. bipolar, major depression with psychotic features). No significant differences emerged between diagnoses, although there was a trend-like difference for delaed recall and baseline expressivity scores, where SSZ patients tended to have worse delayed recall performance and a higher level of expressivity deficits.

aVerbal memory missing for one SSZ patient. Adjusted sample size is included in square brackets.

bCovaried by test version (Pen and Paper neuropsychological test battery vs. and CogState Research Battery (Pietrzak *et al.* 2009)).

cNon-normally distributed, thus a Mann White U non-parametric test was used to test for group differences.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Battery Version** | **Verbal Memory Metric** | **FEP** | | **Control** | |
| **N** | **Mean (SD)** | **N** | **Mean (SD)** |
| Pen and Paper | Immediate | 76 | -1.2 (1.2) | 65 | 0.04 (1.0) |
| Delayed | -0.8 (0.9) | 0.02 (0.8) |
| CogState | Immediate | 39 | -1.4 (1.7) | 15 | 0.0 (1.0) |
| Delayed | -1.0 (1.3) | 0.0 (1.0) |

**Supplementary Table 3**. Baseline verbal memory performance characteristics by neuropsychological test battery and by group, using larger cross-sectional sample of 116 patients that passed quality control and 80 controls (see **Supplementary Table 3** for demographic/clinical characterization of this sample). One-way ANOVAs were used to test an interaction effect between group and test version on verbal memory performance. There were no significant interaction effects of group by test version on either immediate (F1,191=0.179, p=0.67) or delayed (F1,191=0.50, p=0.48) recall. Further, there was no significant main effect of version on either immediate (F1,191=0.38, p=0.54) or delayed (F1,191=0.72, p=0.40) recall. As reported in the main manuscript, there is a significant main effect of group for both verbal memory metrics (F1,191>28.71, p<0.001). Note, the F-statistics reported in the main manuscript are slightly different as the main effect of group is controlling for test version, and only included FEP patients with longitudinal data included in the main manuscript. Finally, independent t-tests found no significant differences in verbal memory performance within the FEP group when comparing data from pen and paper vs. computerized test battery versions (t58.9=0.75, p=0.46 and t57.9=1.15, p=0.25 for immediate and delayed recall, respectively; degrees of freedom adjusted for unequal variance, as assessed by Levene’s test for Equality of Variances).

Abbreviations: FEP, first episode of psychosis. SD, standard deviation.

|  |  |  |
| --- | --- | --- |
|  | **FEP N=115** | **Controls N=80** |
|  |
|  |
|  | **N (%)** | |
| Male | 82 (71) | 52 (65) |
| Right Handed | 95 (83) | 73 (91) |
| *Diagnosisa* | |  |
| Schizophrenia/ Schizophreniform | 79 (69) |  |
| Affective Disorder | 22 (19) |  |
| Delusional Disorder | 3 (3) |  |
| Psychosis Not Otherwise Specified | 10 (9) |  |
|  | **Mean (SD)** | |
| Age at Baseline | 24.1 (3.9) | 24.3 (3.3) |
| **Education in Years\*** | 11.8 (2.6) | 14.3 (2.4) |
| Socioeconomic Status | 3.2 (1.1) [105] | 3.0 (1.0) [76] |
| **Performance IQ\***b | 98.4 (16.4) | 106.9 (12.7) |
| **Verbal IQ\***b | 99.8 (15.8) | 110.0 (14.9) |
| ***Cognitionb,c*** | | |
| **Verbal Memory - Immediate\*** | -1.3 (1.4) | 0.03 (1.0) |
| **Verbal Memory - Delayed\*** | -0.8 (1.0) | 0.02 (0.8) |
| **Attention\*** | -0.8 (1.2) | 0.06 (1.0) |
| **Executive Function\*** | -0.7 (1.1) | 0.05 (0.7) |
| Speed of Processing | -0.4 (1.1) | 0.03 (0.8) |
| **Working Memory\*** | -0.6 (0.9) | -.01 (0.8) |
| **Visual Memory\*** | -0.7 (1.2) | 0.09 (0.7) |
| **General Cognitive Index\*** | -0.6 (0.8) | 0.04 (0.6) |
| ***Clinical Information*** | | |
| Cumulative CPZ equivalent (in mg) | 803.5 (737.6) |  |
| Adherence (%) | 83.0 (27) |  |
| Duration Untreated Psychosis (weeks) | 70.9 (130.8) [106] |  |
| Duration Untreated Illness (years) | 7.0 (6.1) [109] |  |
| Amotivation | 11.4 (6.7) |  |
| Expressivity | 7.1 (7.3) |  |
| SAPS | 10.0 (12.5) |  |
| CDSS | 2.4 (3.1) |  |
| Window |Scan - Symptom Eval| (months) | 0.7 (0.5) |  |

**Supplementary Table 4.** General Demographics for *cross-sectional sample* with verbal memory data collected at baseline. This sample was used for cross-sectional analyses of the relationship between verbal memory and negative symptom dimensions, presented here in Supplementary Material. All data represented as Mean (SD), unless otherwise specified. Square brackets [] include adjusted sample size included in statistical analysis due to missing datapoints. Immediate and delayed verbal memory scores were obtained as described in the methods of the main manuscript. Five other cognitive domains are presented here, and merged to form a “general cognitive index”, which was used as a covariate in **Supplementary Figure 3**. Z-scores of general cognitive ability comparing FEP patients against healthy controls were calculated based on five domains of cognition: attention, executive function, speed of processing, visual memory, and working memory. A composite cognitive index was calculated by averaging the scores obtained for the five cognitive domains. Similar to the verbal memory scores, cognition was assessed with two different batteries (i.e. pen and paper, and the computerized CogState battery). Additional information on the tests comprising each cognitive domain for each battery can be found in Benoit et al (2015). All antipsychotic totals are presented as chlorpromazine equivalents in mg, as prescribed by a psychiatrist, and are reported along with a percentage of medication adherence. SAPS totals are presented as mean scores of the sum of item-level scores.

