Intersection of verbal memory and expressivity on cortical contrast and thickness in first episode psychosis

Carolina Makowski1,2,3,4, John D. Lewis1,2,3, Claude Lepage1,2,3, Ashok K. Malla1,3, Ridha Joober4,5, Alan C. Evans1,2,3,* and Martin Lepage4,5*

1McGill Centre for Integrative Neuroscience, McGill University, Montreal, Canada; 2McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada; 3Ludmer Centre for Neuroinformatics and Mental Health, Montreal, Canada; 4Department of Psychiatry, McGill University, Verdun, Canada and 5Prevention and Early Intervention Program for Psychiatry, Douglas Mental Health University Institute, Verdun, Canada

Abstract

Background. Longitudinal studies of first episode of psychosis (FEP) patients are critical to understanding the dynamic clinical factors influencing functional outcomes; negative symptoms and verbal memory (VM) deficits are two such factors that remain a therapeutic challenge. This study uses white-gray matter contrast at the inner edge of the cortex, in addition to cortical thickness, to probe changes in microstructure and their relation with negative symptoms and possible intersections with verbal memory.

Methods. T1-weighted images and clinical data were collected longitudinally for patients (N = 88) over a two-year period. Cognitive data were also collected at baseline. Relationships between baseline VM (immediate/delayed recall) and rate of change in two negative symptom dimensions, amotivation and expressivity, were assessed at the behavioral level, as well as at the level of brain structure.

Results. VM, particularly immediate recall, was significantly and positively associated with a steeper rate of expressivity symptom decline (r = 0.32, q = 0.012). Significant interaction effects between baseline delayed recall and change in expressivity were uncovered in somatomotor regions bilaterally for both white-gray matter contrast and cortical thickness. Furthermore, interaction effects between immediate recall and change in expressivity on cortical thickness rates were uncovered across higher-order regions of the language processing network.

Conclusions. This study shows common neural correlates of language-related brain areas underlying expressivity and VM in FEP, suggesting deficits in these domains may be more linked to speech production rather than general cognitive capacity. Together, white-gray matter contrast and cortical thickness may optimally inform clinical investigations aiming to capture peri-cortical microstructural changes.

Background

Negative symptoms (e.g. amotivation, flattened affect) and cognitive impairments in individuals with psychotic disorders are strong predictors of poor functional outcome (Albert et al., 2011; Lepage et al., 2014), and are less responsive to currently available medications compared to positive symptoms (e.g. hallucinations, delusions). Verbal memory (VM) is one of the most strongly impacted of all cognitive domains in individuals with psychotic disorders (Jordan et al., 2014; Benoit et al., 2015; Guimond et al., 2016), and is linked to persistent negative symptoms (Hovington et al., 2013). There is some evidence suggesting that VM and negative symptoms, particularly deficits in communication or expressivity, may share common neural substrates. Cohen and Eldevag (2014) have suggested that expressivity is linked to classic language areas in psychiatric disorders, although such a relationship has not been demonstrated in the early stages of psychosis with non-invasive imaging techniques.

Many studies treat VM and negative symptoms as unitary constructs, although these variables are comprised of stable subdomains that are likely to be of clinical relevance; for instance, VM can be broken down into verbal learning and retention, and negative symptoms into amotivation and expressivity, all of which have unique environmental and biological correlates (Malla et al., 2002; Leeson et al., 2009; Millan et al., 2014). For instance, different facets of VM have been differentially associated to white matter microstructure (typically assessed using metrics such as fractional anisotropy from diffusion-weighted imaging) in healthy older adults, where white matter tracts subserving left fronto-parietal regions are related to verbal working memory, whereas bilateral fronto-temporal white matter is linked to long-term episodic memory function (Charlton et al., 2013). Both grey and white matter abnormalities have been observed in relation to amotivation symptoms (e.g. avolition, apathy) in enduring
schizophrenia patients, particularly decreased cortical thickness and fractional anisotropy underlying left orbital/medial frontal cortex and cingulate cortex (Ohtani et al., 2014; Mørch-Johnsen et al., 2015, 2018). A recent study has also shown that the integrity of white matter underlying the right hemisphere homologs of these regions is related to affective flattening (Ohtani et al., 2015); however, beyond this study, very few have further investigated the structural brain correlates of expressivity deficits (Mørch-Johnsen et al., 2018), especially in first episode of psychosis (FEP) patients.

Importantly, and of relevance to the current study, longitudinal brain imaging studies of negative symptoms in FEP are scarce, although altered maturational trajectories of cortical thickness and limbic structure have been previously observed in patients with persistent negative symptoms (Makowski et al., 2016, 2017). Such longitudinal neuroimaging studies are critical to characterizing the emergence of clinical, cognitive, and neuroanatomical markers that may be amenable to intervention early in the course of the disorder. Existing literature examining progressive brain changes after a FEP, irrespective of symptom profiles, have reported widespread accelerated gray matter loss (Andreasen et al., 2011; Gong et al., 2016). Various structural and functional MRI reviews of FEP patients have pinpointed progressive brain changes occurring across frontal and temporal lobes, particularly within the left hemisphere (Radua et al., 2012; Vita et al., 2012; Dazzan et al., 2015). Further, evidence suggests that the stability of medial temporal and prefrontal cortices may be essential predictors of symptomatic and functional outcomes after the first episode of schizophrenia (Dazzan et al., 2015).

