HHS Public Access

Author manuscript

*Neuroimage*. Author manuscript; available in PMC 2017 August 17.

Published in final edited form as:

*Neuroimage*. 2016 January 01; 124(Pt B): 1196–1201. doi:10.1016/j.neuroimage.2015.06.030.

# Northwestern University Schizophrenia Data Sharing for SchizConnect: A Longitudinal Dataset for Large-Scale Integration

**Alex Kogan**[1], **Kathryn Alpert**[1], **Jose Luis Ambite**[2,3,4], **Daniel Marcus**[5], and **Lei Wang**[1,6]

[1]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL

[2]Information Sciences Institute, University of Southern California, Marina del Rey, CA

[3]Digital Government Research Center

[4]Department of Computer Science, University of Southern California, Los Angeles, CA

[5]Neuroinformatics Research Group (NRG) at the Washington University School of Medicine, St Louis, MO

[6]Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL

### Keywords

neuroinformatics; schizophrenia data; research datasets; data mediation and integration; data sharing; XNAT

## Introduction

Schizophrenia is a complex disease with heterogeneous clinical, behavioral, cognitive and genetic manifestations, and sharing of datasets is becoming essential in order to test hypotheses that can capture its variability and complexity (Poline, Breeze et al. 2012). One example is the discovery of microRNA137 that succinctly illustrates the importance of data sharing: using computational biology techniques, (Potkin, Macciardi et al. 2010) combined two previously published, separate datasets and discovered microRNA137 as a risk factor for schizophrenia. It should be noted that neither of the two distinct datasets had identified microRNA137. In a later confirmatory report on 51,695 individuals confirming microRNA137, the International Schizophrenia Consortium proclaimed that a new “cause” of schizophrenia had been found (Ripke, Sanders et al. 2011).

Corresponding Author: Alex Kogan, Northwestern University Feinberg School of Medicine, Department of Psychiatry and Behavioral Sciences, 710 N. Lake Shore Dr, Abbott Hall 1322, Chicago, IL 60611, Phone: 773-368-5244, a-kogan@northwestern.edu.

In this paper, we describe the Northwestern University Schizophrenia Data (NUSDAST) (Wang, Kogan et al. 2013) as part of SchizConnect, an NIH-funded neuroimaging resource for large-scale data sharing for schizophrenia research. With 451 subjects, the majority of whom have archived longitudinal data, NUSDAST is one of the largest single-site, single-platform neuroimaging datasets related to schizophrenia, making it a uniquely important resource to share with the research community. NUSDAST will benefit the neuroscience community in many ways. First, scientists will be able to use these data to generate or test new hypotheses related to abnormalities of brain structures and neural networks in individuals with schizophrenia. Second, scientists will be able to rapidly replicate findings produced using their own datasets. Third, the data could be used to test and validate new brain mapping tools.

## What is available?

The data presented in NUSDAST were collected through the support of two NIH-funded grants on schizophrenia: 1) Neuromorphometry in Schizophrenia (R01-MH056584), and 2) Conte Center for the Neuroscience of Mental Disorders (P50 MH071616). Through these projects, our group has collected high-resolution structural MRI datasets from large cohorts of subjects using the same scanner platform and sequence protocols. We have also collected detailed clinical, cognitive and genetic information from these subjects.

### Subjects

NUSDAST includes de-identified data from 451 individuals with schizophrenia, their non-psychotic siblings, comparison subjects and their siblings. Neuroimaging data exist for 368 individuals. Longitudinal neuroimaging data are also available on 171 individuals with schizophrenia (m/f=114/57, age at baseline=33.8±12.5 years) and 170 controls (m/f=86/84, age at baseline=31.4±13.8 years). Within this group of subjects, 18 individuals with schizophrenia and 30 controls returned for a second follow-up (i.e., 3 time points). The average (SD) follow-up interval was 2.19 (0.82) years for individuals with schizophrenia and 2.28 (0.49) years for the controls. De-identification consisted of stripping HIPAA-mandated identifiable information in research data (such as name, initials, and phone numbers, etc.). Procedures to further anonymize imaging data such as defacing were not performed in order to share the same imaging data that we used in our publications so that others can replicate our findings using their own algorithms if they so desire. See Table 1 below for baseline information.

Clinical data includes information based on specific criteria for clinical stability (Rastogi-Cruz and Csernansky 1997) and clinical rating scales such as the Scale for the Assessment of Positive Symptoms (Andreasen 1984) and Scale for the Assessment of Negative Symptoms (Andreasen 1983) (see Table 1 below for baseline information). Domains of psychopathology (i.e., psychotic symptoms, disorganized symptoms, and negative symptoms) (Andreasen, Arndt et al. 1995) based on raw scales are also included. The reliability and practicality of using these scales in large populations of schizophrenic patients have been demonstrated by Andreasen, et al (1995). Symptom assessments were performed by personnel specially-trained for this purpose. Inter-rater reliability was monitored

regularly for all rating scales, and rater training sessions, including the conjoint assessment of difficult cases, were held weekly. In these sessions, a variety of patients were interviewed in a group. Two established raters reached a consensus of item scores after the interview was completed, and then this “gold standard” score was compared with the rest of the group. New raters were trained by first participating in a minimum of six of these sessions. They were allowed to participate in ratings only after they had demonstrated satisfactory agreement with trained personnel.

