A Subtype of Institutionalized Patients with Schizophrenia Characterized by Pronounced Subcortical and Cognitive Deficits

Supplementary Materials

Qiannan Zhao M.D.[1, 2, 3], Hengyi Cao Ph.D.[1, 4, 5], Wenjing Zhang Ph.D.[1, 2, 3], Siyi Li Ph.D.[1, 2, 3], Yuan Xiao M.D., Ph.D.[1, 2, 3], Carol A. Tamminga M.D.[6], Matcheri S. Keshavan M.D.[7], Godfrey D. Pearlson M.D.[8], Brett A. Clementz M.D.[9], Elliot S. Gershon M.D.[10], Scot Kristian Hill, Ph.D.[11], Sarah K. Keedy Ph.D.[10], Elena I. Ivleva M.D.[6], Rebekka Lencer M.D., Ph.D.[12], John A. Sweeney Ph.D.[1, 13], Qiyong Gong M.D., Ph.D.[1, 2, 3*] and Su Lui M.D., Ph.D.[1, 2, 3*]

[1] Huaxi MR Research Center, Department of Radiology, West China Hospital of Sichuan University, Chengdu, Sichuan Province, China
[2] Research Unit of Psychoradiology, Chinese Academy of Medical Sciences, Chengdu, Sichuan Province, China
[3] Functional and Molecular Imaging Key Laboratory of Sichuan Province, West China Hospital of Sichuan University, Chengdu, Sichuan Province, China
[4] Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY, US
[5] Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, US
[6] Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, US
[7] Department of Psychiatry, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, US
[8] Departments of Psychiatry and Neurobiology, Yale University and Olin Neuropsychiatric Research Center, Hartford, CT, US
[9] Department of Psychology, University of Georgia, Athens, GA, US
[10] Department of Psychiatry, University of Chicago, Chicago, IL, US
[11] Department of Psychology, Rosalind Franklin University of Medicine and Science, Chicago, IL, US
[12] Department of Psychiatry and Psychotherapy, University of Münster, Münster, Germany
[13] Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH, US

Table of Contents

Supplementary Methods

1. Samples and methodological details of structural imaging acquisition

Three independent samples with brain structural images were used in this study, including a sample of institutionalized patients and two samples of community-dwelling patients with schizophrenia. Demographically matched healthy controls were included at a 1:1 ratio from corresponding data sets for comparison purposes. The scanning parameters, demographics, and clinical profiles of these samples are displayed in **Table S1 – S2**. The Brief Assessment of Cognition in Schizophrenia (BACS) [1] was used to evaluate cognitive function, and test scores were available in the institutionalized sample and the community-dwelling sample with long-term illness (**Table S3**).

1.1 The subtype-discovery set

In the subtype-discovery set, a total of 96 institutionalized patients with long-term schizophrenia (46.48 ± 7.25 years, 31 females) and 96 age-, and sex-matched healthy controls (46.65 ± 7.53 years, 31 females) were recruited.

Brain structural images of participants in this set were acquired on a 3-T scanner using a spoiled gradient recalled (SPGR) sequence with an eight-channel head coil at the Department of Radiology, West China Hospital, Sichuan University, Chengdu, China.

1.2 Community-dwelling sets

The present study used two data sets of community-dwelling individuals with schizophrenia at different illness stages and demographically matched healthy controls to compare their brain-behavior patterns and corresponding heterogeneity with the institutionalized sample.

1.2.1 Community-dwelling set 1 (B-SNIP sample, N = 136). We included 68 community-dwelling patients with long-term schizophrenia (36.94 ± 10.80 years, 31 females) and 68 age-, sex-, site-, and race-matched healthy controls (37.29 ± 10.74 years, 31 females) from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) consortium. The illness duration of these patients was greater than or equal to 5 years.

These participants were scanned on 3-T scanners across multiple sites using a magnetization-prepared rapid gradient-echo (MP-RAGE) sequence or an inversion recovery-prepared spoiled gradient-echo (IR-SPGR) sequence, following the Alzheimer's Disease Neuroimaging Initiative (ADNI1) protocol (http://adni.loni.usc.edu/methods/documents/mri-protocols/).

1.2.2 Community-dwelling set 2 (FES sample, N = 252). We also included 126 community-dwelling drug-naïve patients with first-episode schizophrenia (22.98 ± 7.40 years, 74 females) and 126 age-, and sex-matched healthy controls (23.82 ± 7.20 years, 74 females). These participants had illness duration less than or equal to 2 years.

T1-weighted images were acquired from all participants in the FES sample, using the same scanner and protocol applied in the institutionalized sample.

1.3 Details of quality assessment and preprocessing for structural imaging

An experienced neuroradiologist inspected the quality of all images, and participants with any gross abnormalities or scanning artifacts in brain regions were excluded. T1-weighted images from all participants were processed by the FreeSurfer (https://surfer.nmr.mgh.harvard.edu) software (version 6.0) with a 'recon-all' pipeline [2]. Euler number is used to assess cortical reconstruction as a quality measure produced in the FreeSurfer pipeline [3]. Euler number was calculated per hemisphere, averaged across the bilateral hemisphere, and converted into z-scores across all participants in each sample. Z-scores of Euler number were used for statistical inferences.

2. Methodological details of cluster analyses in the institutionalized sample

2.1 K-means++ algorithm

K-means++ algorithm [4] was applied for cluster analysis in 96 institutionalized patients. K-means algorithm can have limitations in clustering accuracy, while the K-means++ algorithm combines typical k-means technique and the randomized initial centers setting to achieve better accuracy.

K-means++ clustering was performed based on volumes of 14 subcortical regions in institutionalized patients with schizophrenia. For these subcortical features, variance related to age, sex, and intracranial volume (ICV) were removed and z-transformations were subsequently performed before clustering. The open-source software R (version 4.0.2) [5] and Python (version 3.8.8) [6] were used for cluster analysis and validation.

2.2 Identification of the optimal number of clusters

When clustering was performed, parameters including center initialization and the number of clusters (k) were tuned to achieve better model performance, and the optimal number of clusters was identified based on the silhouette coefficient [7].

For a given individual, the silhouette width is used to measure how similar a case is to its own cluster compared to the other clusters, ranging from -1 to 1. A silhouette width close to 1 indicates that the data point is close to other data points in its own cluster but poorly matched to its neighboring cluster, representing that this individual is appropriately clustered. In contrast, a silhouette width close to -1 means that the object is dissimilar to other data points in its own cluster but similar to those in its neighboring cluster, indicating that this object is more appropriate if it was clustered in the neighboring cluster. Subsequently, the average silhouette width for the entire data set can be calculated to select the optimal number of clusters. The silhouette coefficient is defined as the largest value of the average silhouette width for the entire data set, taken over all predefined k for which the silhouette values could be constructed (in this work, k = 2, 3, …, 21).

2.3 Clustering performance evaluation and validation

2.3.1 Adjusted Rand index. The adjusted Rand index (ARI) [8] is a measure used to assess the robustness of clustering findings. As a corrected-for-chance form of the Rand index [9], the ARI assesses the agreement between two clustering results by counting all pairs of observations assigned into the same or different clusters based on the predicted and actual clustering findings. The ARI ranges between -1.00 to 1.00. A value of ARI close to 0.00 indicates random labeling, and a value of 1.00 represents that the clustering results are identical. To evaluate the robustness of the primary clustering findings in institutionalized patients, we calculated the ARI in 5-fold cross-validation with k ranging from 2 to 21. The ARI and V-measure considered next look at the correctness of clustering, so require two or more clusters for their computation.

2.3.2 V-measure. We also included the V-measure [10], which is calculated as the weighted harmonic mean of homogeneity and completeness scores, to evaluate the goodness of our clustering results. A clustering solution satisfies homogeneity when each cluster has observations belonging to a single class label. A clustering solution satisfies completeness if all observations of a given cluster are clustered into the same cluster. V-measure ranges from 0.00 to 1.00, and a value of 1.00 indicates perfectly complete labeling. We also calculated the V-measure in 5-fold cross-validation with k ranging from 2 to 21 to validate the primary clustering findings.

2.4 Secondary cluster analyses

To compare clustering results of our primary analysis based on subcortical volumes to those using other features, we performed four additional cluster analyses based on regional cortical volumes, regional cortical and subcortical volumes, global cortical volumes, and global brain volumes, named Model 1, Model 2, Model 3, and Model 4, respectively. Features used in these models include 1) volumes of 68 neocortical regions based on Desikan-Killiany atlas [11] for Model 1; 2) volumes of 68 neocortical regions mentioned above and volumes of 14 subcortical regions for Model 2; 3) left cortical gray matter volume (GMV), right cortical GMV, and cortical GMV for Model 3; and 4) cortical GMV, subcortical GMV, total GMV, cerebral white matter volume (WMV), and total brain volume (TBV) for Model 4.

3. Sensitivity analyses for the primary cluster analysis in the institutionalized sample

3.1 Illness duration for subtyping

To determine whether the identified subtypes of institutionalized patients would be driven by different illness duration, we additionally conducted cluster analysis based on subcortical volumes with prior removal of variance related to age, sex, ICV, and illness duration.

To assess whether patient proportions would be statistically different between the primary subtyping results and the one identified in this sensitivity analysis, we employed z-tests to test differences in patient proportions of Subtype 1. Illness duration was compared between the additionally identified subtypes of patients, and neuroanatomic and cognitive patterns found in this sensitivity analysis were compared with those in primary cluster findings.

3.2 ICV for brain volume comparisons

We conducted several analyses to assess whether ICV would differ across identified subtypes in institutionalized patients and whether ICV would influence between-group differences in brain volumes. First, we compared ICV with ANOVA between identified subtypes and corresponding healthy controls. Second, we re-conducted between-group comparisons in brain volumes when ICV was removed from the covariates.

3.3 Associations between medication or illness duration with brain-behavior measures

Correlation analyses between cognition and subcortical volumes with illness duration or the daily dose of antipsychotics were performed where age, sex, education level, and ICV were considered as covariates. Variance related to these covariates were removed before correlation analyses.

3.4 The importance of using subcortical volumes as subtyping features

We conducted correlation analyses between regional or global brain volumes and cognitive function in the whole group of institutionalized patients with schizophrenia. These analyses were used to investigate whether the associations between cognition and subcortical measures would be stronger than those with cortical measures. Regional volume from 68 neocortical regions and 14 subcortical structures, and global measures including cortical GMV, cerebral WMV, subcortical GMV, total GMV, and TBV, were employed as neuroanatomic features to be analyzed. Age, sex, and education level were treated as covariates for cognitive scores, while age, sex, and ICV were included as covariates for brain volumes. Variance related to corresponding covariates was removed before correlation analyses. Such correlation analyses were repeatedly conducted without removing the ICV variance.

4. Methodological details of classifier training, validation, and patient assignment for community-dwelling samples

4.1 Random-Forest algorithm

A brain-based classifier was trained in institutionalized patients using the random forest algorithm [12], a powerful machine learning method that could be applied for both regression and classification. Random forest algorithms combine bootstrap aggregating with decision trees to reduce the influence of noise and overfitting [13]. Many decision trees are created by applying bootstrap aggregating, and each decision tree is trained on a bootstrap sample of cases from the original training set. The number of predictive variables is selected at each split for a particular tree. Multiple decision trees are part of the model and make their independent predictions. The most frequent prediction is subsequently generated after new data are passed through the model by taking the majority vote.

4.2 Random-Forest classifier training and validation

R packages ‘randomForest’ [14] and ‘caret’ [15] were used to train and validate the brain-based classifier in institutionalized patients and subsequent patient assignment

in community-dwelling samples. Institutionalized patients with schizophrenia were randomly split into training and test sets (70%/30%). Subcortical volumes, the same features used in cluster analysis, were used for classifier modeling. Feature standardization steps for classifier modeling were the same with those employed in clustering, including the removal of variance related to age, sex, and ICV, as well as z-transformations. Clustering results (i.e., subtype labels) were the categorical variable to be predicted. In the training set (i.e., 70% of institutionalized cases), the random forest algorithm following with 100 repetitions of 10-fold cross-validation (CV) and tuning parameters (i.e., the number of sampled predictors at each split) was applied to avoid overfitting. The model with the highest average accuracy was identified as the optimal model. First, the model performance was assessed based on the average accuracy and Cohen's kappa in repeated CV. The prediction using the optimal model was then performed in the test set (i.e., the 30% of institutionalized cases held out from the training set), and measures such as accuracy, Cohen's kappa, sensitivity, specificity, the area under the receiver operating characteristic (ROC) curve (AUC) were also reported to assess the performance based on held-out data. Once the model achieved satisfactory performance, it was subsequently used for patient assignment in community-dwelling samples.

4.3 Feature importance assessment

Feature importance for the optimal classifier was measured by the mean decrease in accuracy using a permutation strategy, computed using the 'importance' function in the 'randomForest' package [14] by setting the 'scale' argument at 'FALSE' status. For each tree, the baseline accuracy is recorded by passing the out-of-bag (OBB) data through the random-forest model, which is done after every single predictive variable is permuted. The raw mean decrease in accuracy for the particular feature is calculated, defined as the difference between the baseline and the drop in averaged accuracy over all trees by permuting the predictive variable.

4.4 Participant assignment for the community-dwelling samples

To compare classification rates and brain-behavior patterns in institutionalized patients with that of other independent samples of interest (the community-dwelling samples), we assigned these community-dwelling patients based on the classifier developed in institutionalized patients. Each patient in community-dwelling samples was classified into identified subgroups, using the primary classifier based on subcortical volumes, where variance related to age, sex, and ICV were removed and z-transformations were performed before the assignment.

5. Within-sample statistical analyses in community-dwelling samples

5.1 Between-group comparisons

Participants in the B-SNIP sample (i.e., community-dwelling patients with long-term illness and controls) were scanned by six different scanners. The race of these participants was recoded as three levels, including Caucasian, African American, and others. Before subtyping, case-control comparisons in demographics were conducted with two-sample t-tests for continuous variables and chi-square tests for categorical variables. After patient assignment based on the classifier developed in

institutionalized patients, comparisons in community-dwelling samples were performed among identified subgroups of patients and healthy controls. Between-group differences in age, education level, Euler number, ICV were tested using ANOVA with Tukey HSD post hoc tests. Between-subgroup differences in illness duration, the daily dose of antipsychotics, PANSS scores were tested by two-sample t-tests. The distributions of sex, site, and race were compared using chi-square tests. Brain-behavior between-group differences were tested with ANCOVA and Tukey HSD post hoc tests. Brain volume comparisons were conducted with age, sex, and ICV as covariates. Cognitive comparisons were conducted in the B-SNIP sample with education level as the covariate because their BACS scores were already age-, and sex-corrected. FDR corrections were applied for comparisons in PANSS scores, regional and global brain volumes, and cognitive function. Glass's delta (Δ) effect sizes were calculated with the prior removal of corresponding covariates to display case-control differences in identified subgroups of patients. It is worth noting that Glass's delta effect sizes were calculated by dividing the standard deviation of the control group. The z-test, used to compare population proportions in two samples, was employed to test differences in patient proportions of Subtype 1 between each community-dwelling sample with the institutionalized sample.