\***Bolded variables with asterisk** reflect significant differences between FEP patients and controls. Similar to the longitudinal sample presented in the main manuscript, FEP patients had significantly lower verbal memory performance, IQ, and levels of education compared to HC (p<0.05). FEP patients had significantly lower performance on all cognitive subdomains and the general cognitive index (p<0.001), with the exception of speed of processing, which showed a trend-like group difference (p=0.06).

aOne patient did not have enough clinical data to reliably define a diagnosis.

bCovaried by test version. Specifically, IQ was collected with WAIS-III (Weschler 1997) and WASI (Weschler 1999) and verbal memory was collected with a Pen and Paper neuropsychological test battery, and CogState Research Battery (Pietrzak *et al.* 2009).

cNote, Mean and Standard Deviation of cognitive domains in Controls does not equate to exactly 0 and 1, respectively, as the norms for cognitive domains were calculated before exclusion of a subset of controls for this study due to imaging quality control.

Abbreviations: CDSS, Calgary Depression Scale for Schizophrenia. CPZ, chlorpromazine. FEP, First episode of psychosis. SAPS, Scale for Assessment of Positive Symptoms

**SUPPLEMENTARY REFERENCES**

**Addington D, Addington J, Schissel B** (1990). A depression rating scale for schizophrenics. *Schizophrenia research* **3**, 247–51.

**Akaike H** (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike, Springer Series in Statistics*, pp199–213. Springer New York: New York, NY.

**Andreasen N** (1984a). *Scale for the Assessment of Negative Symptoms (SANS)*. University of Iowa: Iowa City.

**Andreasen N** (1984b). *Scale for the Assessment of Positive Symptoms (SAPS)*. University of Iowa: Iowa City.

**Bartzokis G, Lu PH, Stewart SB, Oluwadara B, Lucas AJ, Pantages J, Pratt E, Sherin JE, Altshuler LL, Mintz J, Gitlin MJ, Subotnik KL, Nuechterlein KH** (2009). In vivo evidence of differential impact of typical and atypical antipsychotics on intracortical myelin in adults with schizophrenia. *Schizophrenia Research* **113**, 322–331.

**Benoit A, Malla AK, Iyer SN, Joober R, Bherer L, Lepage M** (2015). Cognitive deficits characterization using the CogState Research Battery in first-episode psychosis patients. *Schizophrenia Research: Cognition* **2**, 140–145.

**Bodnar M, Malla AK, Makowski C, Chakravarty MM, Joober R, Lepage M** (2016). The effect of second-generation antipsychotics on hippocampal volume in first episode of psychosis: longitudinal study. . The Royal College of Psychiatrists *British Journal of Psychiatry Open* **2**, 139–146.

**Cassidy CM, Rabinovitch M, Schmitz N, Joober R, Malla A** (2010). A comparison study of multiple measures of adherence to antipsychotic medication in first-episode psychosis. *Journal of clinical psychopharmacology* **30**, 64–7.

**First M, Spitzer R, Gibbon M, JBW W** (1998). *Structured clinical interview for DSM-IV patient edition (SCID-I/P V and SCID-I/NP Version 2.0)*. New York: Biometric Research Department.

**Hamilton M** (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology* **32**

**Hollingshead A** (1965). *A Two-Factor Index of Social Position*. Yale University Press: New Haven, CN.

**June SK, Singh V, Jun KL, Lerch J, Ad-Dab’bagh Y, MacDonald D, Jong ML, Kim SI, Evans AC** (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage* **27**, 210–221.

**Leucht S, Samara M, Heres S, Patel MX, Woods SW, Davis JM** (2014). Dose equivalents for second-generation antipsychotics: The minimum effective dose method. *Schizophrenia Bulletin* **40**, 314–326.

**Lewis JD, Evans AC, Tohka J** (2018). T1 white/gray contrast as a predictor of chronological age, and an index of cognitive performance. *NeuroImage* **173**, 341–450.

**Oldfield R** (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113.

**Pietrzak RH, Olver J, Norman T, Piskulic D, Maruff P, Snyder PJ** (2009). A comparison of the CogState Schizophrenia Battery and the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Battery in assessing cognitive impairment in chronic schizophrenia. *Journal of Clinical and Experimental Neuropsychology* **31**, 848–859.

**Roiz-Santiáñez R, Tordesillas-Gutiérrez D, Ortíz-García de la Foz V, Ayesa-Arriola R, Gutiérrez A, Tabarés-Seisdedos R, Vázquez-Barquero JL, Crespo-Facorro B** (2012). Effect of antipsychotic drugs on cortical thickness. A randomized controlled one-year follow-up study of haloperidol, risperidone and olanzapine. . Elsevier B.V. *Schizophrenia Research* **141**, 22–28.

**Vita A, De Peri L, Deste G, Barlati S, Sacchetti E** (2015). The Effect of Antipsychotic Treatment on Cortical Gray Matter Changes in Schizophrenia: Does the Class Matter? A Meta-analysis and Meta-regression of Longitudinal Magnetic Resonance Imaging Studies. *Biological Psychiatry* **78**

**Weschler D** (1997). *Weschler Adult Intelligence Scale - 3rd Edition*. The Psychological Corporation: San Antonio, TX.

**Weschler D** (1999). *Weschler Abbreviated Scale of Intelligence*. The Psychological Corporation: San Antonio, TX.