White matter abnormalities have also been a key focus of studies of schizophrenia and related disorders, across different disease stages (Kubickei et al., 2005; Whitford et al., 2007, 2012; Carletti et al., 2012; Lee et al., 2013; Birur et al., 2017; Klausner et al., 2017; Kelly et al., 2018). Evidence from post-mortem studies and diffusion tensor imaging suggests that white matter abnormalities, particularly within prefrontal regions, are more significantly correlated with negative symptoms compared with positive symptoms (Uranova et al., 2011; Asami et al., 2014). Other neuroimaging studies have obtained proxy measures of myelin to better understand the nature of white matter alterations in psychosis (Andreasen et al., 1991; Lang et al., 2014; Ganzetti et al., 2015; Iwatani et al., 2015). There is certainly a need to shift focus to earlier stages of psychosis, obtaining measures more proximal to a FEP, to better understand the cascade of brain structural alterations that follow.

Although measures of cortical thickness (CT) are often the choice of methodology for investigation of cortical structure in psychiatric disorders from T1-weighted MRI scans, inconsistencies have arisen in a number of studies of longitudinal cortical thickness trajectories in early psychosis, with a handful of studies contesting the evidence of progressive brain change after a FEP (Nesvåg et al., 2012; Roiz-Santiáñez et al., 2015; Haukvik et al., 2016), emphasizing the need for novel approaches to analyze structural data in clinical cohorts. Obtaining a measure of white-gray matter contrast (WGC) from T1-weighted MRI may provide a meaningful marker of myelin content and other biophysical properties that may complement measures of cortical thickness (Salat et al., 2009; Westlye et al., 2009; Andrews et al., 2017; Lewis et al., 2018). We have recently shown that WGC captures brain architectural features and putative network-level abnormalities in FEP patients more robustly than measures of cortical thickness (Makowski et al., 2019). Thus, using multiple measures to assess the integrity of the cortical mantle, including a proxy measure of peri-cortical myelin, could provide a better understanding of the neurobiological correlates underlying subdomains of negative symptoms and VM, particularly in FEP patients. This in turn could be paramount in developing more effective treatments in early psychosis, and other disorders that are characterized by similar deficits.

The aims of the current study were to (i) examine the relationship between two negative symptom dimensions (expressivity and amotivation) and deficits in two VM processes (immediate and delayed recall), (ii) relate expressivity and amotivation negative symptom dimensions to changes in CT and WGC across time, and (iii) see how relationships in (ii) interact with VM capacity. We expect expressivity and amotivation to be uniquely correlated with progressive changes in white-gray contrast, and to be moderated by VM abilities. Specifically, we postulate that immediate recall, a measure of verbal learning, will be more strongly correlated with expressivity both at the levels of brain and behavior, given that both constructs have been described as ‘core’ traits in psychosis and less influenced by external factors (Leeson et al., 2009; Lutgens et al., 2014). On the other hand, we expect delayed recall, a measure of verbal retention, to be more related to amotivation, due to these factors’ higher potential to be influenced by the environment.

**Methods**

**Sample**

Patients were recruited from the Prevention and Early Intervention Program (PEPP) at the Douglas Institute in Montreal, Canada, and were part of a longitudinal naturalistic outcome study. Details are outlined elsewhere (Iyer et al., 2015). Inclusion criteria at PEPP include a diagnosis of affective (e.g. bipolar disorder, depression with psychotic features) or non-affective psychosis (e.g. schizophrenia, schizoaffective), an IQ above 70, and limited (<1 month) to no previous exposure to antipsychotic medication. Patients recruited to PEPP (ages 18–35) were invited to take part in a neuroimaging study, comprising three time points (baseline, one/two-year follow-ups) as described in previous work (Makowski et al., 2016, 2017). Patients were recruited from February 2003 to February 2014. The first baseline scan took place in May 2004. It should be noted that many of the initial patients recruited for the study did not meet criteria for this neuroimaging investigation, as more than 6 months passed between entry to the PEPP clinic and their baseline scan. The last two-year follow-up scan and clinical assessment for the last patient took place in April 2016.

Non-clinical healthy controls were recruited through advertisements within the same local catchment area. All participants provided written informed consent, and the research protocol was approved by the Douglas Institute human ethics review board. From 150 FEP patients recruited for the study, 88 patients (male, N = 62) were included in analysis. Data from 80 healthy controls were included for the calculation of standardized VM scores, as described below. See online Supplementary Material for more detailed information on the final included sample and demographic/clinical data collection. Also, see online Supplementary Table S1 for a comparison of patients included in the study, compared to those who were excluded.
We opted to include both affective and non-affective psychoses in our sample of FEP patients, as our group has previously shown that clinical diagnoses play a negligible role in findings related to MRI-derived metrics, including cortical thickness (Makowski et al., 2016, 2017). Further, our aims are to investigate individual clinical/cognitive profiles, rather than work within a diagnostic framework, which is being recognized as a clear setback in the field of psychiatry (Insel et al., 2010; Owen, 2018; Wolflers et al., 2018). However, we also explored differences in key variables of interest, namely immediate/delayed recall, and amotivation and expressivity deficits, between patients on the schizophrenia spectrum and patients with an affective disorder. No significant differences emerged between groups as can be seen in online Supplementary Table S2.