## MRI Data

All MR scans were collected using the same 1.5 T Vision scanner platform (Siemens Medical Systems) at each time point. The Vision scanner had actively shielded gradients and echo-planar capability with very high gradient linearity (<0.4% over a 22-cm diameter spherical volume compared to 2–5% over 22-cm for our other scanners), which yielded anatomical images with virtually no distortion (< 0.4% voxel displacement), critical to analyses of neuroanatomical structures. Using the same scanner provided stable longitudinal MR data throughout the entire period of data collection from 1998 to 2006.

Acquisition of all scans was performed at the Mallinckrodt Institute of Radiology at Washington University School of Medicine, where scanner stability (e.g., frequency, receiver gain, transmitter voltage, SNR) and artifacts were regularly monitored. Phantoms of known size were scanned to confirm image dimensions. Further tests and adjustments (shims, gradient calibrations, EPI switch delays, etc.) were made as needed. During each scan session, a small standardization object (i.e., vitamin-E gelcap) was placed on the left side of the forehead for each subject to clearly indicate laterality in the scans. Each scan session included a high-resolution T1-weighted turbo-FLASH scan (Venkatesan and Haacke 1997), multiple (2–4) MPRAGE scans, and MPRAGE average. Source MR scan data were in Siemens MAGNETOM VISION IMA format and subsequently converted into Analyze format using in-house software. Since Analyze-format images may cause confusion with regard to laterality, even though the abovementioned vitamin-E gelcap information may help verify laterality, all Analyze-format images are being converted into NIFTI format and uploaded. The multiple MPRAGE images for each subject are aligned with the first image and averaged to create a low-noise image volume (Buckner, Head et al. 2004). See Table 2 for detailed scan protocol parameters.

## Neuroimaging Meta-Data

In our template-based brain mapping applications, we have focused on a network of structures previously implicated in the pathophysiology of schizophrenia (Weinberger, Berman et al. 1992; Csernansky and Bardgett 1998; Goldman-Rakic 1999). This network included regions with the prefrontal cortex (e.g., middle frontal gyrus - Brodman area 46) (John, Wang et al. 2006; Harms, Wang et al. 2010), the cingulate gyrus (Qiu, Younes et al. 2007; Wang, Hosakere et al. 2007), and the hippocampus (Wang, Joshi et al. 2001; Csernansky, Wang et al. 2002), the parahippocampalgyrus (Karnik-Henry, Wang et al. 2012), as well as the thalamus (Csernansky, Schindler et al. 2004; Harms, Wang et al. 2007; Smith, Wang et al. 2011) and the basal ganglia (Mamah, Wang et al. 2007; Wang, Mamah et al. 2008), which directly or indirectly link these structures via cortical-subcortical

connections. We have constructed manual segmentation datasets for all these structures, which can be used for the validation of new computational methods. In addition, we have also used FreeSurfer (Desikan, Segonne et al. 2006) to generate cortical surface parcellations and measures of cortical regional volume, thickness and surface area (Cobia, Csernansky et al. 2011).

**Template Data—**The templates for the hippocampus and amygdala were generated using a T1-weighted MR scan collected in a healthy subject (Wang, Mamah et al. 2008). The templates for the thalamus and basal ganglia (caudate nucleus, putamen, nucleus accumbens and globus pallidus) were generated using a seven-time averaged T1-weighted MR scan collected in another healthy subject (Wang, Lee et al. 2007). These segmentations were manually performed using Analyze software in these scans by consensus of experts using atlas guidelines (Duvernoy 1988; Duvernoy 1991; Mai, Assheuer et al. 1997). Surfaces (byu format) of each structure were generated using the marching cubes algorithm (Lorensen and Cline 1987; Claudio and Roberto 1994). The left and right surfaces have corresponding nodes so that analyses of shape asymmetry can be performed. These templates and subject-level landmark and surface data (below) have been shared here for the purpose of replication and facilitating potential, further modeling work (Haller, Banerjee et al. 1997; Csernansky, Schindler et al. 2004; Wang, Lee et al. 2007).

**Landmark and Surface Data—**Mapping of the template MR scan occurred in a two-step process. First, it was coarsely aligned to each target scan using landmarks, and then the diffeomorphic map was applied. Surfaces for subcortical structures in the target scans were generated by carrying the template surfaces through these maps (Joshi, Miller et al. 1997; Csernansky, Wang et al. 2004).