5.2 Sensitivity analyses of site

To assess the influence of site for subtyping in the B-SNIP sample, we performed additional patient assignments in the community-dwelling individuals with long-term schizophrenia based on the classifier developed in institutionalized patients. In the subtyping procedures, besides variance related to age, sex, and ICV, we additionally removed the site variance for subcortical volumes for assignment.

To assess whether patient proportions would be statistically different between the primary subtyping and the one in this sensitivity analysis, we employed z-tests to test differences in patient proportions of Subtype 1. Neuroanatomic and cognitive patterns found in this sensitivity analysis were compared with those in primary subtyping findings in the community-dwelling sample with long-term illness.

We also conducted brain volume comparisons with site as the additional covariate with-in the B-SNIP sample.

5.3 Sensitivity analyses of race and education level

We conducted several models with different combinations of covariates for the BACS comparisons in the B-SNIP sample: 1) Model 1: no covariates were included; 2) Model 2: education level was included as the covariate; 3) Model 3: race was included as the covariate; 4) Model 4: race and education level were included as covariates.

5.4 Sensitivity analysis of Euler number

For samples that showed differences in Euler number between identified subtypes, brain volume comparisons included Euler number as the additional covariate was performed to evaluate the influence of cortical reconstruction quality measured by Euler number.

6. Statistical analyses across institutionalized and community-dwelling patients

6.1 Neuroanatomic between-sample analyses

We pooled healthy controls from each of the three samples to a single group for neuroanatomic profiles. We subsequently harmonized neuroanatomic features using the neuroCombat R package (https://github.com/Jfortin1/neuroCombat_Rpackage) [16] to handle the between-sample/within-sample scanner effects. At the same time, the information of age, sex, diagnostic group, and patient subtype were reserved during harmonization. Then, harmonized neuroanatomic features were employed in ANCOVA and Tukey HSD post hoc tests. For each subtype, the independent variable was the participant group, including the institutionalized patient group, the B-SNIP patient group, the FES patient group, and the pooled healthy control group. Dependent variables include averaged subcortical volumes, cortical volumes, and global brain volumes. Age, sex, and ICV were included as covariates. FDR corrections were applied for *P*-values generated in main and post hoc pairwise tests. Only features significantly different across groups in the main tests were tested in post hoc tests. We also reported Glass's delta effect sizes and the information of significance in pairwise comparisons.

6.2 Cognitive between-sample analyses

Considering cognitive function, measured by BACS scores in two of three samples, we performed standardization procedures before pooling due to their inconsistency. We transformed BACS raw scores into z-scores based on published norms [17] in the institutionalized sample, further combined with data from the B-SNIP sample. Healthy control subjects from the two samples were pooled as a single group for cognitive comparisons. ANCOVA with post hoc Tukey HSD tests were conducted to detect between-sample differences in cognition between institutionalized and B-SNIP patients and pooled healthy controls. Education level was the only covariate because of the prior age- and sex-corrections. FDR adjustment for P-values and reporting Glass's delta effect sizes and significance information was also made.

6.3 Brain-behavior associations across samples

To quantify the severity of brain-behavior profiles across subtypes or samples, we firstly pooled brain-behavior data from institutionalized and B-SNIP patients (i.e., community-dwelling patients with long-term schizophrenia) (see descriptions in **Sections 6.1 – 6.2**). We performed standardization procedures for brain-behavior data in this pooled patient group, including removing variance related to covariates (age, sex, and ICV for brain volumes, and education level for age- and sex-corrected BACS z-scores), as well as z-score transformations. Thus, the relative brain-behavior severity for individuals from this group could be quantified as the distance from zero in the negative direction.

In this pooled group, we carried out correlation analyses between regional (including volumes from cortical and subcortical regions) and global brain volumes and cognitive function to explore which pairs of features could better represent brain-behavior severity from the dimensional perspective. FDR adjustment was applied on generated *P*-values because of multiple analyses. For correlations to be shown significant, we

further conducted correlation analyses in each sample to identify whether they would display different patterns in terms of brain-behavior relationships.

Supplemental Results

1. Primary cluster analysis in the institutionalized sample

1.1 Identification of the optimal number of clusters

The silhouette coefficient was calculated for the k-means++ model with k ranging from 2 to 21. It achieved its highest value when k was equal to 2, suggesting that the optimal number of clusters is 2 for our primary cluster analysis (see **Figure S1**).

1.2 Cluster validation

In 5-fold CV with k ranging from 2 to 21, both the ARI and the V-measure reached their highest value when k was equal to 2, indicating that our 2-subtype solution provides the greatest precision in classification (see **Figure S2**).

2. Sensitivity analysis for the primary clustering findings in the institutionalized sample

2.1 Illness duration for subtyping

2.1.1. Subtyping. Another cluster analysis was performed after the variance related to illness duration was additionally removed for subcortical volumes. In this sensitivity analysis, institutionalized patients were also clustered into two subtypes, including 42 patients (43.75%) and 54 patients (56.25%) in Subtypes 1 and 2, respectively. The patient proportion in Subtype 1 was not statistically different from that in primary cluster analysis, suggesting a considerable agreement of two classifications (**Table S4**).

2.1.2. Illness duration differences. In primary cluster findings, two subtypes of institutionalized patients showed significant illness duration differences (M ± SD: 21.96 ± 8.22 years in Subtype 1 and 18.08 ± 8.92 years in Subtype 2; t = 2.21; P = .029), while no significant differences in illness duration were found in this additional clustering (M ± SD: 19.88 ± 8.45 years in Subtype 1 and 20.13 ± 9.07 in Subtype 2; t = -0.14; P = .890).

2.1.3. Neuroanatomic and cognitive patterns. Neuroanatomic and cognitive patterns were highly similar between the two classifications with (**Figure S3**) or without (**Figure 2 in the main text**) prior removal of variance related to illness duration before subtyping. Such consistency of brain-behavior abnormalities and findings displayed above suggest that the subtyping and corresponding brain-behavior comparisons in institutionalized patients with schizophrenia were not driven by their slightly different illness duration.

2.2 ICV for brain volume comparisons

In our primary clustering findings, the abnormal patterns of subcortical and global brain volumes were the same for the situation that ICV was included as one of the covariates (**Figure 2 in the main text**) or not (**Figure S4**) during comparisons.

2.3 Associations between medication or illness duration with brain-behavior measures

Neither significant correlations between antipsychotic dose or illness duration and cognitive function nor significant correlations between antipsychotic dose or illness duration and subcortical volumes were found in both identified subtypes of institutionalized patients (**Table S11 – S12**).

2.4 The importance of using subcortical volumes as subtyping features

Correlations between regional or global brain volumes and cognitive function were conducted with or without the prior removal of variance related to ICV for brain volumes in the whole group of institutionalized patients with schizophrenia (**Table S10**).

Concerning correlations involving regional brain measures, cognitive scores were significantly associated with subcortical volumes instead of any cortical measures. Specifically, volume in the right thalamus was significantly associated with Tower of London test scores (with ICV as the covariate: r = .41, P_{FDR} = .020; without ICV as the covariate: r = .41, P_{FDR} = .019). Volumes in the bilateral nucleus accumbens, the left caudate, and the right hippocampus were marginally correlated with verbal memory test scores when the ICV variance was not removed before analyzing (r ranged from .34 to .37, P_{FDR} = .049).

About correlations involving global brain measures, subcortical GMV showed widespread associations with verbal memory test scores, Tower of London test scores, and BACS composite scores (with ICV as the covariate: r ranged from .26 to .38, P_{FDR} ranged from .004 to .026; without ICV as the covariate: r ranged from .28 to .38, P_{FDR} ranged from .003 to .019). Note the similar ranges with or without ICV covariate.

Thus, cognitive scores had stronger associations with regional or global subcortical volumes than with cortical measures in institutionalized patients, indicating the importance and robustness of utilizing subcortical volumes to subtyping these individuals.

3. Secondary clustering analyses in the institutionalized sample

Based on regional cortical volumes (Model 1), regional cortical and subcortical volumes (Model 2), global cortical volumes (Model 3), and global brain volumes (Model 4), institutionalized patients with schizophrenia were clustered into 2 to 5 subtypes (**Table S13**).

These subtypes significantly differed in regional and global brain volumes in each of these models. However, no significant between-subtype differences in cognitive function were found in these models (**Table S13**).

4. Classifier training, validation, and patient assignment in community-dwelling samples

4.1 Model optimization

In the training set, the model achieved the highest average accuracy (0.97) when the number of sampled variables at each split was set at 1, which means that 1 of 14 predictive variables randomly sampled at each split achieved the optimal model (**Figure S5**).

4.2 Model performance

4.2.1 In the training set with repeated CV. In training set with the optimized model, the brain-based classifier achieved an average accuracy of 0.97 (SD = 0.07) and an average Cohen's kappa of 0.94 (SD = 0.13) across 100 repetitions of 10-fold CV (**Figure S5**).

4.2.2 In the test set. In the test set, the brain-based classifier performance was tested and reached an accuracy of 1.00 (95% CI ranged from 0.88 and 1.00, $P < .001$), a Cohen's kappa of 1.00, a sensitivity of 1.00, and a specificity of 1.00. The AUC for the ROC curve is 1.00 (**Figure S6**).

4.3 Feature importance

The top 5 features contain volumes in the left putamen, the left thalamus, the left pallidum, the right thalamus, and the right pallidum. While the gaps between different features are subtle (**Figure S7**), consistent with the truth that the optimal random-forest model was achieved as the number of sampled predictors at each split was set at 1 (**Figure S5**).

4.4 Patient assignment in community-dwelling samples

Based on the brain-based classifier developed in the institutionalized sample, patients in each of the community-dwelling samples were classified into two subgroups. The community-dwelling sample with long-term illness was composed of 33 patients (48.53%) in Subgroup 1 and 35 patients (51.47%) in Subgroup 2, and the community-dwelling sample with first-episode illness included 67 patients (53.17%) in Subgroup 1 and 59 patients (46.83%) in Subgroup 2. The classification rate in each community-dwelling sample did not significantly differ from that of the institutionalized sample by clustering (χ^2 ranged from < 0.01 to 0.11, P ranged from .739 to .978) (**Table S4**).

5. Within-sample comparisons and sensitivity analyses in community-dwelling samples

5.1 Neuroanatomic and cognitive comparisons in the B-SNIP sample

Neuroanatomic and cognitive comparisons within the B-SNIP sample (i.e., community-dwelling patients with long-term illness and controls) are demonstrated in **Figure S8**.

5.1.1 Brain volume comparisons. Regrading between-subtype comparisons in brain volumes, Subgroup 1 demonstrated significantly smaller volumes in almost all subcortical measures (*t* ranged from -7.17 to -2.47, P_{FDR} ranged from < .001 to .038) except the right hippocampus volume, and significantly smaller subcortical GMV (*t* = -7.15, P_{FDR} < 0.001), cerebral WMV (*t* = -3.26, P_{FDR} = .006), and TBV (*t* = -3.01, P_{FDR} = .009) than Subgroup 2. No significant differences in regional cortical volumes were found between two subtypes.

In terms of patient-control differences in brain volumes, Subgroup 1 did not significantly differ from controls in regional cortical and subcortical volumes but showed significantly smaller subcortical GMV (Δ = -0.51, P_{FDR} = .042), and Subgroup 2 showed significantly decreased volumes in the right superior frontal gyrus (Δ = -0.83, P_{FDR} < .001) but increased volumes in the bilateral basal ganglia (Δ ranged from 0.71 to 1.47, P_{FDR} ranged from < .001 to .002) and thalamus (Δ ranged from 0.52 to 0.66, P_{FDR} ranged from .006 to .036) and increased subcortical GMV (Δ = 0.97, P_{FDR} < .001) and cerebral WMV (Δ = 0.61, P_{FDR} = .028) relative to controls.

5.1.2 Cognitive comparisons. Although both subgroups displayed significant cognitive deficits with medium-to-large effect sizes relative to controls, no significant between-subtype cognitive differences were found (*t* ranged from -0.71 to 0.79, P_{FDR} = .999). Specifically, Subgroup 1 displayed significant cognitive deficits in BACS composite scores (Δ = -0.67, P_{FDR} = .005) and in symbol coding test scores (Δ = -0.86, P_{FDR} < .001), token motor test scores (Δ = -0.65, P_{FDR} = .010), and verbal fluency test scores (Δ = -0.53, P_{FDR} = .014) relative to controls. Compared with controls, Subgroup 2 also showed significant cognitive deficits, involving BACS composite scores (Δ = -0.68, P_{FDR} = .002) and symbol coding test scores (Δ = -0.71, P_{FDR} = .002), token motor test scores (Δ = -0.62, P_{FDR} = .008), and verbal memory test scores (Δ = -0.55, P_{FDR} = .018).

5.2 Neuroanatomic comparisons in the FES sample

Brain volume comparisons in the FES sample (i.e., community-dwelling patients with first-episode illness and controls) are demonstrated in **Figure S9**.

5.2.1 Between-subtype differences. In terms of between-subtype comparisons, Subgroup 1 displayed significantly smaller volumes than Subgroup 2, involving the bilateral rostral middle frontal gyrus (*t* ranged from -4.99 to -4.50, P_{FDR} < .001), almost all included subcortical measures (*t* ranged from -6.52 to -2.64, P_{FDR} ranged from < .001 to .024) except the left amygdala volume, and almost all included global measures (*t* ranged from -8.03 to -3.90, P_{FDR} < .001) except cerebral WMV.

5.2.2 Patient-control differences. With regards to patient-control differences in regional brain volumes, Subgroup 1 displayed significantly decreased volumes in the bilateral thalamus (Δ ranged from -0.68 to -0.63, P_{FDR} < .001) and pallidum (Δ ranged from -0.36 to -0.35, P_{FDR} = .048), and the right nucleus accumbens (Δ = -0.40, P_{FDR} = .048) but did not show any significant regional cortical volume alterations relative to controls. Subgroup 2 demonstrated significant increased volumes in the bilateral rostral middle frontal gyrus (Δ ranged from 0.53 to 0.55, P_{FDR} = .006), basal ganglia (Δ ranged from 0.32 to 0.84, P_{FDR} ranged from < .001 to .047), thalamus (Δ ranged from 0.41 to 0.44, P_{FDR} = .020), and hippocampus (Δ ranged from 0.33 to 0.39, P_{FDR} ranged from .020 to .039) compared with controls.