**Negative symptom dimensions and change over time**

Negative symptoms were assessed using the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen, 1984), which has been shown to have good inter-rater reliability ($\kappa = 0.71$) at PEPP-Montreal (Jordan et al., 2018). Item-level scores from the SANS were used to define two principal dimensions of negative symptoms: amotivation and (lack of) expressivity. These dimensions were based on a consistent body of literature reporting a two-factor model of negative symptoms (Malla et al., 2002; Blanchard and Cohen, 2006; Kirkpatrick et al., 2006; Messinger et al., 2011; Jang et al., 2016; Marder and Galderisi, 2017). Items were assigned to either the amotivation or expressivity dimension based on a recent confirmatory factor analysis published by our group (Lutgens et al., 2019). Although other models have been proposed to categorize negative symptoms, for instance, a recent investigation uncovering a five-factor latent structure of negative symptoms in schizophrenia (Strauss et al., 2018), it is important not to dismiss a large breadth of literature that has linked motivation and expressivity dimensions of negative symptoms to functional outcomes in patients. Finally, a two-factor solution minimizes the burden of multiple comparisons and potential false positives for the purposes of our neuroimaging analysis.

Thus, based on Lutgens et al. (2019), amotivation in this study represented summed items from ‘Avolition-Apathy’ and ‘Anhedonia–Asociality’ domains, while the expressivity dimension represented summed items from ‘Affective Flattening/Blunting’ and ‘Alogia’ domains, excluding Item 6: ‘Inappropriate Affect’ and Item 10: ‘Poverty of Content of Thought’, given that these items do not effectively map onto the clinical construct of expressivity (Lutgens et al., 2019). To assess the mean rate of change in symptoms over time, a linear model was fit to each subject’s longitudinal amotivation and expressivity symptom data against the participants’ age, across all available time points for that participant. A slope was calculated and extracted for each subject to represent a single metric of change over time in symptoms.

**Verbal memory performance**

VM data were collected from a larger neuropsychological battery database for patients followed at PEPP. Due to a change in neuropsychological testing protocol, VM data were compiled from two different protocols: (1) a pen and paper format administered to patients who took part in the study from 2003 to 2010, using logical memory subtests of the Wechsler Memory Scale-Third Edition (WMS-III) (Wechsler, 1997); and (2) theCogState Research Battery (Pietrzak et al., 2009), administered from September 2010 onwards, using the International Shopping List task. Further details of both testing protocols and tasks have been described before (Benoit et al., 2015). Z-Scores were calculated for each neuropsychological test battery separately, using the mean and standard deviation of controls for immediate/delayed recall measures. Previous work from our group (Benoit et al., 2015) has shown that participants tested with theCogState Research Battery tended to perform better. Thus, in addition to using standardized scores, the test version was used as a covariate in analyses using VM data. Online Supplementary Table S3 demonstrates there are no significant version or group*version effects on verbal memory performance.

**MRI acquisition**

Scans were collected at the Montreal Neurological Institute, all on the same 1.5-Tesla Siemens Sonata MRI scanner. Structural T1-weighted volumes were acquired for each participant using a 3D gradient echo pulse sequence with sagittal volume excitation (resolution = $1 \text{ mm}^3$, repetition time = 22 ms, echo time = 9.2 ms, flip angle = 30°, 180 1 mm contiguous sagittal slices). The rectangular field of view (FOV) for the images was 256 mm (AP) 204 mm (SI).

**MRI post-processing**

**Cortical thickness**

Raw 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. Main processing steps include: (1) Registration of T1-weighted images to standardized space (Collins et al., 1994) and correction for non-uniformity artifacts (Sled et al., 1998); (2) segmentation of gray, subcortical gray and white matter, and cerebral spinal fluid (Zijdenbos et al., 2002; Tohka et al., 2004); (3) extraction of the white matter surface using a marching-cubes algorithm and extraction of the gray matter surface using the CLASP algorithm (Kim et al., 2005); (4) surface registration to a template for inter-subject correspondence (Lytton et al., 2007); (5) reverse transformation (initially done in step 1) to estimate CT in native space for each subject at 81 924 vertices using the t-laplace metric (Lerch and Evans, 2005); and (6) smoothing the data in native space with a 20 mm FWHM Gaussian kernel to diminish the impact of noise (Boucher et al., 2009). CT was estimated using the Laplacian distance between the two surfaces (Jones et al., 2000) across 81 924 vertices.

**White-gray matter contrast**

Measures of WGC were generated as follows, similarly to what is described in Makowski et al. (2019): (1) additional surfaces were created 1 mm on either side of the surface at the gray-white matter boundary (i.e. +1 mm corresponds to supra-white surface, and −1 mm corresponds to sub-white surface); (2) surface maps of the intensity of the T1-weighted MRI were generated and smoothed; and (3) a ratio was calculated, by dividing the intensity of the −1 mm point by the corresponding +1 mm point, as defined by Lewis et al. (2018). WGC values ranged from approximately 1.15 to 1.35 where lower values, closer to 1, reflect lower contrast (i.e. reduced gray-white matter distinction) whereas higher values reflect higher contrast (i.e. clearer distinction between gray and white matter). The WGC method is depicted...
in Fig. 1. More details of imaging methods and quality control procedures are outlined in online Supplementary Material.