For patient-control differences in global brain volumes, Subgroup 1 displayed significant decreased subcortical GMV relative to controls (Δ = -0.45, P_{FDR} = .009), and

Subgroup 2 showed significant increased subcortical GMV (Δ = 0.76, P_{FDR} < .001), total GMV (Δ = 0.52, P_{FDR} = .011), TBV (Δ = 0.47, P_{FDR} = .019), and cortical GMV (Δ = 0.41, P_{FDR} = .044) compared with controls.

5.3 Sensitivity analyses of site

The removal of the site variance before subtyping did not affect neuroanatomic and cognitive patterns (**Figure S10**) or the classification rates (**Table S4**) much in identified subtypes for the B-SNIP sample. Moreover, including site as the additional covariate for regional and global brain volume comparisons did not affect abnormal neuroanatomic patterns much in two subtypes of B-SNIP patients (**Figure S11**).

5.4 Sensitivity analyses of race and education level

For these different models of cognitive comparisons in the B-SNIP sample (i.e., community-dwelling patients with long-term illness and demographically matched controls), the term of education level was constantly significant in the ANCOVA main tests, but the race term became marginal when education level was added together (race in Model 3: P = .003; race in Model 4, P = .048; education level in Model 4: P = .002). Moreover, adding race as the covariate did not affect cognitive deficit patterns revealed by two subgroups of patients. The number of cognitive domains decreased when education level was included as a covariate relative to models without this covariate (**Figure S12**). Thus, in the final model displayed in the main text, we included education level as the only covariate for cognitive comparisons in two subgroups of patients and healthy controls.

5.5 Sensitivity analysis of Euler number

The brain volume comparisons with or without the Euler number as one of the covariates between two subgroups of community-dwelling patients with first-episode illness and controls are displayed in **Figure S13** and **Figure S9**, respectively. For all regional subcortical volumes and the most global brain features included in analyses, features with significant between-subtype differences kept constant no matter whether Euler number was included as the covariates (with Euler number as the covariate: t ranged from -6.93 to -2.68, P_{FDR} ranged from < .001 to .021 for subcortical volumes, and t ranged from -8.69 to -4.37, P_{FDR} < .001 for global volumes except cerebral WMV; without Euler number as the covariate: t ranged from -6.52 to -2.64, P_{FDR} ranged from < .001 to .024 for subcortical volumes, and t ranged from -8.03 to -3.90, P_{FDR} < .001 for global volumes except cerebral WMV). Thus, the significant differences in subcortical and global brain volumes between the two subgroups of FES patients were not driven by the slightly different quality in cortical reconstruction.

6. Brain-behavior associations across institutionalized and community-dwelling patients

In the pooled group composed of institutionalized and B-SNIP patients, we carried out correlation analyses between regional and global brain volumes and cognitive function to explore which pairs of features could better represent brain-behavior severity from the dimensional perspective.

In the pooled group, only the correlation between subcortical GMV and general cognitive function measured by BACS composite scores survived in FDR corrections (r = .23, P_{FDR} = .035) (see **Figure S14A).** In the correlation analysis further conducted between these two features for each sample, institutionalized patients (r = .27, P = .019), rather than B-SNIP patients (r = .06, P = .660), displayed a significant relationship (see **Figure S14B – S14C**). In **Figure S14B**, individuals from the institutionalized sample are located within -3 to 3 in both subcortical GMV and general cognitive function dimensions; the lower left quadrant is mainly composed of Subtype 1 individuals while the upper right quadrant is almost composed of Subtype 2 individuals.

This finding indicates that institutionalized patients rather than B-SNIP patients drive the significant subcortex-cognition relationship in the pooled group. Brain-behavior profiles in institutionalized patients could be characterized in both categorical and dimensional perspectives. That is, institutionalized patients could be classified into two subtypes with distinct subcortex-cognition abnormal patterns, and individuals with more significant subcortical GMV had a better general cognitive function.

Supplemental Figures

Part 1. Primary clustering analysis in the institutionalized sample

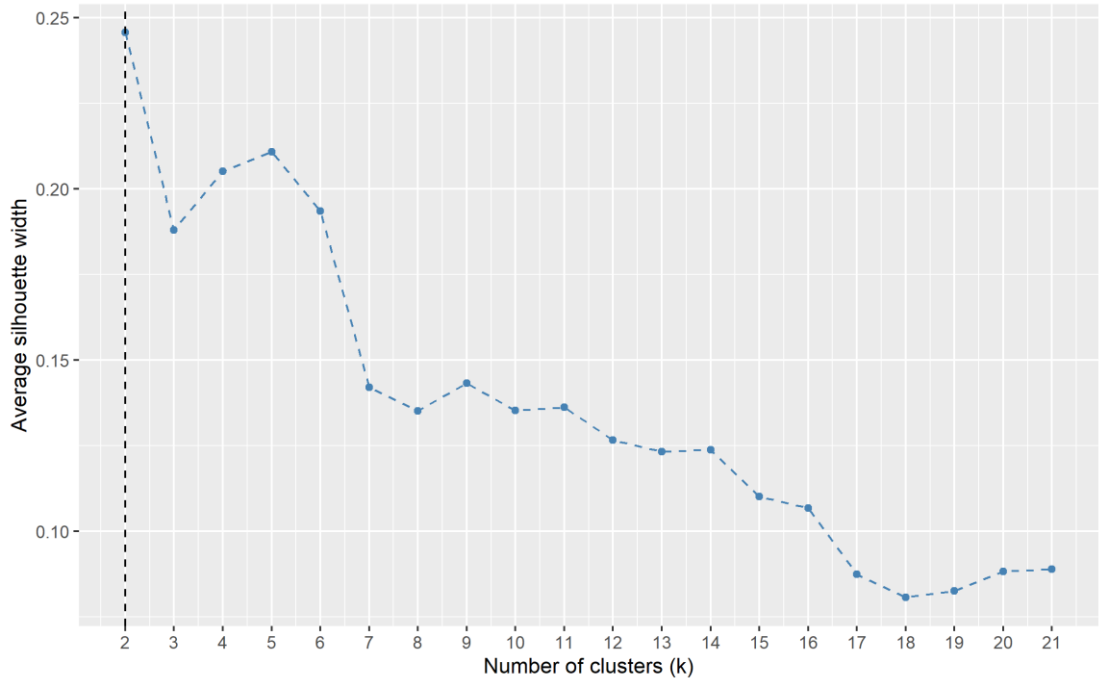

Figure S1. Identification of the optimal number of clusters in K-means++ clustering analysis based on subcortical volumes in institutionalized patients with schizophrenia

K-mean++ cluster analysis was used to stratify institutionalized patients with schizophrenia based on subcortical volumes from 14 regions. We employed the silhouette coefficient (i.e., the largest value of the average silhouette width of the entire data set) to identify the optimal number of clusters with k ranging from 2 to 21, which was achieved in the 2-cluster solution ($k = 2$).

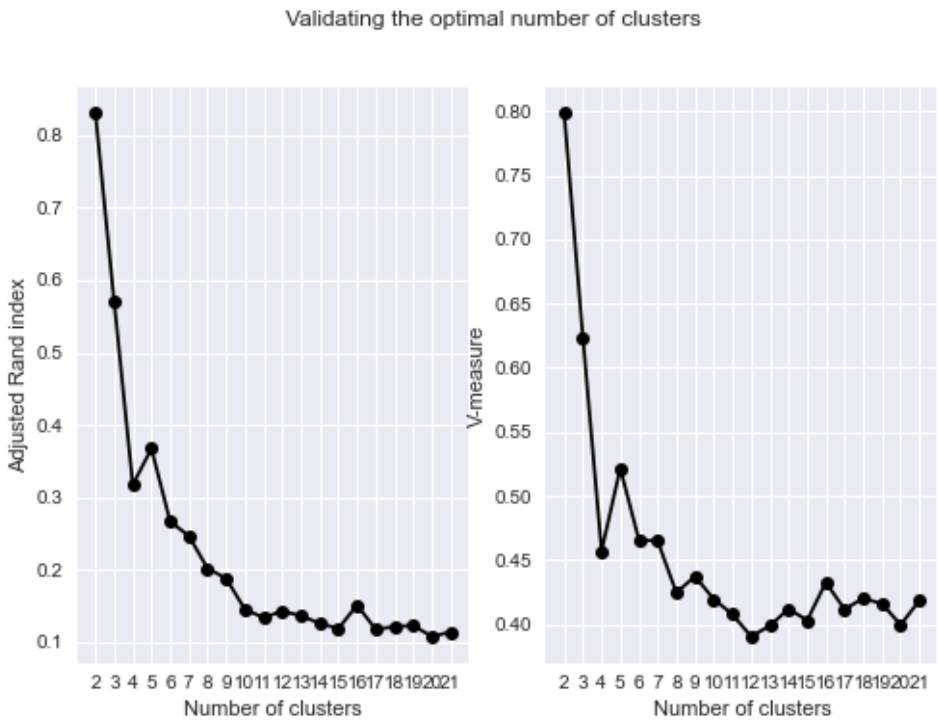

Figure S2. Clustering validation based on the ARI and V-measure for the primary cluster analysis of subcortical volumes in institutionalized patients with schizophrenia

K-means++ cluster analysis was performed in institutionalized patients (N = 96) using 14 subcortical features with prior removal of variance related to age, sex, and ICV, and the optimal number of clusters ($k = 2$) were identified based on the silhouette coefficient. To validate our primary findings, the ARI and V-measure were calculated in 5-fold cross-validation with k ranging from 2 to 21. The two validating metrics both achieved their highest values when the number of clusters was set at 2 ($k = 2$), highlighting the robustness of the clustering in the institutionalized sample.

Part 2. Sensitivity analyses for the primary cluster analysis and brain volume comparisons in the institutionalized sample

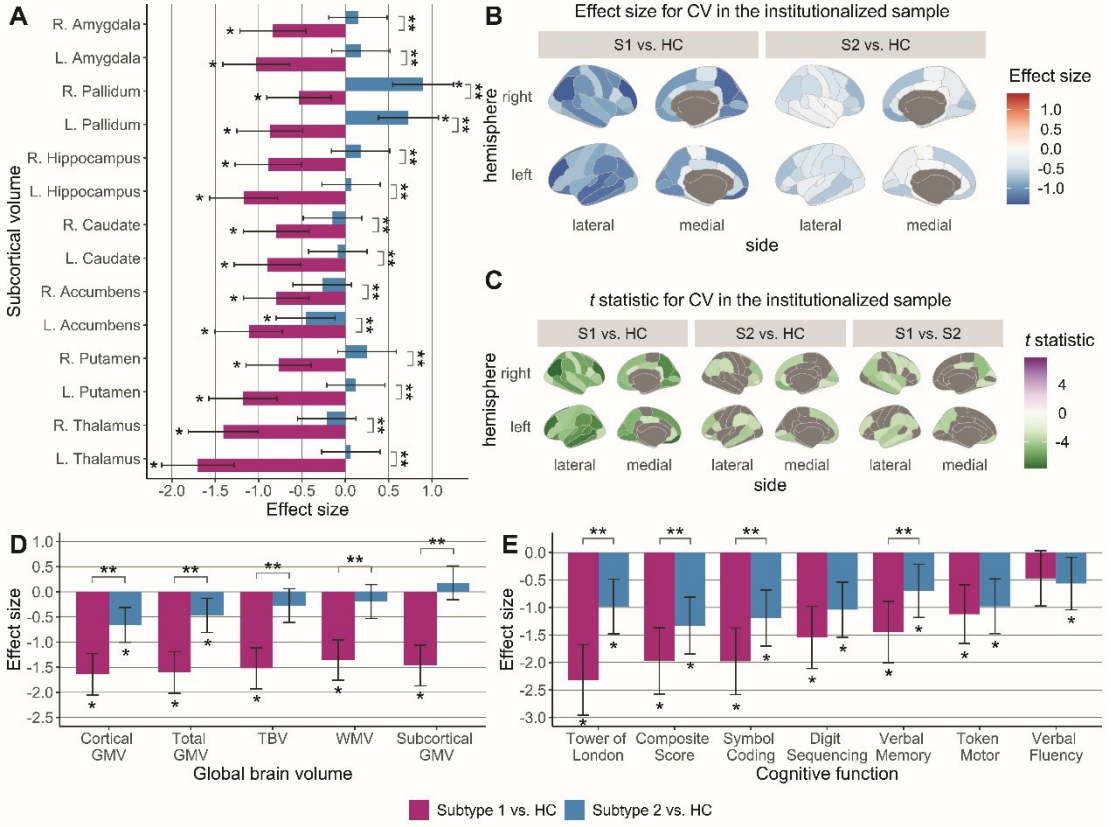

Figure S3. Sensitivity analysis of illness duration for subtyping with subcortical volumes in institutionalized patients with schizophrenia

In this sensitivity analysis, the cluster analysis was performed in institutionalized patients using subcortical volumes, with the additional removal of the variance related to illness duration. Institutionalized patients were also clustered into two subtypes. ANCOVA and post hoc Tukey HSD tests were subsequently used to detect between-group differences in (**A**) subcortical volumes, (**B – C**) cortical volumes, (**D**) global brain volumes, and (**E**) cognitive function in two subtypes of institutionalized patients and demographically matched healthy controls. Age, sex, and ICV were included as covariates for brain volumes, and age, sex, and education level were covariates for BACS raw scores. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (Δ) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of Δ. In cortical maps, only regions that survived FDR corrections are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of institutionalized patients; S2, Subtype 2 of institutionalized patients; TBV, total brain volume; WMV, white matter volume.

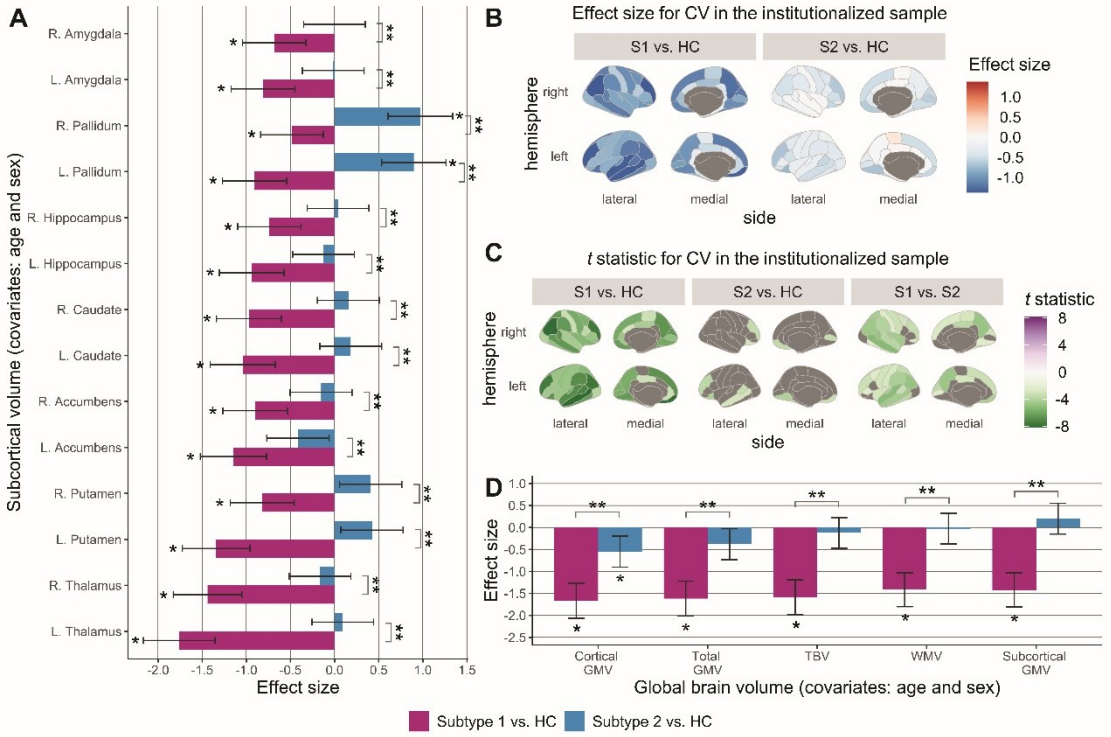

Figure S4. Sensitivity analysis of ICV for brain volume comparisons within the institutionalized sample

Institutionalized patients were classified into two subtypes based on the primary cluster analysis. In this sensitivity analysis, brain volume comparisons were conducted using ANCOVA and Tukey's HSD tests between two subtypes of institutionalized patients and demographically matched healthy controls without ICV as the covariate, including comparisons in (**A**) subcortical volumes, (**B**) cortical volumes, and (**C**) global brain volumes. Age, sex, and ICV were included as covariates for brain volumes. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (*Δ*) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of *Δ*. In cortical maps, only regions that survived FDR corrections are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of institutionalized patients; S2, Subtype 2 of institutionalized patients; TBV, total brain volume; WMV, white matter volume.

Part 3. Classifier developed in the institutionalized sample

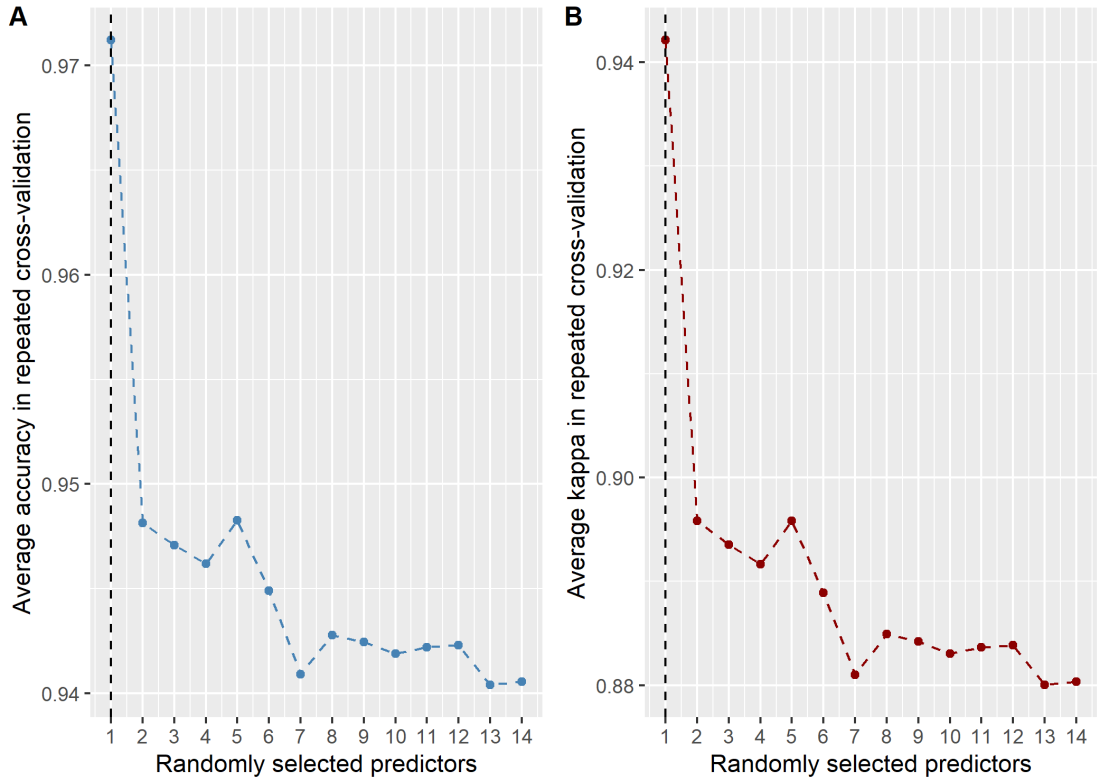

Figure S5. The average accuracy and Cohen's kappa of the brain-based classifier across 100 repetitions of 10-fold cross-validation (CV) with tuning parameters in the training set of institutionalized patients with schizophrenia

The brain-based classifier was trained in the training set (i.e., the 70% of institutionalized patients) using the Random-Forest algorithm with 14 subcortical features. The subtype labels identified in the cluster analysis of the institutionalized sample were the categorical variable to be predicted. To avoid overfitting, model optimization was performed using 100 repetitions of 10-fold CV with tuning parameters. The number of sampled predictors at each split was tuned using a grid search approach to achieve better model performance. (A) The average accuracy and (B) Cohen's kappa was calculated to evaluate model performance. The model with the highest average accuracy was selected as the optimal model. After hyperparameter tuning with CV, the model achieved the highest average accuracy when the number of sampled predictors at each split was set at 1, as the average Cohen's kappa also achieved its highest value.

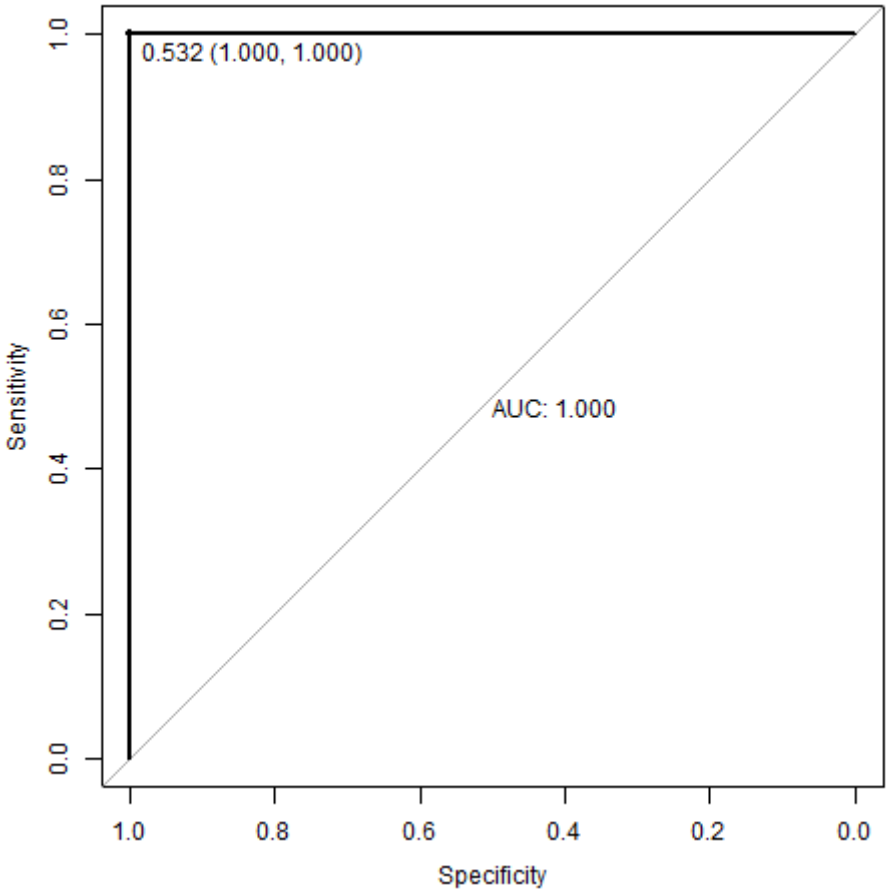

Figure S6. The ROC curve of the optimal brain-based classifier developed in institutionalized patients with schizophrenia

The prediction performance of the optimal model was evaluated by comparing the predicted results with the clustering results in the test set (i.e., the 30% of institutionalized patients not included in the training set). The AUC is 1.00, and the optimized point had a threshold of 0.53, and corresponding sensitivity of 1.00, and a specificity of 1.00.

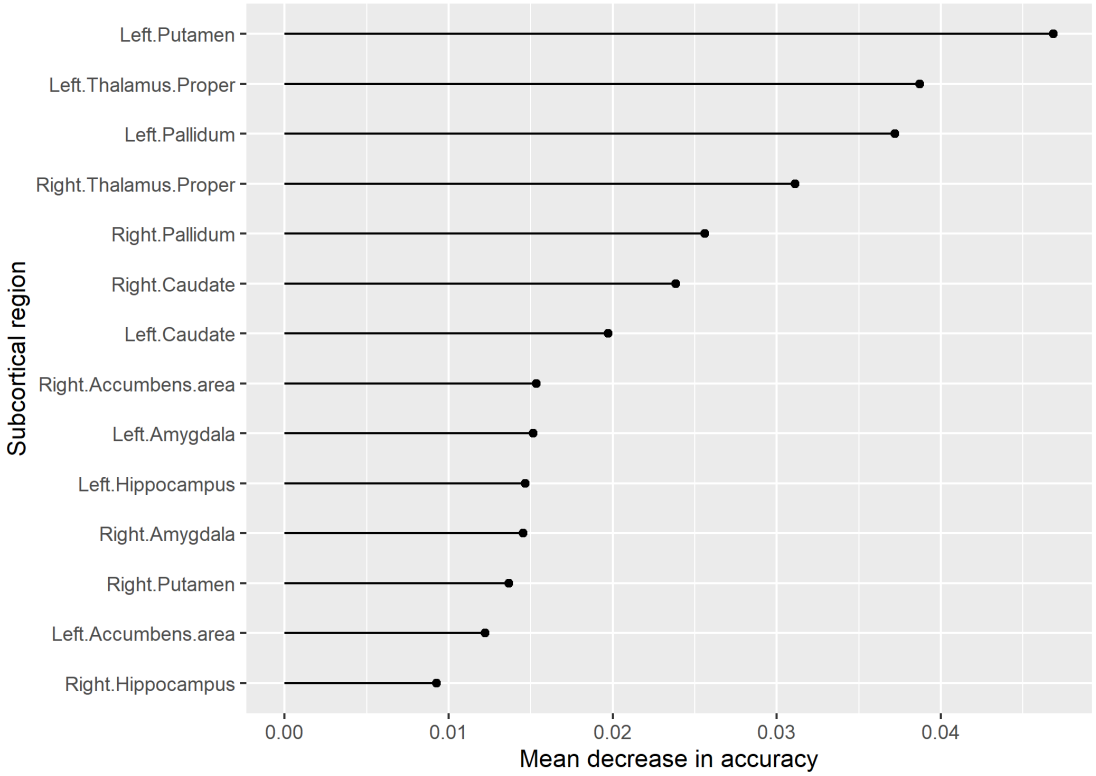

Figure S7. Feature importance measured by the mean decrease in accuracy based on permutation strategy in the optimal brain-based classifier developed in institutionalized patients with schizophrenia

For each tree, the baseline accuracy is recorded by passing the out-of-bag (OBB) data through the random forest model, which is done after every single predictive variable is permuted. The raw mean decrease in accuracy for the particular feature is defined as the difference between the baseline and the drop in averaged accuracy over all trees by permuting the predictive variable.

Part 4. Neuroanatomic and cognitive patterns in community-dwelling patients

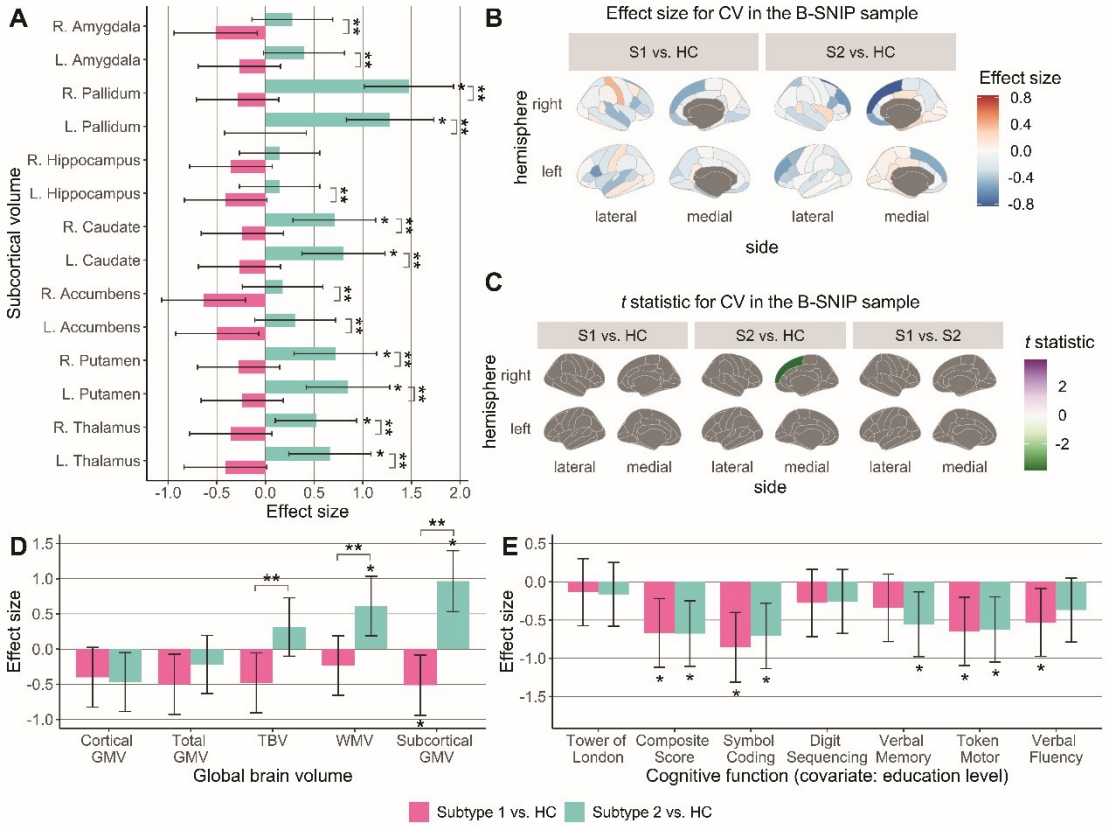

Figure S8. Between-group comparisons in brain-behavior profiles within the B-SNIP sample

In the patient assignment procedure, B-SNIP patients (i.e., community-dwelling patients with long-term schizophrenia) were classified into two subgroups based on the classifier developed in institutionalized patients. ANCOVA and post hoc Tukey HSD tests were used to detect between-group differences in (**A**) subcortical volumes, (**B – C**) cortical volumes, (**D**) global brain volumes, and (**E**) cognitive function in two subtypes of B-SNIP patients and demographically matched healthy controls. Age, sex, and ICV were included as covariates for brain volumes, while education level was the covariate for age- and sex-corrected BACS z-scores. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (Δ) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of Δ. In cortical maps, only regions that survived FDR corrections

are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of B-SNIP patients; S2, Subtype 2 of B-SNIP patients; TBV, total brain volume; WMV, white matter volume.