**Statistical analyses of behavioral/clinical data**

Demographic and clinical variables were analysed with one-way ANOVAs for continuous variables or Kruskal–Wallis H tests for nominal variables. An ANCOVA was used to analyse differences in VM performance at baseline between patients and controls, covarying for test version and years of education. Additional tests to compare the effect of version on results with VM data, as well as exploratory analyses between change in negative symptoms and change in verbal memory scores for a subset of patients with longitudinal data available (N = 49), are in online Supplementary Material. Associations between rates of change in amotivation/expressivity negative symptom domains and immediate/delayed recall in FEP patients were evaluated with Pearson r-correlations, adjusting verbal memory values for age, sex, and test version, and taking into account multiple comparisons with a false discovery rate (FDR) correction (Benjamini and Hochberg, 1995). Analyses of behavioral and clinical variables were conducted using PASW Statistics 21 (SPSS Inc., 2009, Chicago, IL, USA) and were two-tailed with a critical p value of 0.05. Normality of continuous data was assessed with the Shapiro-Wilk test of normality. Non-parametric tests were used to compare demographic variables between patients and controls that were found not to be normally distributed. Demographic and clinical information was also compared to controls that were found not to be normally distributed. Non-parametric tests were used to compare demographic variables between patients and controls that were found not to be normally distributed. Non-parametric tests were used to compare demographic variables between patients and controls that were found not to be normally distributed.

**Surface-based brain analyses**

Vertex-wise analysis of WGC and CT were conducted using SurfStat in Matlab (http://www.math.mcgill.ca/keith/surfstat/). As with the negative symptom data and as described in Raznahan et al. (2011), a linear model was fit to each vertex for the WGC and CT data, yielding a single metric for each subject describing the rate of change in years in WGC and CT. To determine which covariates should be used for analyses, the Akaike Information Criterion (Akaike, 1998) was used to determine the best model. AIC values were compared to a ‘baseline’ model that included centered age and sex as covariates, given the well-documented impact of age and sex on neuroanatomy, as well as their influence on the onset and progression of psychosis (Tamnes et al., 2010; Ochoa et al., 2012; Makowski et al., 2016, 2017). Details are included in the online Supplementary Material, including an analysis testing the effect of antipsychotic medication on WGC and CT (online Supplementary Fig. S1) and justification on the exclusion of medication as a covariate.

The following model was used to assess the main effect of change in negative symptoms (η, representing either amotivation or expressivity deficits) on change in Y (representing either WGC or CT), covarying for centered age, sex, and mean σ (reflecting mean WGC or mean CT across the entire cortical surface for each participant):

\[
\Delta Y = \text{intercept} + \beta_1(\Delta \eta) + \beta_2(\text{Centered Age}) + \beta_3(\text{Test Version}) + \beta_4(\text{Mean}(\sigma)) + \epsilon
\]

Here, \( \beta_1 \) is the slope for the main effect of interest, namely, rate of change (\( \Delta \)) in negative symptoms per year. Note, \( \beta_2 \) and \( \beta_4 \) reflect measures at baseline. Positive values of the rate of Y change reflect an increase in either WGC and CT per year, whereas positive values in negative symptom reflect an increase (i.e. worsening) in symptoms per year.

The interaction between baseline VM abilities (\( \nu \), reflecting immediate recall or delayed recall) and change in negative symptoms over time (\( \eta \)) on rate of change in WGC and CT (\( \Delta Y \)) was tested with the following model:

\[
\Delta Y = \text{intercept} + \beta_1(\Delta \eta) + \beta_2(\nu) + \beta_3(\text{Test Version}) + \beta_4(\text{Centered Age}) + \beta_5(\text{Sex}) + \beta_6(\text{Mean}(\sigma)) + \beta_7(\Delta \eta'\nu) + \epsilon
\]

where \( \beta_7(\Delta \eta'\nu) \) represented the slope for the main predictor of interest; that is, the interaction between change in negative
symptoms and baseline verbal memory. Here, \( \beta_2, \beta_3, \beta_4, \) and \( \beta_6 \) were measures taken at baseline. For visualization purposes only, patients were divided using a median split into two groups, on the basis of their verbal memory performance: ‘mild to moderate VM deficits’ and ‘high VM deficits’. Specific details of these groups are outlined in the Results section.

Random field theory (RFT) (Worsley et al., 2004) was used for multiple comparison correction using a stringent p-cluster threshold of \( p = 0.001 \). Significant results are also presented with a more liberal cluster threshold of \( p = 0.01 \) (p-corrected <0.05), to show the extent of sub-threshold results. For all significant clusters found with the more stringent cluster threshold (\( p = 0.001 \)), the WGC and CT, or rate of change in WGC and CT, at the peak \( t \)-statistic was extracted and adjusted for centered age, sex, and mean (\( \sigma \)), and used to generate scatterplots to visualize results.

### Results

The final sample comprised 88 FEP patients, of which 59 patients completed all three scans, and the remainder had two scans. See Table 1 for descriptive statistics and clinical information. Importantly, FEP patients had a median duration of untreated psychosis of \( \sim \)20 weeks, highlighting the efforts that have been put forth by the PEPP-clinic to minimize patients’ pathways to care within our early intervention service (Iyer et al., 2015; MacDonald et al., 2018). This also emphasizes that our FEP patients are indeed highly representative of the first-episode time period. Included patients were also compared to 47 patients that were excluded from the current study on variables collected at baseline, and these results can be found in online Supplementary Table S1. In summary, included patients did not differ largely to excluded patients.

Summary statistics for the rate of change in negative symptoms domains for 88 FEP patients with longitudinal data were as follows: mean \( \Delta \)Amotivation = −1.80 (S.D. = 4.22, range = −11.72−13.61), mean \( \Delta \)Expressivity = −1.22 (S.D. = 5.19, range = −19.63−14.66). For amotivation symptoms, five patients had the same level (or absence) of symptoms over time (mean \( \Delta = -0.0038, S.D. = 0.0086 \)), 25 patients had worsening symptoms (mean \( \Delta = 3.12, S.D. = 2.91 \)), and 58 patients improved (mean \( \Delta = -4.08, S.D. = 2.72 \)). For expressivity symptoms, 16 patients had the same level (or absence) of symptoms over time (mean \( \Delta = 0, S.D. = 0 \)), 22 patients had worsening symptoms (mean \( \Delta = 4.64, S.D. = 4.02 \)), and 50 patients improved (mean \( \Delta = -4.18, S.D. = 4.01 \)). See online Supplementary Fig. S2 for the spread of longitudinal amotivation and expressivity negative symptom data.