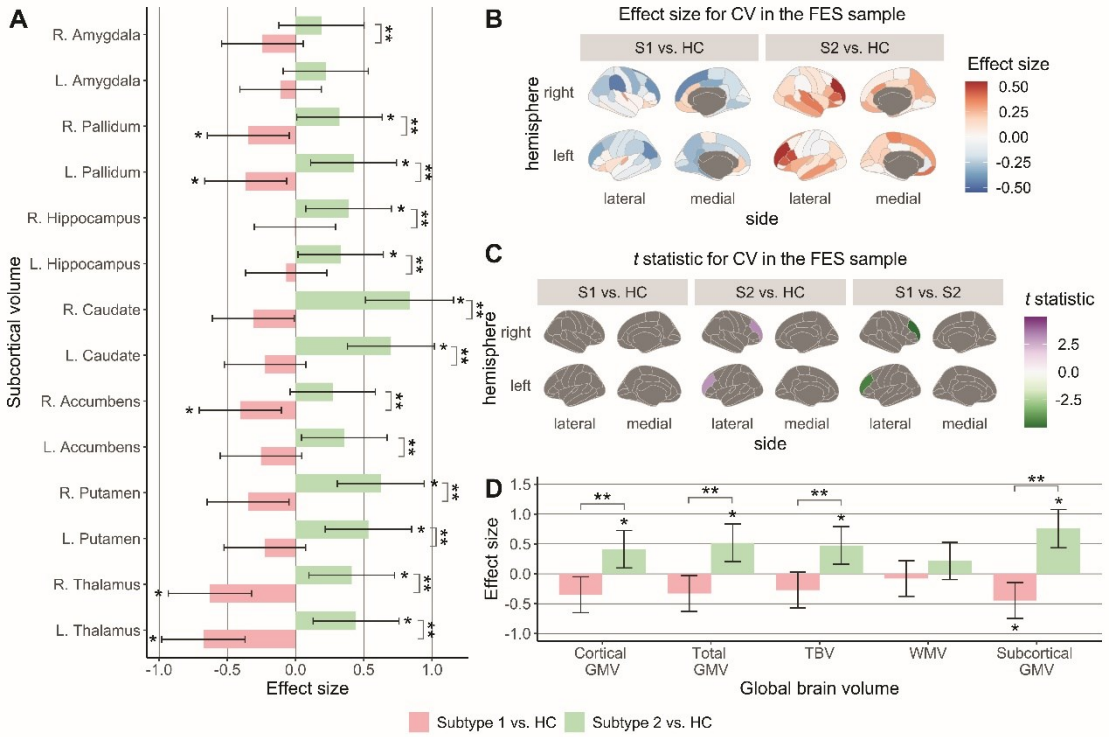

Figure S9. Between-group comparisons in neuroanatomic profiles within the FES sample

In the patient assignment procedure, FES patients (i.e., community-dwelling patients with first-episode schizophrenia) were classified into two subgroups based on the classifier developed in institutionalized patients. ANCOVA and post hoc Tukey HSD tests were used to detect between-group differences in (**A**) subcortical volumes, (**B – C**) cortical volumes and (**D**) global brain volumes in two subtypes of FES patients and demographically matched healthy controls. Age, sex, and ICV were included as covariates for brain volumes. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (*Δ*) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of *Δ*. In cortical maps, only regions that survived FDR corrections are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of FES patients; S2, Subtype 2 of FES patients; TBV, total brain volume; WMV, white matter volume.

Part 5. Sensitivity analyses for the subtyping and comparisons in community-dwelling samples

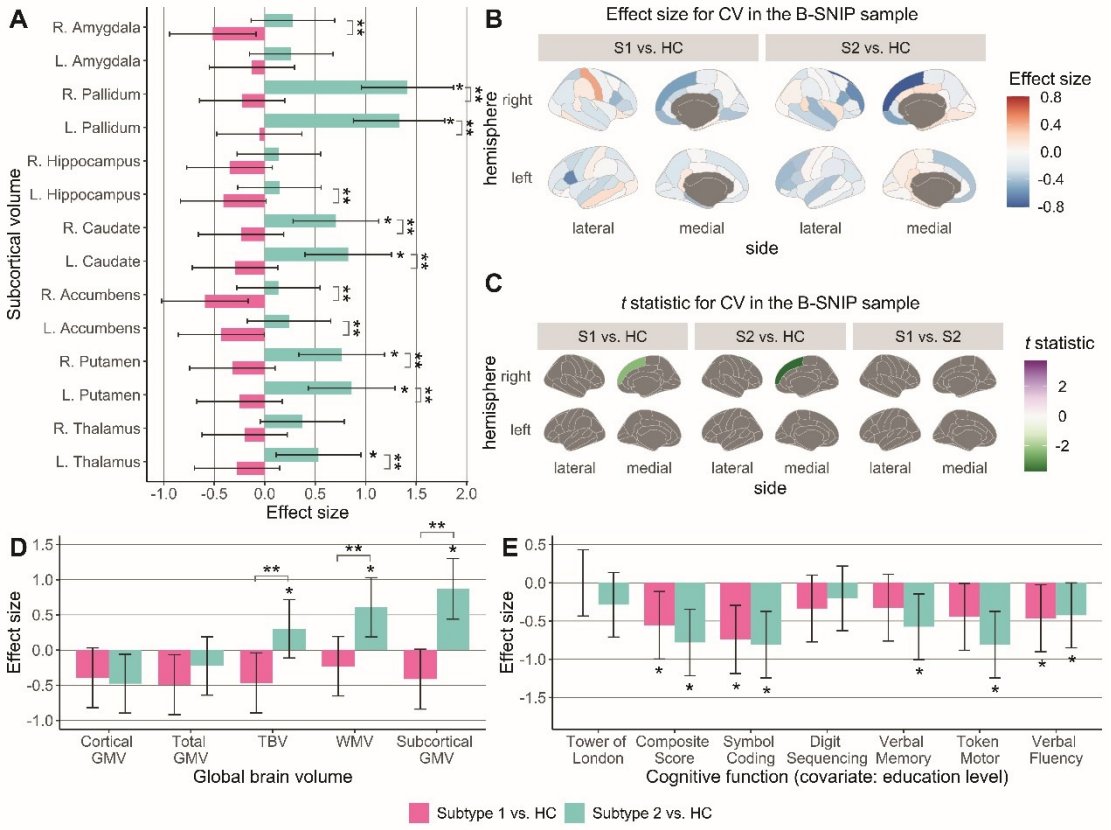

Figure S10. Sensitivity analysis of site for subtyping in B-SNIP patients

In order to evaluate the influence of site for subtyping, the site variance was additionally removed for subcortical volumes in B-SNIP patients (i.e., community-dwelling patients with long-term illness) before patient assignment. These community-dwelling patients were subsequently classified into two subgroups based on the classifier developed in institutionalized patients. ANCOVA and post hoc Tukey HSD tests were used to detect between-group differences in (**A**) subcortical volumes, (**B – C**) cortical volumes, (**D**) global brain volumes, and (**E**) cognitive function in two subtypes of B-SNIP patients and demographically matched healthy controls. Age, sex, and ICV were included as covariates for brain volumes, while education level was the covariate for age- and sex-corrected BACS z-scores. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (Δ) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of Δ. In cortical maps, only regions that survived FDR corrections are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of B-SNIP patients; S2, Subtype 2 of B-SNIP patients; TBV, total brain volume; WMV, white matter volume.

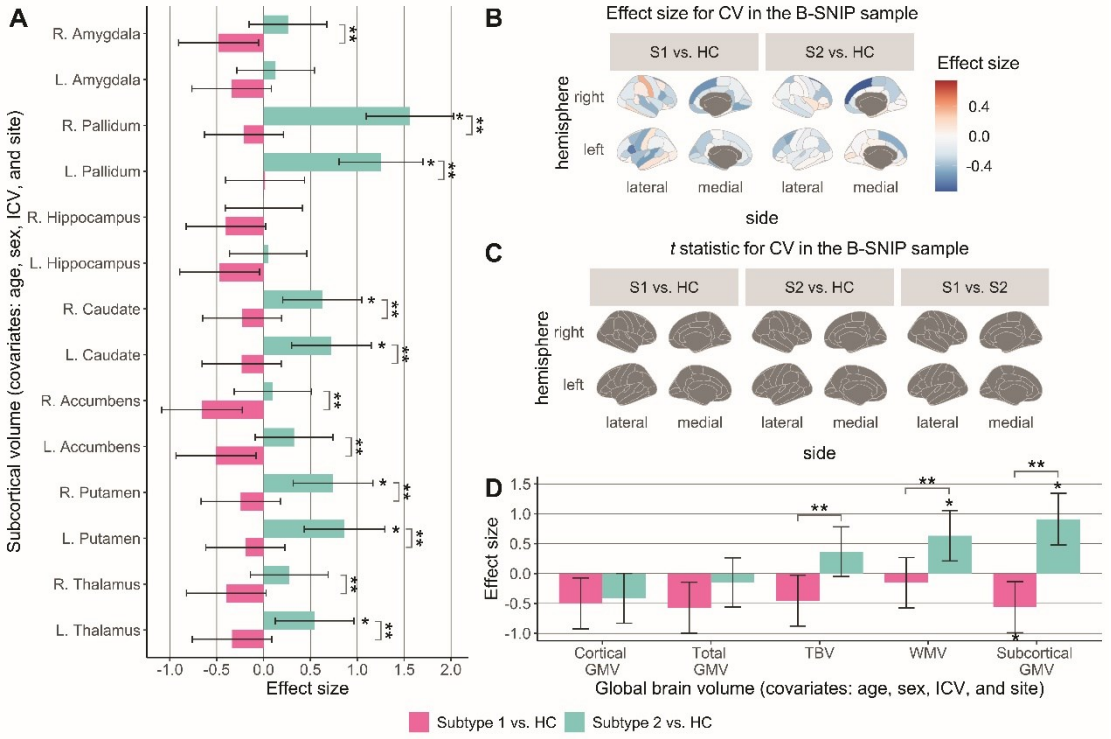

Figure S11. Sensitivity analysis of site for brain volume comparisons within the B-SNIP sample

In order to evaluate the influence of site in detecting between-group differences in brain volumes, ANCOVA and Tukey's HSD tests were conducted between two subtypes of B-SNIP patients (i.e., community-dwelling patients with long-term schizophrenia) and demographically matched healthy controls in (**A**) subcortical volumes, (**B**) cortical volumes, and (**C**) global brain volumes, with the additional inclusion of site as one of the covariates. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (Δ) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of Δ. In cortical maps, only regions that survived FDR corrections are colored by *t* statistics from post hoc tests. CV, cortical volume; GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of B-SNIP patients; S2, Subtype 2 of B-SNIP patients; TBV, total brain volume; WMV, white matter volume.

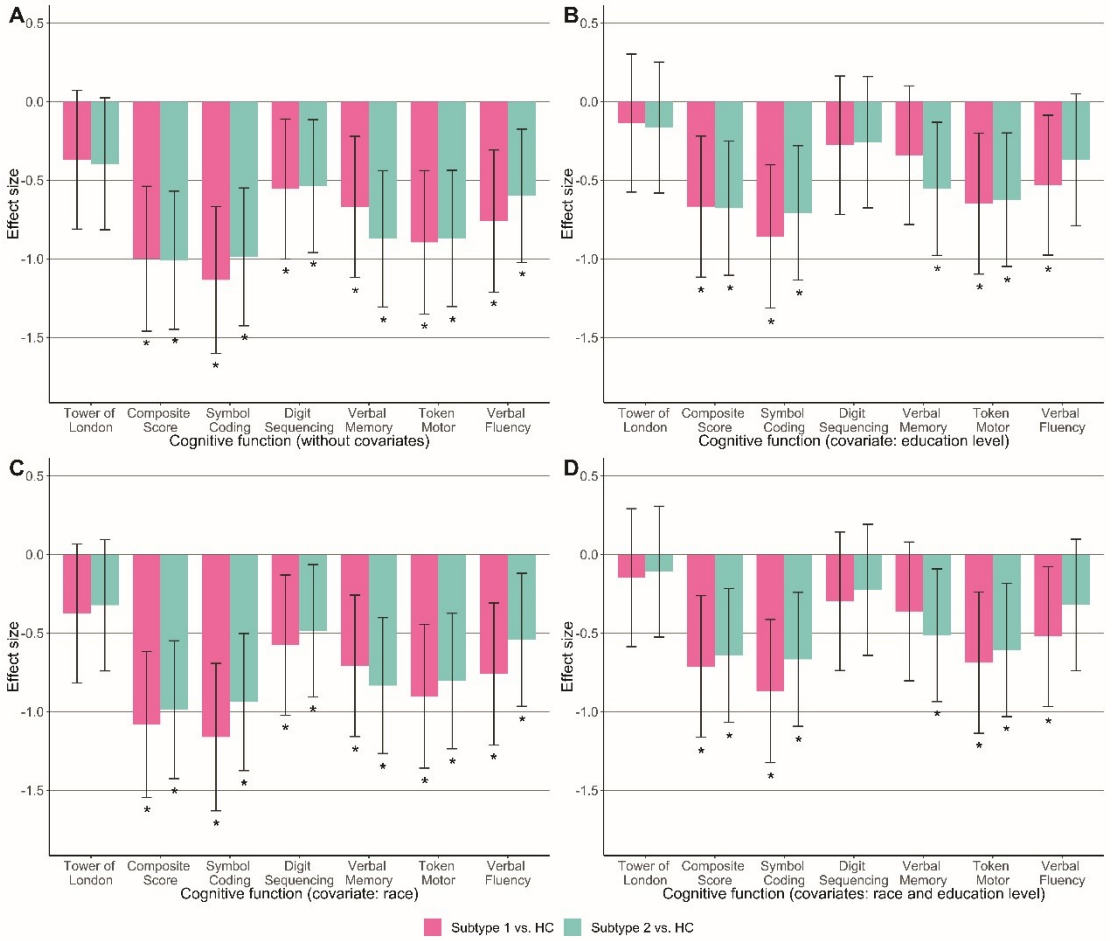

Figure S12. Sensitivity analyses of race and education level for cognitive comparisons within the B-SNIP sample

In order to evaluate the influence of race and education level in detecting between-group differences in cognition, ANCOVA and Tukey's HSD tests were conducted between two subtypes of B-SNIP patients (i.e., community-dwelling patients with long-term schizophrenia) and demographically matched healthy controls with different combinations of covariates, including (**A**) no covariates; (**B**) education level as the covariate; (**C**) race as the covariate; and (**D**) race and education level as covariates. We did not include age and sex as covariates because the cognitive scores of these participants had been corrected for age and sex. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (Δ) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of Δ. HC, healthy controls.