### Relationships between verbal memory and \( \Delta \)Negative symptoms

Cross-sectional relationships between VM and negative symptoms at baseline are reported in online Supplementary Material, where significant negative associations between baseline expressivity and immediate recall (\( r = -0.35, q < 0.001 \)) and delayed recall (\( r = -0.28, q = 0.036 \)) were found; no relationships were found with amotivation. The relationship between VM (i.e. \( z \)-scores of immediate and delayed recall) at baseline and \( \Delta \)Negative symptoms was assessed in 87 patients (degrees of freedom = 82), for which baseline VM and longitudinal negative symptom data were available, using Pearson r correlations, adjusting verbal memory data for age, sex, and test battery. The only significant finding that survived FDR correction for multiple comparisons was a positive relationship between immediate recall at baseline and \( \Delta \)Expressivity (\( r = 0.32, q = 0.012 \)). In other words, better VM (i.e. less severe deficits) at baseline was associated with worsening expressivity deficits over time. A trend-like association in the same direction was also found between delayed recall and \( \Delta \)Expressivity (\( r = 0.19, p\)-uncorrected = 0.083). No significant associations were found with either immediate or delayed recall and \( \Delta \)Amotivation (immediate: \( r = 0.079, q = 0.476 \); delayed: \( r = 0.077, q = 0.486 \)). Finally, we assessed the relationship between change in VM and change in negative symptoms over the first year after a FEP for a subset of patients (\( N = 49 \)) for which data were available. No significant relationships were found, as reported in online Supplementary Material.

### Associations between \( \Delta \)Negative symptoms and \( \Delta \)WGC and \( \Delta \)CT

Worsening of amotivation symptoms was associated with greater increases in WGC over time within the left superior parietal lobe and right dorsal primary motor cortex and paracentral lobule. Overlapping findings were found with respect to \( \Delta \)Expressivity, with additional peaks in the left dorsal precentral gyrus and right cuneus. With respect to CT, \( \Delta \)Expressivity was negatively associated with \( \Delta \)CT (i.e. cortical thinning with worsening symptoms) within left frontal regions. \( \Delta \)Amotivation was not found to be significantly associated with changes in CT. Results are shown in Fig. 2, along with the brain regions that survived correction for multiple comparisons with a relaxed p-cluster threshold of \( p = 0.01 \). Plots were generated for regions that survived correction for multiple comparison with a stringent threshold of \( p < 0.001 \).

### Interaction between verbal memory and \( \Delta \)Expressivity on WGC and CT

Given that no significant associations were found between verbal memory and \( \Delta \)Amotivation at the behavioral level, we did not explore this contrast. No significant regions survived the stringent threshold for WGC, however, results with a relaxed threshold are in Fig. 3. For CT, both immediate and delayed recall interacted significantly and uniquely with changes in expressivity. The interaction between immediate recall and \( \Delta \)Expressivity was significantly associated with changes in thickness along the left hemisphere ventrally, including the left orbital and medial frontal, insular, temporal pole, and middle temporal regions. The interaction term between delayed recall and \( \Delta \)Expressivity showed significant effects on the right cuneus/primary visual cortex. Results are shown in Fig. 3, along with brain regions that survived correction for multiple comparisons with a relaxed p-cluster threshold of \( p = 0.01 \). Plots were generated for regions that survived correction for multiple comparison with a stringent threshold of \( p < 0.001 \). Even though no results survived correction for multiple comparisons with WGC with our stringent threshold, we also generated two additional plots with WGC data for right retrosplenial cortex and left central sulcus, to compare the nature of the associations against CT data.

To better visualize the direction of results, patients were divided into two groups based on their verbal memory performance using a median split, as described in the methods. Specifically, the groups were broken down as follows: ‘mild to moderate VM deficits’ (\( N = 43 \); for immediate recall patients had a \( z \)-score higher than \(-1.15; \)
Table 1. Demographic and clinical information

<table>
<thead>
<tr>
<th></th>
<th>FEP ($N = 88$)</th>
<th>Controls ($N = 80$)</th>
<th>Statistic</th>
<th>df</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>62 (70)</td>
<td>52 (65)</td>
<td>$\chi^2 = 0.57$</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>Right handed</td>
<td>73 (83)</td>
<td>73 (91)</td>
<td>$\chi^2 = 2.53$</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia/</td>
<td>63 (72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>schizophreniaform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective disorder</td>
<td>16 (18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delusional disorder</td>
<td>3 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosis not otherwise specified</td>
<td>6 (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at baseline$^a$</td>
<td>24.3 (4.1)</td>
<td>24.27 (3.3)</td>
<td>$t = -0.009$</td>
<td>164.1</td>
<td>0.99</td>
</tr>
<tr>
<td>Verbal memory – immediate$^{b,c}$</td>
<td>$-1.3 (1.4)$</td>
<td>$[87]$, 0.03 (1)</td>
<td>$F = 23.38$</td>
<td>1, 163</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Verbal memory – delayed$^{b,c}$</td>
<td>$-0.9 (1.1)$</td>
<td>$[87]$, 0.02 (0.8)</td>
<td>$F = 16.77$</td>
<td>1, 163</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Education in years$^*$</td>
<td>12.0 (2.5)</td>
<td>14.3 (2.4)</td>
<td>$U = 1887.00$</td>
<td></td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td>3.2 (1.1)</td>
<td>3.0 (1.0)</td>
<td>$U = 2817.00$</td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td>Performance IQ$^{b}$</td>
<td>98.7 (17.1)</td>
<td>106.9 (12.7)</td>
<td>$U = 2447.50$</td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>Verbal IQ$^{b}$</td>
<td>99.8 (16.1)</td>
<td>110.0 (14.9)</td>
<td>$U = 2345.50$</td>
<td></td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>Cumulative CPZ equivalent (in mg)</td>
<td>4753.1 (4770.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherence (%)</td>
<td>80 (24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration untreated psychosis (weeks)</td>
<td>76.4 (142.0)</td>
<td>Median: 20.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration untreated illness (years)</td>
<td>7.3 (6.5)</td>
<td>Median: 6.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 year FUP symptoms (N = 80)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amotivation</td>
<td>11.3 (6.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressivity</td>
<td>7.1 (7.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAPS</td>
<td>9.5 (12.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDSS</td>
<td>2.4 (2.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Window [Scan – Symptom Eval] (months)</td>
<td>0.7 (0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2 year FUP symptoms (N = 65)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amotivation</td>
<td>7.4 (6.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressivity</td>
<td>5.1 (6.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAPS</td>
<td>7.9 (13.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
for delayed recall, patients had a $z$-score higher than $-0.91$ and ‘high VM deficit’ ($N = 44$; for immediate recall, patients had a $z$-score lower than or equal to $-1.15$; for delayed recall, patients had a $z$-score lower than or equal to $-0.91$). Note, one patient had missing verbal memory data at baseline, thus 87 patients were included in this analysis. Visualization of these groups generally revealed that patients with mild to moderate VM deficits (i.e. more preserved VM ability) at baseline drove the positive association between changes in expressivity symptoms and changes in WGC. For CT, patients with mild to moderate VM deficits at baseline predominantly drove the negative association between changes in expressivity and changes in thickness (Fig. 3).

Results were also generated excluding mean ($\sigma$) as a covariate; results were largely similar (online Supplementary Fig. S3). To ensure results in the final analysis were specific to verbal memory, a general cognitive index was calculated across five domains (i.e. attention, executive function, speed of processing, working memory, visual memory) and controlled for in analyses, yielding similar results (online Supplementary Fig. S4). Additional methods pertaining to cognitive data can be found in captions of online Supplementary Fig. S4/Table S3.

**Discussion/conclusions**

These results provide insight into the relationship between surface-based brain metrics and two behavioral domains that contribute strongly to outcome in FEP patients, namely negative symptoms and VM. Two novel findings emerged: (1) rate of change in expressivity and amotivation negative symptoms over the two-year period following a FEP are associated with both overlapping and distinct changes in WGC, and are also associated with changes in left prefrontal regions in relation to CT; and (2) links between baseline VM and change in expressivity deficits are found both at the behavioral and neuroanatomical level, where significant interactions on the rate of change over time in neuroanatomy were uncovered, most pronounced with CT.

We also extend a key finding of cortical thinning of the prefrontal cortex in relation to negative symptoms in psychosis (Galderisi et al., 2015; Walton et al., 2017), a finding that supports long-standing evidence that the frontal lobes are key contributors underlying negative symptom severity (Turetsky et al., 1995; Wible et al., 2001). Importantly, our findings suggest thinning is more specific to changes in expressivity over a one to two year period. However, no associations were uncovered in the prefrontal cortex when examining WGC and negative symptom progression, suggesting these cortical thinning patterns may be due to neuroanatomical changes within more superficial layers. This idea is supported by several studies that cortical thinning of superficial layers is a key biological mechanism underlying altered neuroanatomy of the prefrontal cortex in schizophrenia (Wagstyl et al., 2016; Lake et al., 2017). Significant associations still emerged between expressivity deficits and WGC, but within primary sensory and motor regions. Similar patterns were uncovered with amotivation. We have previously shown that WGC of primary sensory and motor regions is related to general psychopathology in FEP patients (Makowski et al., 2019), and maybe related to alterations within the high levels of intracortical myelin that typically characterize these regions (Glasser and Van Essen, 2011). It is possible that the association between contrast changes in primary sensory and motor regions and negative symptom severity is not specific to negative symptoms, but could also be related to general positive symptoms and cognitive deficits more generally. Although there are few studies that have looked at the neural correlates of expressivity and amotivation separately, one of the key regions that have emerged in several investigations is that of the association between avolition and activation of the striatum (Galderisi et al., 2015). Given our conjecture of WGC as a proxy measure of peri-cortical myelin, putatively tapping into integrity of cortico-cortical and cortico-subcortical connections, it is interesting to consider this finding in light of potential aberrancies between cortico-striatal circuitry (Haber, 2016).