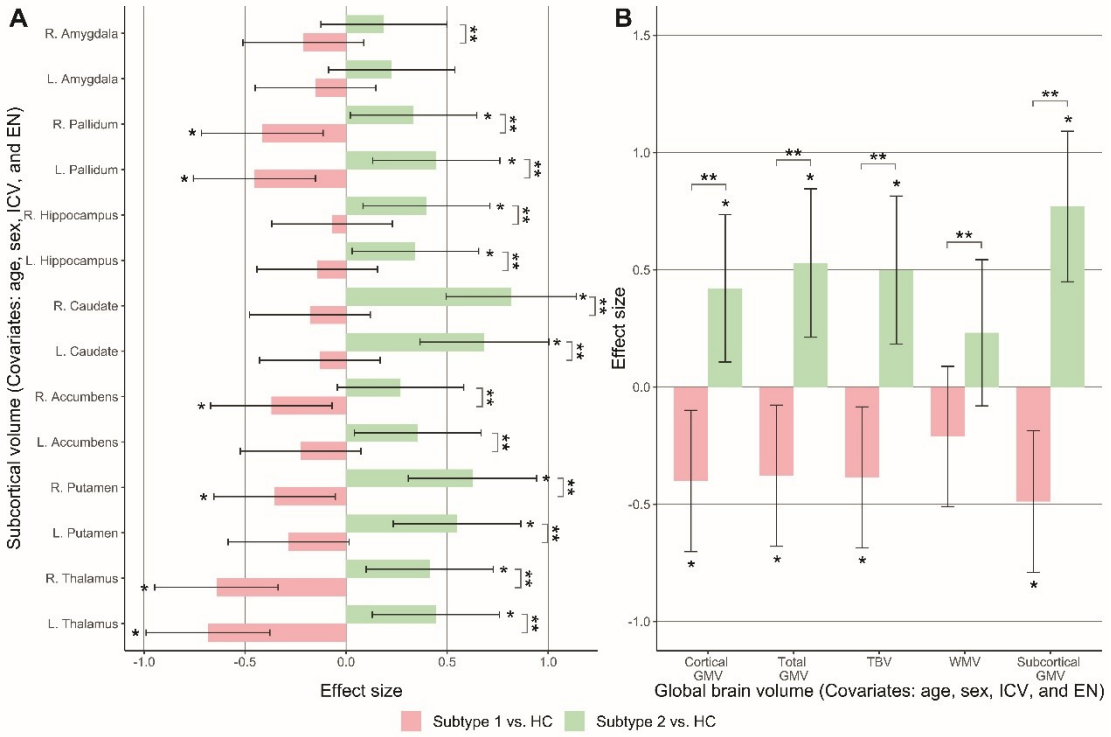

Figure S13. Sensitivity analysis of Euler number for brain volume comparisons within the FES sample

In order to evaluate the influence of Euler number in detecting between-group differences in brain volumes, ANCOVA and Tukey's HSD tests were conducted between two subtypes of FES patients (i.e., community-dwelling patients with first-episode illness) and demographically matched healthy controls in (**A**) subcortical volumes and (**B**) global brain volumes, with the additional inclusion of Euler number as one of the covariates. In bar charts, significant patient-control and between-subtype differences, determined by FDR-corrected *P*-values generated in post hoc pairwise tests, are marked by one and two asterisks, respectively. Shading bars represent Glass's delta (*Δ*) effect sizes, which were calculated after removing variance related to corresponding covariates and used to demonstrate patient-control differences for each subtype. Error bars mean 95% confidence interval of *Δ*. GMV, gray matter volume; HC, healthy controls; L, the left hemisphere; R, the right hemisphere; S1, Subtype 1 of FES patients; S2, Subtype 2 of FES patients; TBV, total brain volume; WMV, white matter volume.

Part 6. Brain-behavior analyses across institutionalized and B-SNIP patients

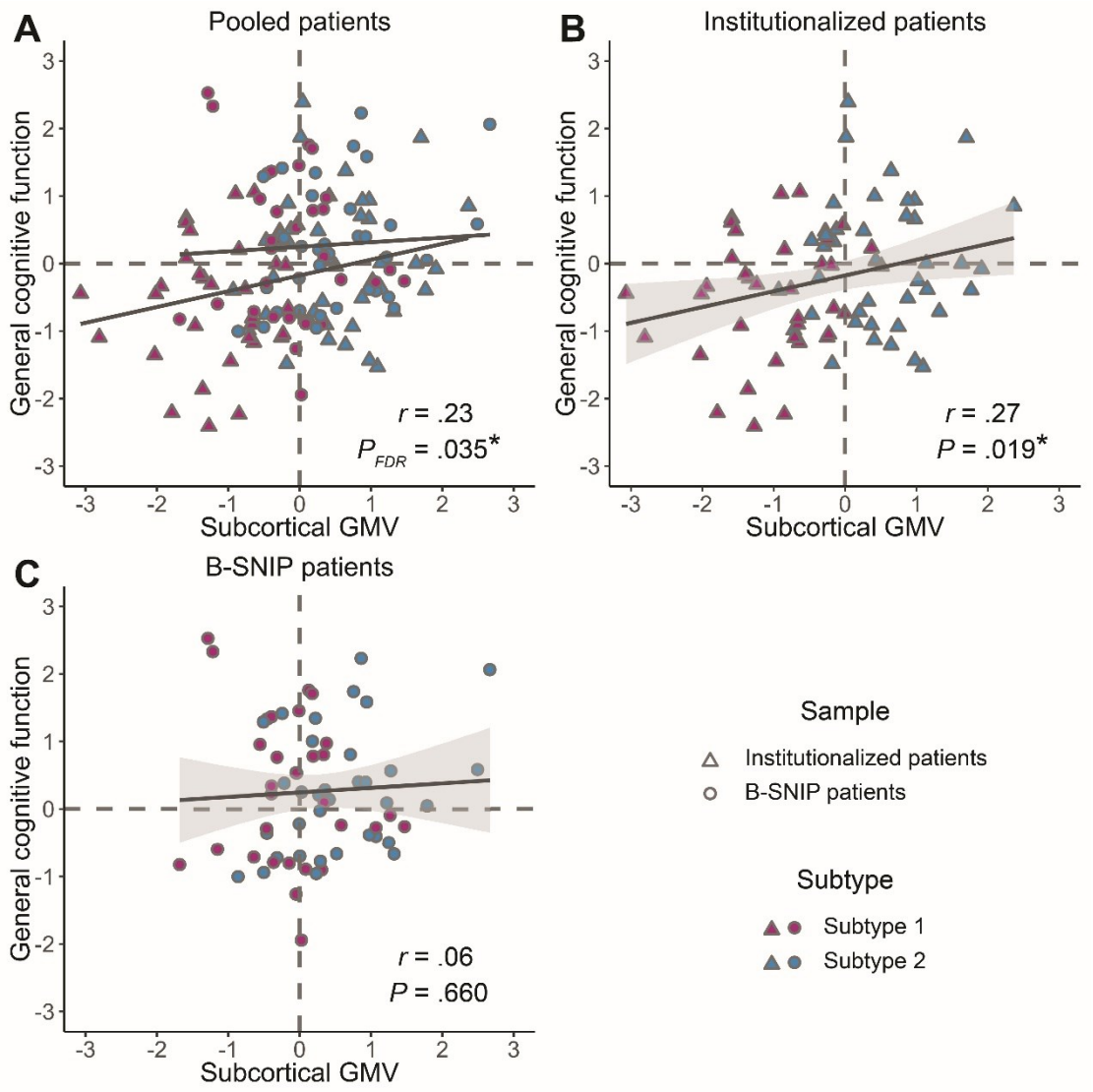

Figure S14. Subcortex-cognition relationships in institutionalized and B-SNIP patients with schizophrenia

We pooled brain-behavior data from institutionalized and B-SNIP patients to better characterize subcortex-cognition relationships across samples. After regressing out covariates and standardizing them into z-scores, we carried out correlation analyses between regional (including volumes from cortical and subcortical regions) or global brain volumes and cognitive function in this pooled group. FDR adjustment was applied on generated *P*-values because of multiple analyses. (**A**) In pooled patients, only the

correlation between subcortical GMV and general cognitive function survived FDR corrections. We further performed correlation analyses between this pair of features in each sample. (**B**) Institutionalized patients, instead of (**C**) B-SNIP patients, showed a significant correlation between subcortical GMV and general cognitive function ($P < .025$; the significance level was set at .05/2 = .025 because two correlations were being conducted at the same time). Solid lines represent regression lines for each sample, and shading bands represent corresponding 95% confidence intervals. Asterisks mean significant subcortex-cognition correlations.

Supplemental Tables

Part 1. Scanning parameters, demographics, and clinical profiles for all included samples

Table S1. Scanning parameters of 3D T1-weighted images for participants in all included samples

Site	TR (ms)	TE (ms)	Flip angle (°)	Slices (N)	Matrix (mm^2)	Voxel size (mm^3)	Scanner vendor
Subtype discovery set (N = 192): institutionalized patients and healthy controls							
West China Hospital	8.5	3.4	12	156	240×240	1×1×1	GE Signa EXCITE
Community-dwelling set 1 (B-SNIP sample, N = 136): community-dwelling patients with long-term illness and healthy controls							
Baltimore	6.80	2.91	9	160	256×240	1×1×1.2	Siemens Trio
Boston	7.0	3.00	8	166	256×256	1×1×1.2	GE Signa HDxt
Chicago	6.99	2.85	8	166	256×256	1×1×1.2	GE Signa HDx
Dallas	6.80	3.10	8	170	256×240	1×1×1.2	Philips Achieva
Detroit	6.80	2.74	8	160	256×240	1×1×1.2	Siemens Trio
Hartford	7.20	2.91	9	160	256×240	1×1×1.2	Siemens Allegra
Community-dwelling set 2 (FES sample, N = 252): community-dwelling drug-naïve patients with first-episode illness and healthy controls							
West China Hospital	8.5	3.4	12	156	240×240	1×1×1	GE Signa EXCITE

TR, repetition time; TE, echo time.

Table S2. Demographical and clinical profiles for participants in all included samples

Measure	Subtype discovery set: institutionalized patients and controls				Community-dwelling set 1 (B-SNIP sample): community-dwelling patients with long-term illness and controls			
	Patient (N = 96)	HC (N = 96)	t/χ^2	P/P_{FDR}	Patient (N = 68)	HC (N = 68)	t/χ^2	P
Age (year, M ± SD)	46.48 ± 7.25	46.65 ± 7.53	-0.16	.876	36.94 ± 10.80	37.29 ± 10.74	-0.19	.849
Sex (female: N/%)	31 (32.29%)	31 (32.29%)	0.00	1.000	31 (45.59%)	31 (45.59%)	0.00	1.000
Education level (year, M ± SD)	9.92 ± 2.66	9.56 ± 3.24	0.82	.411	12.81 ± 2.21	14.76 ± 2.49	-4.84	< .001*
Site (6 sites: N)	-	-	-	-	11/11/13/20/4/9	11/12/5/28/8/4	8.19	.146
Race (CA/AA/others: N)	-	-	-	-	33/32/3	33/29/6	1.48	.563
Illness duration (year, M ± SD)	20.02 ± 8.76				16.56 ± 9.71			
CPZ equivalent (mg/day, M ± SD)	482.67 ± 204.10				462.21 ± 304.60			
PANSS								
Positive score (M ± SD)	10.49 ± 4.34				17.57 ± 4.96			
Negative score (M ± SD)	16.27 ± 5.45				16.04 ± 5.68			
General score (M ± SD)	26.06 ± 5.52				33.30 ± 8.53			
Total score (M ± SD)	52.83 ± 13.06				66.91 ± 16.60			

Measure	Community-dwelling set 2 (FES sample): community-dwelling patients with first-episode illness and controls			
	Patient (N = 126)	HC (N = 126)	t/χ^2	P
Age (year, M ± SD)	22.98 ± 7.40	23.82 ± 7.20	-0.91	.366
Sex (female: N/%)	74 (58.73%)	74 (58.73%)	0.00	1.000
Education level (year, M ± SD)	11.97 ± 3.31	12.90 ± 2.95	-2.37	.019*
Illness duration (month, M ± SD)	6.59 ± 7.85			
PANSS				
Positive score (M ± SD)	25.10 ± 6.11			
Negative score (M ± SD)	17.58 ± 7.23			
General score (M ± SD)	46.18 ± 9.48			
Total score (M ± SD)	88.87 ± 16.07			

AA, African American; CA, Caucasian; CPZ equivalent, the daily dose of antipsychotics transformed into chlorpromazine equivalent; HC, healthy controls; M, mean value; PANSS, the Positive and Negative Syndrome Scale; P_{FDR}, False discovery rate (FDR) adjusted *p*-value; SD, standard deviation; *t*, *t*-statistic in two-sample t-tests; χ^2, chi-squared statistic.

Age and education level were compared using two-sample t-tests, and sex, race, and site distribution were compared using chi-squared tests between patients and healthy controls. Asterisks demonstrate significant between-group differences.

Table S3. Within-sample cognitive comparisons between patients and healthy controls in the institutionalized sample and the community-dwelling sample with long-term illness

BACS score	**Subtype-discovery set: institutionalized patients with long-term illness and controls**				**Community-dwelling set 1 (B-SNIP sample): community-dwelling patients with long-term illness and controls**			
	Patient vs. HC	*F*	P_{FDR}	η^2	Patient vs. HC	*F*	P_{FDR}	η^2
Verbal memory (*Δ* [95%CI])	-1.03 [-1.47, -0.58]	19.43	< .001*	0.16	-0.46 [-0.81, -0.10]	8.14	.007*	0.06
Digit sequencing (*Δ* [95%CI])	-1.26 [-1.72, -0.81]	24.75	< .001*	0.19	-0.27 [-0.61, 0.08]	3.29	.084	0.02
Token motor (*Δ* [95%CI])	-1.04 [-1.48, -0.60]	36.19	< .001*	0.26	-0.64 [-0.99, -0.28]	14.29	.001*	0.10
Verbal fluency (*Δ* [95%CI])	-0.52 [-0.95, -0.09]	6.74	.011*	0.06	-0.44 [-0.80, -0.09]	9.58	.004*	0.07
Symbol coding (*Δ* [95%CI])	-1.54 [-2.01, -1.07]	54.80	< .001*	0.35	-0.78 [-1.14, -0.42]	24.09	< .001*	0.16
Tower of London (*Δ* [95%CI])	-1.57 [-2.04, -1.10]	16.18	< .001*	0.14	-0.15 [-0.50, 0.20]	0.89	.346	0.01
Composite score (*Δ* [95%CI])	-1.61 [-2.09, -1.14]	65.31	< .001*	0.39	-0.67 [-1.03, -0.32]	18.74	< .001*	0.13

BACS, the Brief Assessment of Cognitive in Schizophrenia; *Δ*, Glass's delta; HC, healthy controls; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; 95% CI, 95% confidence interval.