A key discovery in our behavioral findings was the stronger association between VM specifically with expressivity, as opposed to both negative symptom dimensions. Consistent with previous literature, we uncovered a negative association between VM performance and negative symptoms in psychosis cross-sectionally (Hartmann-Riemer et al., 2015). However, when assessing baseline VM performance against the rate of change in symptoms, we identified a relationship that, at first glance, runs counter to intuition. Namely, it was found that patients with better VM at
baseline (i.e. a mild level of VM deficits, particularly with immediate recall) had worsening of expressivity symptoms, a relationship that has largely been unexplored longitudinally in psychosis. Our results suggest that having less striking VM deficits at intake does not necessarily protect against progressive changes in expressivity. Other key cognitive factors, such as insight, may also contribute to this relationship. Based on previously identified positive relationships between cognitive insight and VM (Lysaker et al., 2005; Lepage et al., 2008), we can consider that patients with a better ability to recall stored information are also likely to have lower self-certainty and more self-reflectiveness (a profile that characterizes high levels of cognitive insight), which could set the stage for a trajectory of worsening expressivity as the patient internalizes their mental illness. It is also important to note that many of the patients included in this study had an improving course of negative symptoms after the first two years after a

Fig. 2. Change in amotivation and expressivity deficits associated with a change in WGC/CT across time. Results are RFT-corrected, where blue colors represent significant results cluster-thresholded at a 'stringent' threshold of \( p = 0.001 \), whereas red/pink color represent significant results at a 'relaxed' threshold of \( p = 0.01 \). Selected regions for scatterplots are those that survived multiple comparisons with the more stringent cluster threshold of \( p = 0.001 \). Plots with amotivation are plotted with a dotted line and crosses, whereas expressivity data are plotted with a solid line and circular markers. For overlapping cluster within right paracentral gyrus of the effect of negative symptoms on WGC (top two rows), a common peak was selected where amotivation and expressivity are included on the same plot. SPL, superior parietal lobule. Orientation: Surfaces from left to right in each row: left lateral, right lateral, left medial, right medial, dorsal.
Fig. 3. Interaction between baseline verbal memory and change in Expressivity on WGC and CT. Results are RFT-corrected, where blue colors represent significant results cluster-thresholded at ‘original’ stringent threshold of $p = 0.001$, whereas red/pink color represent significant results at a ‘relaxed’ threshold of $p = 0.01$. ‘Immediate’ and ‘Delayed’ labels of 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. Selected regions for scatterplots are those that survived multiple comparisons with the more stringent cluster threshold of $p = 0.001$ for cortical thickness data. Although nothing survived correction after stringent correction for WGC analyses, peaks within left central sulcus and right retrosplenial cortex were plotted to explore the direction of results with WGC. To better visualize the interaction effect of two continuous variables (rate of change in expressivity and baseline VM), patients were divided based on low/moderate VM deficit (purple) and high VM deficit (yellow). Plots separating patients on the basis of immediate recall are depicted with crosses, and delayed recall with diamonds. 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 last surface is a ventral view).
FEP, consistent with a previous study (Lutgens et al., 2019). In this context, several studies have suggested that negative symptoms can improve in the absence of any change in verbal memory performance (Nopoulos et al., 1994; Cantor-Graae et al., 1995; Hoff et al., 1999). In a related vein, we also conducted a supplementary analysis exploring the potential relationship between change in verbal memory and change in negative symptoms over the first year after a FEP, for a subset of patients (N = 49) for which data were available. No significant association was found, consistent with the studies cited above, but it is still possible that more time points would be needed to elucidate such longitudinal relationships. Many studies have suggested that cognition, including verbal memory, is a set of static traits, as suggested by evidence showing the stability of cognitive performance from the clinical high-risk state to transition to psychosis (Green and Harvey, 2014; Carrion et al., 2015). However, future studies are encouraged to further investigate longitudinal cognition-symptom relationships after psychosis onset; a future direction that our group is currently investigating as well.

It is also worth discussing the granularity of results with respect to immediate vs. delayed recall in relation to expressivity. It has been shown that immediate recall is highly related to performance on other measures of cognition in schizophrenia (Leeson et al., 2009). Immediate recall may also have a stronger biological and/or genetic predisposition; deficits in immediate recall predict conversion to psychosis (Lindgren et al., 2017) and are present in non-affected siblings of schizophrenia twins (Goldberg et al., 1993). Together, this suggests that immediate recall may be a more ‘hard-coded’ feature of psychosis compared to retention of information (i.e. delayed recall), with stronger biological and genetic influences. Further, expressivity has also been recently posited to be a more primary negative symptom domain, and less influenced by environmental factors, such as community resources, compared to amotivation (Lutgens et al., 2019). Our data present a feasible model whereby expressivity and immediate recall or verbal learning may share common etiology and a potential endophenotype of interest for future studies.

More preserved immediate recall abilities contributed to the relationship between worsening of expressivity deficits and cortical thinning across higher-order language processing areas in ventral frontal areas and middle/inferior temporal gyri of the left hemisphere, which contributes to the semantic encoding of language (Saur et al., 2008) and for keeping verbal information active in working memory (Smith and Jonides, 1999). These areas are subserved by white matter tracts such as the uncinate and superior longitudinal fasciculi (Friederici, 2011), which have been shown to be disrupted in schizophrenia patients with predominant negative symptoms (Sigmundsson et al., 2001). With WGC, an interesting relationship also emerged between expressivity deficits and increases in contrast between patients with different levels of immediate recall performance within the retrosplenial cortex bilaterally. Although these results were only significant with a relaxed threshold when correcting for multiple comparisons, the retrosplenial cortex is interesting to consider in this context as it is a region that has been strongly linked to memory function (Vann et al., 2009), as well as more recently, to the reinforcement of salient behavioral cues (Smith et al., 2012, 2018), which is likely to contribute during immediate learning phases of memory. A previous study has also shown that white matter volume abnormalities exist within the retrosplenial cortex in schizophrenia patients (Mitelman et al., 2005).

Meanwhile, the interaction between delayed recall and expressivity uncovered regions of the dorsal language processing stream, namely superior parietal lobule, dorsal primary somatomotor and premotor areas, which relates more to language production (Saur et al., 2008). Disruptions in these cortical regions could plausibly support a mechanistic explanation of the lowered rate of speech production and/or increased psychomotor retardation observed in patients with expressivity deficits (Cohen et al., 2014; Marder and Galderisi, 2017), as well as the necessity for preserved structure/function of the left posterior dorsolateral prefrontal cortex in delayed retrieval (Alexander et al., 2003). Our results also did not change when accounting for a general cognitive index that comprised five cognitive domains. Together, these results suggest that the relationship between expressivity deficits and VM may be more closely linked to speech articulation or language production mechanisms, as opposed to general cognitive abilities.

It is worth noting that WGC and CT are related metrics, but they also tap into different neurobiological properties (Westlye et al., 2009). WGC may take myelination patterns into account more strongly than CT, and may shed light on a compartment of the cortex that may be more amenable to short-term changes (Wenger et al., 2017), which in turn may be a more viable biomarker target in FEP patient samples as measured on MRI.

Several limitations should be considered. Our model did not capture patients who had a consistently high or low level of negative symptoms, as well as the distinction between primary and secondary negative symptoms. These individual patterns of negative symptoms have been touched upon by some of our previous work (Makowski et al., 2016, 2017), but future investigations are encouraged to parse apart such trajectories in myelin and/or other microstructural compartments. The rates of change in both cortical metrics and negative symptoms should also be interpreted as a proxy measure of negative symptom/cortical change, as subtle non-linear fluctuations over time are not captured by the measures presented in this manuscript. Limitations in accurately and objectively assessing negative symptoms should also be considered, where rater bias may preclude the treatment of amotivation and expressivity as completely independent domains. Automated unbiased assessments of negative symptoms, such as software that has been developed to assess communication deficits and natural language in patients with schizophrenia (Cohen and Elvevåg, 2014), may provide data with higher construct validity in this respect. It should also be noted that many patients ultimately showed an improvement in negative symptoms; although this improvement has been reported before in the first two years after a FEP (Lutgens et al., 2019), we cannot rule out that patients with improvement in negative symptoms ultimately comprise a subset of patients that are more motivated to participate in longitudinal studies such as the one described here. With respect to cognitive data, we collected VM from two batteries, although examination of the tests and standardization of our data suggests test version likely did not significantly impact our findings. Our VM data was also limited by the fact that it was largely cross-sectional; as mentioned, future studies would benefit from studying potential temporal changes in such cognitive performance after a FEP. Finally, although our patient sample was largely antipsychotic-naïve at the beginning of the study, many of these patients were prescribed antipsychotic medications. Our exploration of antipsychotic medication on the rate of change in WGC and CT showed no notable effect, however individual medication classes could have differential effects on gray/white matter...
Although further work is needed to clarify the longitudinal relationships between cognition and negative symptoms, this avenue is promising in understanding domains of psychosis which have largely gone untreated. These findings may also hold implications for other neurological and psychiatric disorders characterized by negative symptom presentation and verbal memory deficits. Given that structural MRI is one of the most commonly used imaging techniques in psychiatric neuroimaging studies, extracting WGC and CT together may be fruitful in the broader search for a valid and easy-to-measure biomarker in psychiatry.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291719002071.

Acknowledgements. We would like to thank PEPP-Montreal and Lepage Lab research staff for their efforts in recruitment and clinical data collection. We are also grateful to all patients and families for participating in the study.

Financial support. The study was supported by operating grants from the Canadian Institutes of Health Research (CIHR, 68961, MCT-94189); the Fonds de Recherche du Quebec – Santé (FRSQ); Sackler Foundation; and by grant ANRP-MIR13-3388 from the Aziëti Neurolodëvelopmental Research Program in partnership with the Brain Canada Multi-Investigator Research Initiative (to A.C.E.). It also benefited from computational resources provided by Compute Canada (www.computecanada.ca) and Calcul Quebec (www.calculquebec.ca). Salary awards include: Canadian Institutes for Health Research (CIHR; C.M., R.J., M.L., A.C.E.), Fonds de la Recherche en Santé du Quebec (FRSQ; C.M., M.L., and R.J.), funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives Initiative (C.M., A.C.E.), James McGill Professorship (M.L., A.C.E.) and Canada Research Chairs Programme (A.K.M.). The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of interest. A.C.E. has received consulting fees from Biospective. M.L. reports grants from Otsuka Lundbeck Alliance, personal fees from Otsuka Canada, personal fees from Lundbeck Canada, grants and personal fees from Janssen, and personal fees from MedAvante-Prophouse, outside the submitted work. R.J. reports receipt of grants, speaker’s fees, or honoraria from AstraZeneca, BMS, Janssen, Lundbeck, Otsuka, Pfizer Canada, Shire, and Sunovion and royalties from Henry Stewart Talks. A.M. reports receipt of grants, fees, or honoraria from BMS, Lundbeck, and Otsuka. The authors have no other competing interests to disclose.

References
Birur B, Kraguljac NV, Shelton RC and Lahti AC (2017) Brain structure, function, and neurochemistry in schizophrenia and bipolar disorder—a systematic review of the magnetic resonance neuroimaging literature. NPJ Schizophrenia 3, 15.