Analysis of covariance (ANCOVA) was used to test cognitive differences. For participants from the subtype-discovery set, age, sex, and education level were treated as covariates. For B-SNIP data (i.e., participants from Community-dwelling set 1), education level was included as the covariate, because test scores were already corrected for age and sex. Glass's delta effect sizes for cognitive patient-control differences were calculated after variance related to corresponding covariates were removed within each cohort. Asterisks demonstrate significant patient-control differences.

Part 2. Patient membership and comparisons in all included samples

Table S4. Patient membership identified based on subcortical volumes and z tests of patient proportions

Main subtyping analyses	Subtype-discovery set	Subtype 1	Subtype 2		
	Institutionalized patients with long-term illness duration (N = 96)	48 (50.00%)	48 (50.00%)		
	Community-dwelling set	Subgroup 1	Subgroup 2	χ^2	*P*
	B-SNIP sample: community-dwelling patients with long-term illness duration (N = 68)	33 (48.53%)	35 (51.47%)	< 0.01	.978
	FES sample: community-dwelling drug-naïve patients with first-episode illness (N = 126)	67 (53.17%)	59 (46.83%)	0.11	.739
Sensitivity analysis of illness duration for subtyping	Subtype-discovery set	Subtype 1	Subtype 2	χ^2	*P*
	Institutionalized patients with long-term illness duration (N = 96)	42 (43.75%)	54 (56.25%)	0.52	.470
Sensitivity analysis of site for subtyping	Community-dwelling set	Subgroup 1	Subgroup 2	χ^2	*P*
	B-SNIP sample: community-dwelling patients with long-term illness duration (N = 68)	33 (48.53%)	35 (51.47%)	0.00	1.000

P, p-value; χ^2, chi-squared statistic.

Institutionalized patients were clustered into two subtypes based on subcortical volumes. Patients in each community-dwelling sample were assigned to the identified subtypes using the brain-based classifier developed in institutionalized patients.

The z-test, used to compare population proportions in two samples, was employed to test differences in patient proportions of Subtype 1 between each community-dwelling sample with the institutionalized sample. We also employed the z-test to compare patient proportions of Subtype 1 between primary clustering and corresponding sensitivity analysis for the institutionalized sample or the B-SNIP sample.

Table S5. Demographical and clinical comparisons in two identified subgroups of patients and healthy controls from community-dwelling sets

Community-dwelling set 1 (B-SNIP sample, N = 136): patients with long-term illness and controls					
Measure	Subgroup 1 (N = 33)	Subgroup 2 (N = 35)	HC (N = 68)	$F/\chi^2/t$	P/P_{FDR}
Age (year, M ± SD)	35.91 ± 11.94	37.91 ± 9.68	37.29 ± 10.74	0.31	.733
Sex (female: N/%)	15 (45.46%)	16 (45.71%)	31 (45.59%)	<0.01	≈ 1.000
Educational level (year, M ± SD)	12.82 ± 1.79	12.80 ± 2.58	14.76 ± 2.49	11.65	< .001[a, b]
Site (6 sites: N)	2/6/5/11/4/5	9/5/8/9/0/4	11/12/5/28/8/4	16.51	.086
Race (CA/AA/others: N)	17/13/3	16/19/0	33/29/6	4.23	.376
Illness duration (year, M ± SD)	15.45 ± 9.79	17.60 ± 9.67	-	-0.09	.367
CPZ equivalent (mg/day, M ± SD)	460.65 ± 296.29	463.90 ± 320.42	-	-0.04	.972
PANSS (M ± SD)					
Positive score	16.94 ± 4.59	18.14 ± 5.29	-	-1.00	.321
Negative score	15.25 ± 6.47	16.77 ± 4.83	-	-1.08	.321
General score	32.00 ± 8.62	34.49 ± 8.39	-	-1.19	.321
Total score	64.19 ± 17.45	69.40 ± 15.62	-	-1.28	.321
Community-dwelling set 2 (FES sample, N = 252): patients with first-episode illness and controls					
Measure	Subgroup 1 (N = 67)	Subgroup 2 (N = 59)	HC (N = 126)	$F/\chi^2/t$	P/P_{FDR}
Age (year, M ± SD)	23.45 ± 6.69	22.46 ± 8.15	23.82 ± 7.20	0.70	.498
Sex (female: N/%)	40 (59.70%)	34 (57.63%)	74 (58.73%)	0.06	.973
Educational level (year, M ± SD)	12.24 ± 3.10	11.66 ± 3.53	12.90 ± 2.95	3.34	.037[b]
Illness duration (month, M ± SD)	6.25 ± 7.67	6.97 ± 8.10	-	-0.51	.613
PANSS (M ± SD)					
Positive score	24.29 ± 6.31	26.02 ± 5.79	-	-1.60	.150
Negative score	17.65 ± 7.06	17.51 ± 7.48	-	0.11	.913
General score	44.20 ± 8.75	48.41 ± 9.83	-	-2.52	.053
Total score	86.14 ± 14.96	91.93 ± 16.84	-	-2.02	.090

AA, African American; CA, Caucasian; CPZ equivalent, the daily dose of antipsychotics transformed into chlorpromazine equivalent; HC, healthy controls; M, mean value; PANSS, the Positive and Negative Syndrome Scale; P_{FDR}, False discovery rate (FDR) adjusted *p*-value; SD, standard deviation.

Age and education level were compared using analysis of variance, and sex, site, and race distributions were compared using the chi-squared test in two subtypes of patients and healthy controls. Between-subtype comparisons in illness duration, the daily dose of antipsychotics, and PANSS scores were performed with two-sample t-tests. P-values assessed PANSS differences were adjusted by FDR because of multiple comparisons.

[a] Patients in Subgroup 1 showed significantly shorter educated years than healthy controls ($P < .001$).

[b] Patients in Subgroup 2 showed significantly shorter educated years than healthy controls ($P < .001$ in the B-SNIP set; $P < .05$ in the FES set).

Table S6. Comparisons of ICV and Euler number between two identified subgroups and healthy controls in all included samples

Subtyping analysis	Sample	Measure	Subtype 1	Subtype 2	HC	*F*	*P*
Main subtyping analysis	Subtype discovery set: institutionalized patients with long-term illness	ICV	1530268.11 ± 460280.20	1592135.74 ± 301968.30	1510468.58 ± 178373.19	1.20	.302
		Euler number	0.05 ± 0.95	-0.23 ± 1.03	0.09 ± 1.00	1.76	.176
	Community-dwelling set 1 (B-SNIP sample): patients with long-term illness	ICV	1489715.37 ± 181277.16	1459469.39 ± 197268.07	1481013.99 ± 176943.92	0.25	.775
		Euler number	0.00 ± 0.71	0.15 ± 0.97	-0.08 ± 1.13	0.58	.562
	Community dwelling set 2 (FES sample): patients with first-episode illness	ICV	1427691.65 ± 144624.53	1416355.05 ± 153220.80	1460798.42 ± 137270.95	3.94	.021[b]
		Euler number	0.51 ± 0.67	-0.23 ± 1.00	-0.16 ± 1.06	13.17	< .001[a, c]
Sensitivity of illness duration for subtyping	Subtype-discovery set: institutionalized patients with long-term illness	ICV	1565114.46 ± 485739.37	1558158.84 ± 296523.09	1510468.58 ± 178373.19	0.76	.470
		Euler number	0.10 ± 0.96	-0.24 ± 1.01	0.09 ± 1.00	2.14	.121
Sensitivity of site for subtyping	Community-dwelling set 1 (B-SNIP sample): patients with long-term illness	ICV	1493659.28 ± 196881.43	1455750.85 ± 181934.13	1481013.99 ± 176943.92	0.48	.618
		Euler number	0.03 ± 0.71	0.12 ± 0.97	-0.08 ± 1.13	0.47	.625

F, *F* statistic in the analysis of variance (ANOVA) and analysis of covariance (ANCOVA); ICV, intracranial volume;

ICV was compared using ANCOVA with age and sex as covariates. Z-scores of Euler number were compared with ANOVA. Corresponding post hoc tests were conducted with Tukey HSD tests.

[a] Patients in Subgroup 1 were significantly greater than healthy controls ($P < .001$).

[b] Patients in Subgroup 2 were significantly smaller than healthy controls ($P < .05$).

[c] Patients in Subgroup 1 were significantly greater than those in Subgroup 2 ($P < .001$).

Part 3. Analyses based on primary cluster findings in the institutionalized sample

Table S7. Analyses based on primary cluster findings: Subcortical volume comparisons between two identified subtypes of institutionalized patients with schizophrenia and healthy controls

Subcortical region	ANCOVA main test				Post hoc test				
						Left hemisphere		Right hemisphere	
	Hemisphere	*F*	P_{FDR}	η^2	Comparison	*t*	P_{FDR}	*t*	P_{FDR}
Thalamus	Left	48.24	< .001*	0.34	S1 vs HC	-8.86	< .001*	-8.57	< .001*
	Right	39.59	< .001*	0.30	S2 vs HC	1.09	.945	-0.57	.945
					S1 vs S2	-8.56	< .001*	-6.88	< .001*
Caudate	Left	23.66	< .001*	0.20	S1 vs HC	-6.09	< .001*	-5.87	< .001*
	Right	21.37	< .001*	0.19	S2 vs HC	1.02	.945	0.79	.945
					S1 vs S2	-6.12	< .001*	-5.73	< .001*
Putamen	Left	36.34	< .001*	0.28	S1 vs HC	-6.97	< .001*	-4.91	< .001*
	Right	22.01	< .001*	0.19	S2 vs HC	2.34	.147	2.60	.126
					S1 vs S2	-8.02	< .001*	-6.47	< .001*
Pallidum	Left	36.79	< .001*	0.28	S1 vs HC	-4.77	< .001*	-2.70	.020*
	Right	29.63	< .001*	0.24	S2 vs HC	5.16	< .001*	5.92	< .001*
					S1 vs S2	-8.58	< .001*	-7.45	< .001*
Hippocampus	Left	13.75	< .001*	0.13	S1 vs HC	-5.02	< .001*	-4.01	< .001*
	Right	10.46	< .001*	0.10	S2 vs HC	-0.24	.970	0.75	.945
					S1 vs S2	-4.12	< .001*	-4.10	< .001*
Amygdala	Left	11.29	< .001*	0.11	S1 vs HC	-4.30	< .001*	-3.84	< .001*
	Right	9.20	< .001*	0.09	S2 vs HC	0.49	.945	0.53	.945
					S1 vs S2	-4.12	< .001*	-3.77	.001*
Accumbens	Left	25.51	< .001*	0.22	S1 vs HC	-7.14	< .001*	-5.38	< .001*
	Right	15.18	< .001*	0.14	S2 vs HC	-2.34	.147	-0.67	.945

	S1 vs S2	-4.11	< .001*	-4.05	< .001*

Accumbens, nucleus accumbens; ANCOVA, analysis of covariance; η^2, eta squared, an effect size that was used to assess the effect of the independent variable (group) in the analysis of covariance; *F*, *F* statistic in ANCOVA; HC, healthy controls; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; S1, subtype 1 of patients; S2, subtype 2 of patients; *t*, *t* statistic in post hoc tests.

Two subtypes of institutionalized patients were identified based on cluster analysis with subcortical volumes as input. ANCOVA was employed to test between-group differences in regional subcortical volumes in two subtypes of institutionalized patients with schizophrenia and healthy controls, with age, sex, and intracranial volume (ICV) as covariates. Tukey HSD tests were performed for measures to be shown significant differences in ANCOVA main tests. FDR correction was performed on p values generated in both ANCOVA main tests and post hoc tests. Asterisks demonstrate significant differences revealed by main tests of ANCOVA and post hoc tests after FDR correction.

Table S8. Analyses based on primary cluster findings: Global brain volume comparisons between two identified subtypes of institutionalized patients with schizophrenia and healthy controls

Global brain measure	ANCOVA main test			Post hoc test					
	F	P_{FDR}	η^2	Group	Mean	SD	Comparison	*t*	P_{FDR}
Cortical GMV	47.83	< .001*	0.34	S1	407191.36	41095.02	S1 vs HC	-9.78	< .001*
				S2	458675.18	33176.10	S2 vs HC	-3.11	.030*
				HC	479162.22	48235.32	S1 vs S2	-5.72	< .001*
Cerebral WMV	37.74	< .001*	0.29	S1	391093.60	48273.96	S1 vs HC	-8.20	< .001*
				S2	460777.59	42582.46	S2 vs HC	< 0.01	≈ 1.000
				HC	460944.88	54052.74	S1 vs S2	-7.06	< .001*
Subcortical GMV	47.55	< .001*	0.34	S1	45400.46	6106.25	S1 vs HC	-8.47	< .001*
				S2	55019.83	4551.20	S2 vs HC	1.77	.303
				HC	53738.71	5978.45	S1 vs S2	-8.82	< .001*
Total GMV	46.20	< .001*	0.33	S1	534001.79	49307.71	S1 vs HC	-9.54	< .001*
				S2	606789.68	46001.31	S2 vs HC	-2.05	.256
				HC	624257.15	61405.22	S1 vs S2	-6.43	< .001*
TBV	44.04	< .001*	0.32	S1	959114.94	101652.29	S1 vs HC	-9.01	< .001*
				S2	1106095.15	86059.21	S2 vs HC	-0.48	≈ 1.000
				HC	1112558.10	107099.13	S1 vs S2	-7.33	< .001*

ANCOVA, analysis of covariance; η^2, eta squared, an effect size that was used to assess the effect of the independent variable (group) in analysis of covariance; *F*, *F* statistic in ANCOVA; GMV, gray matter volume; HC, healthy controls; ICV, intracranial volume; SD, standard deviation; S1, subtype 1 of patients; S2, subtype 2 of patients; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; *t*, *t* statistic in post hoc tests; TBV, total brain volume; WMV, white matter volume.

Two subtypes of institutionalized patients were identified based on cluster analysis with subcortical volumes as input. ANCOVA was employed to test between-group differences in global brain volumes in two subtypes of institutionalized patients with schizophrenia and healthy controls, with age, sex, and ICV as covariates. Tukey HSD tests were performed for measures to be shown significant differences in ANCOVA main tests. FDR correction was performed on p values generated in both ANCOVA main tests and post hoc tests. Asterisks demonstrate significant differences revealed by main tests of ANCOVA and post hoc tests after FDR correction.

≈, FDR-corrected *P* values are larger enough to close to 1.

Table S9. Analyses based on primary cluster findings: Cognitive comparisons between two identified subtypes of institutionalized patients with schizophrenia and healthy controls

BACS raw score	ANCOVA main test			Post hoc test					
	F	P_{FDR}	η^2	Group	Mean	SD	Comparison	*t*	P_{FDR}
Verbal memory	18.19	< .001*	0.26	S1	21.17	8.36	S1 vs HC	-5.91	< .001*
				S2	29.64	10.18	S2 vs HC	-2.51	.046*
				HC	33.26	9.81	S1 vs S2	3.80	.005*
Digit sequencing	15.29	< .001*	0.23	S1	13.92	5.01	S1 vs HC	-5.52	< .001*
				S2	16.64	5.11	S2 vs HC	-3.55	.003*
				HC	18.87	4.71	S1 vs S2	2.21	.103
Token motor	18.05	< .001*	0.26	S1	51.61	15.16	S1 vs HC	-5.51	< .001*
				S2	55.00	13.06	S2 vs HC	-5.17	< .001*
				HC	68.71	16.46	S1 vs S2	0.43	.933
Verbal fluency	3.40	.037*	0.06	S1	21.14	7.26	S1 vs HC	-2.13	.089
				S2	21.45	6.07	S2 vs HC	-2.47	.046*
				HC	23.71	7.39	S1 vs S2	-0.35	.933
Symbol coding	31.02	< .001*	0.38	S1	23.42	13.39	S1 vs HC	-7.73	< .001*
				S2	31.14	12.15	S2 vs HC	-5.76	< .001*
				HC	40.03	13.81	S1 vs S2	2.25	.103
Tower of London	14.55	< .001*	0.22	S1	12.42	6.25	S1 vs HC	-5.30	< .001*
				S2	16.07	3.68	S2 vs HC	-2.28	.063
				HC	17.71	2.10	S1 vs S2	3.36	.011*
Composite score	38.39	< .001*	0.43	S1	23.94	6.90	S1 vs HC	-8.64	< .001*
				S2	28.33	5.88	S2 vs HC	-6.25	< .001*
				HC	33.72	6.86	S1 vs S2	2.72	.048*

ANCOVA, analysis of covariance; BACS, the Brief Assessment of Cognitive in Schizophrenia; η^2, eta squared, an effect size that was used to assess the effect of the independent variable (group) in analysis of covariance; *F*, *F* statistic in ANCOVA; HC, healthy controls; SD, standard deviation; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; S1, Subtype 2 of patients; S2, Subtype 2 of patients; *t*, *t* statistic in post hoc tests.

Two subtypes of institutionalized patients were identified based on cluster analysis with subcortical volumes as input. BACS was used to measure cognitive function. ANCOVA was employed to test between-group differences in cognitive function in two subtypes of institutionalized patients with schizophrenia and healthy controls, with age, sex, and education level as covariates. Tukey HSD tests were performed for measures to be shown significant differences in ANCOVA main tests. FDR correction was performed on p values generated in both ANCOVA main tests and post hoc tests. Asterisks demonstrate significant differences revealed by main tests of ANCOVA and post hoc tests after FDR correction.

Table S10. Analyses based on primary cluster findings: Correlations between brain volumes and cognitive function in the whole group of institutionalized patients with schizophrenia

Brain volume	**Model 1: Variance related to age, sex, education level, and ICV were removed before analyzing.**						**Model 2: Variance related to age, sex and education level were removed before analyzing.**							
	Verbal memory		Tower of London		Composite score		Verbal memory		Symbol coding		Tower of London		Composite score	
	r	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}
Regional brain volume (68 cortical and 14 subcortical features)														
L. Caudate	.35	.055	.29	.110	.29	.156	**.35**	**.049***	.13	.568	.30	.100	.30	.091
L. Accumbens	.36	.055	.33	.063	.29	.156	**.37**	**.049***	.06	.864	.35	.065	.32	.071
R. Thalamus	.24	.170	**.41**	**.020***	.28	.176	.25	.163	.29	.098	**.41**	**.019***	.32	.071
R. Hippocampus	.35	.055	.04	.891	.30	.156	**.34**	**.049***	.35	.050	.07	.760	.35	.054
R. Accumbens	.34	.055	.19	.424	.23	.225	**.34**	**.049***	.07	.798	.20	.367	.27	.123
Global brain volume (5 global features)														
Cortical GMV	.21	.108	.19	.102	**.23**	**.049***	.22	.094	.18	.110	.20	.079	**.25**	**.026***
Cerebral WMV	.19	.115	**.30**	**.023***	**.28**	**.034***	.20	.099	.23	.057	**.31**	**.016***	**.31**	**.012***
Subcortical GMV	**.38**	**.004***	**.26**	**.026***	**.34**	**.011***	**.37**	**.005***	**.33**	**.018***	**.28**	**.019***	**.38**	**.003***
Total GMV	**.27**	**.043***	**.27**	**.026***	**.26**	**.034***	**.28**	**.038***	.24	.057	**.28**	**.019***	**.30**	**.012***
TBV	.18	.115	**.30**	**.023***	**.25**	**.034***	.19	.099	.26	.057	**.31**	**.016***	**.29**	**.012***

Accumbens, nucleus accumbens; GMV, gray matter volume; ICV, intracranial volume; L, left hemisphere; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; R, right hemisphere; *r*, Pearson's correlation coefficient; TBV, total brain volume; WMV, white matter volume.

Correlation analyses were conducted between regional or global brain volumes and cognitive function in the whole group of institutionalized patients with schizophrenia. These analyses were conducted with or without ICV as the covariate for brain volumes to investigate the influence of ICV. For brain volumes, age, sex, and (or) ICV were treated as covariates, while age, sex and education level were included as covariates for cognitive function. Variance related to corresponding covariates were removed before correlation analyses were conducted. FDR correction was employed among regional or global brain features per cognitive score. Asterisks and bold text demonstrate significant correlations between brain volumes and cognition. Only features involved in significant associations are displayed due to the space restriction.

Table S11. Analyses based on primary cluster findings: Correlation analyses between clinical measures and cognitive function in two identified subtypes of institutionalized patients with schizophrenia

Clinical measure	BACS score	S1		S2	
		r	P_{FDR}	*r*	P_{FDR}
Illness duration	Verbal memory	-.37	0.192	-.15	.712
	Digit sequencing	-.08	.745	-.04	.818
	Token motor	-.03	.875	-.13	.712
	Verbal fluency	-.17	.657	.19	.712
	Symbol coding	-.11	.726	.09	.712
	Tower of London	-.15	.657	-.40	.058
	Composite score	-.21	.657	-.08	.712
The daily dose of antipsychotics	Verbal memory	-.25	.346	.08	.993
	Digit sequencing	.09	.786	.22	.565
	Token motor	-.44	.056	< .01	.993
	Verbal fluency	-.05	.786	-.29	.421
	Symbol coding	.05	.786	-.01	.993
	Tower of London	-.15	.684	.18	.616
	Composite score	-.26	.346	.03	.993

BACS, the Brief Assessment of Cognitive in Schizophrenia; S1, Subtype 1 of patients; S2, Subtype 2 of patients; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; *r*, Pearson's correlation coefficient.

Variance related to age, sex, and education level was removed before correlation analyses were performed. FDR correction was employed among BACS scores per clinical measure. The significant level of FDR-corrected *p* values was set at two-tailed .025 (.5/2) due to correlation analyses separately performed in two subtypes. No significant correlations between cognitive function and illness duration or the daily dose of antipsychotics were found in each subtype of institutionalized patients with schizophrenia.

Table S12. Analyses based on primary cluster findings: Correlation analyses between clinical measures and subcortical volumes in two identified subtypes of institutionalized patients with schizophrenia

Region	Illness duration				The daily dose of antipsychotics			
	S1		S2		S1		S2	
	r	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}	*r*	P_{FDR}
L. Thalamus	-.03	.828	-.26	.277	-.17	.577	.05	.824
R. Thalamus	-.12	.600	-.15	.456	-.09	.791	-.04	.824
L. Caudate	-.19	.395	-.28	.277	-.17	.577	.09	.816
R. Caudate	-.14	.529	-.26	.277	-.23	.530	.04	.824
L. Putamen	.09	.663	.01	.951	-.17	.577	.11	.816
R. Putamen	.08	.674	-.02	.951	-.24	.530	.15	.816
L. Pallidum	.16	.470	-.20	.385	-.09	.791	.26	.789
R. Pallidum	.05	.809	-.26	.277	-.10	.791	.23	.789
L. Hippocampus	-.24	.277	-.14	.456	.00	.994	-.07	.824
R. Hippocampus	-.28	.269	-.17	.445	-.04	.915	-.03	.824
L. Amygdala	-.23	.277	-.17	.445	-.01	.994	-.17	.816
R. Amygdala	-.23	.277	-.14	.456	.04	.915	-.14	.816
L. Nucleus accumbens	-.29	.269	-.23	.321	-.25	.530	-.10	.816
R. Nucleus accumbens	-.41	.050	-.11	.551	-.14	.703	-.12	.816

L., left hemisphere; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; *r*, Pearson's correlation coefficient; R., right hemisphere; S1, Subtype 1 of patients; S2, Subtype 2 of patients.

Variance related to age, sex, and ICV was removed before correlation analyses were performed. FDR correction was employed among subcortical regions per clinical measure. The significant level of FDR-corrected *p* values was set at two-tailed 0.025 (0.5/2) due to correlation analyses separately performed in two subtypes. No significant correlations between subcortical volumes and illness duration or the daily dose of antipsychotics were found in each subtype of institutionalized patients with schizophrenia.

Part 4. Analyses based on secondary cluster findings in the institutionalized sample

Table S13. Secondary cluster analyses: Patient membership and cognitive comparisons in the institutionalized sample based on other neuroanatomical features

Features used for subtyping	Subtype					Post hoc test of BACS composite score				
	S1	S2	S3	S4	S5	Group	Mean	SD	Comparison	*t*	P_{FDR}
Model 1 (Regional cortical volumes)	34 (35.42%)	61 (63.54%)	1 (1.04%)	-	-	S1	24.26	7.50	S1 vs. HC	-7.18	< .001*
						S2	27.20	6.17	S2 vs. HC	-7.30	< .001*
						S3	30.67	-	**S1 vs. S2**	**-1.24**	**.977**
						HC	33.72	6.86			
Model 2 (Regional cortical and subcortical volumes)	40 (41.67%)	55 (57.29%)	1 (1.04%)	-	-	S1	24.39	7.74	S1 vs. HC	-7.21	< .001*
						S2	27.31	5.89	S2 vs. HC	-7.21	< .001*
						S3	30.67	-	**S1 vs. S2**	**-1.01**	**.948**
						HC	33.72	6.86			
Model 3 (Global cortical volumes)	5 (5.21%)	18 (18.75%)	26 (27.08%)	42 (43.75%)	5 (5.21%)	S1	21.37	6.78	S2 vs. HC	-4.30	.002*
						S2	28.99	6.32	S3 vs. HC	-6.12	< .001*
						S3	24.59	7.65	S4 vs. HC	-7.65	< .001*
						S4	26.55	6.08	**S2 vs. S3**	**1.75**	**.494**
						S5	27.17	5.49	**S2 vs. S4**	**2.33**	**.309**
						HC	33.72	6.86	**S3 vs. S4**	**0.33**	**.997**
Model 4 (Global brain volumes)	56 (58.33%)	40 (41.67%)	-	-	-	S1	25.02	7.03	S1 vs. HC	-8.08	< .001*
						S2	27.80	6.04	S2 vs. HC	-6.37	< .001*
						HC	33.72	6.86	**S1 vs. S2**	**-1.68**	**.509**

BACS, the Brief Assessment of Cognitive in Schizophrenia; HC, healthy controls; Mean, mean value; P_{FDR}, false discovery rate (FDR) adjusted *p*-value; SD, standard deviation; S1 ~ S5, Subtype 1 ~ Subtype 5 of patients; *t*, *t* statistic in post hoc tests.

This series of secondary subtyping analyses identified different subtypes of institutionalized patients with schizophrenia based on features different from the primary clustering. BACS was used to measure cognitive function. Cognitive comparisons were compared after removing the subtypes with small sample sizes. For Model 1 and Model 2, cognitive comparisons were between Subtype 1, Subtype 2, and controls. As to Model 3, cognitive function was compared among Subtype 2, Subtype 3, Subtype 4, and controls. For Model 1 and Model 2, the mean value of the composite score in Subtype 3 is the original value because there is only one person in this subtype. Asterisks represent significant patient-control differences in cognition. Bold text indicates between-subtype comparisons.

Cognitive comparisons were compared ANCOVA was employed to test between-group differences in different subtypes of patients and healthy controls, with age, sex, and education level as covariates. Tukey HSD tests were performed for measures to be shown significant differences in ANCOVA main tests. FDR correction was performed on p values generated in both ANCOVA main tests and post hoc tests. Although significant case-control cognitive differences were found, no significant between-subtype differences in cognition. Only comparisons in BACS composite scores are displayed in this table due to the space restriction. Asterisks demonstrate significant differences revealed by post hoc tests after FDR correction. Bold texts represent between-subtype comparisons in BACS composite scores.

Supplemental References

1 Keefe RS, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr Res. 2004;68(2-3):283-97.

2 Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774-81.

3 Rosen AFG, Roalf DR, Ruparel K, Blake J, Seelaus K, Villa LP, et al. Quantitative assessment of structural image quality. Neuroimage. 2018;169:407-18.

4 Arthur D, Vassilvitskii S, Siam/Acm. k-means plus plus : The Advantages of Careful Seeding. Siam: Philadelphia; 2007.

5 Team RC. (R Foundation for Statistical Computing, 2020).

6 Van Rossum G, Drake FL. Python 3 Reference Manual. CreateSpace: Scotts Valley, CA; 2009.

7 Kaufman L, Rousseeuw PJ. Finding groups in data: An introduction to cluster analysis. New York: Wiley-Interscience; 1990. p. 83 - 87.

8 Hubert L, Arabie P. Comparing partitions. Journal of Classification. 1985;2(1):193-218.

9 Rand WM. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association. 1971;66(336):846-50.

10 Rosenberg A, Hirschberg J. in *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)* 410-20 (2007).

11 Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968-80.

12 Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.

13 Rhys HI. Machine Learning with R, the tidyverse, and mlr. Manning Publications; 2020.

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

15 Kuhn M. (2020).

16 Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage. 2018;167:104-20.

17 Keefe RS, Harvey PD, Goldberg TE, Gold JM, Walker TM, Kennel C, et al. Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). Schizophr Res. 2008;102(1-3):108-15.