To facilitate our template-based mapping, global and local (i.e., structure-dependent) neuroanatomical landmarks were placed on the MR images. Landmark-based registration (Joshi, Miller et al. 1995) served to adjust the orientation and size for the head (based on global landmarks) and the subcortical structures of interest (based on local landmarks). *Global landmarks:* In each scan, twelve global landmarks were placed following procedures described in (Haller, Banerjee et al. 1997): at the points where the anterior and posterior commissures intersected the midsaggittal plane, and at the external boundaries of the cerebrum (anterior, posterior, superior, inferior and lateral). *Local landmarks:* 1) Hippocampus and amygdala were landmarked separately as follows. The most anterior and posterior boundaries of the structure were identified first and a line connecting these points created an anterior/posterior axis. Then in each of five equally distanced slices along this axis, four landmarks were placed at predetermined points in each slice. 2) Thalamus and basal ganglia were landmarked together as follows. The most anterior boundary of the caudate nucleus and the most posterior boundary of the thalamus were identified and a line connecting these points created an anterior/posterior axis. The region between the two points was then divided into five equally distanced slices along this axis and in each slice five landmarks were placed at predetermined places.

**FreeSurfer Data—**All scans were processed through FreeSurfer Version 3.0.4 (Desikan, Segonne et al. 2006) pipeline, with careful quality assurance as per FreeSurfer

recommendations. All FreeSurfer data, including subcortical segmentation, cortical parcellation and surface, and regional measurement data have been made available.

## Cognitive Data

Schizophrenia subjects demonstrate a wide array of cognitive deficits (Gur, Ragland et al. 2001). Data related to intellect, executive functioning (verbal and visual abstraction), and attention, as well as working and episodic memory are included in this data set. Measures of episodic memory included verbal and visual learning, and also the spontaneous and guided use of memory cues (Jacoby, Ste-Marie et al. 1993; Jacoby, Toth et al. 1993; Jacoby 1999; Jacoby, Debner et al. 2001). Our assessment of working memory included maintenance and manipulation processes across both verbal and visual modalities (Braver, Cohen et al. 1997; Barch, Csernansky et al. 2002). To date, our schizophrenia subjects have demonstrated deficits across all predicted cognitive domains using this battery (Delawalla, Barch et al. 2006; Cobia, Csernansky et al. 2011).

At each visit, the subjects were administered a core battery of neuropsychological measures relevant to areas identified in prior studies of cognition in schizophrenia (Nuechterlein, Barch et al. 2004). Tasks were grouped into the following four domains:

**Crystallized Intelligence—**Scaled scores from the Vocabulary subtest of Wechsler Adult Intelligence Scale (WMS-III; (Wechsler 1997)).

**Working Memory—**Scaled scores based on subtests from the Wechsler Memory Scale – Third Edition (WMS-III; (Wechsler 1997) including Digit Span (total forward and backwards), Spatial Span (total forward and backwards), and Letter-Number Sequencing, and also overall d-prime from the CPT-IP task (Cornblatt, Risch et al. 1988).

**Episodic Memory—**Included scaled scores from the WMS-III Logical Memory and Family Pictures subtests.

**Executive Function—**Included number of novel words generated on phonemic and semantic verbal fluency tasks (Benton 1976; Benton, Hamsher et al. 1984), time to completion on the Trail Making Test Part B (Reitan and Wolfson 1985), scaled scores on the WAIS-III Matrix Reasoning subtest, and number of perseverative errors on the Wisconsin Card Sorting Test (Heaton, Chelune et al. 1993).

Cronbach’s alpha (assessed in the standardization set of subjects) was 0.77, 0.78, and 0.70 for working memory, episodic memory, and executive function, respectively, in the individuals with schizophrenia, and 0.76, 0.65, and 0.67, respectively, in the control individuals.

## Genotyping Data

Blood for the isolation of DNA was collected from each of the subjects. These samples have been genotyped for a panel of 20 gene polymorphisms selected for their association with schizophrenia or their involvement in neurodevelopment. Examples of these genes include BDNF (rs6265), EGFR (rs10228436), FGF-2 (rs1048201), and IL-3 (rs40401).

Morphometric measures (e.g., structural volume) of individuals are compared and contrasted with specific differences (i.e., Single Nucleotide Polymorphisms, or SNPs) in the genes of interest. These differences are qualified by testing whether or not each subject has a particular polymorphism, and then how many copies of that polymorphism they have. A subject can fall into one of three categories: both copies of the gene are polymorphism-free (homozygous), one copy is polymorphism-free whereas the other copy has the polymorphism (heterozygous), or both copies of the gene carry the polymorphism (homozygous).

Presently we have genotyping data on 117 subjects with schizophrenia and 58 controls. DNA samples in additional subjects are being analyzed. All available genotype data will be made available with the scans to users of the database.

### Data Dictionary

Along with the data, we provide a data dictionary of terms. In the dictionary, standard descriptions for which ontologies exist are used. We searched following sources for ontology: NeuroLEX (http://neurolex.org/wiki/Main_Page) – a semantic wiki for terms used in Neuroscience, the Neuroscience Information Framework (http://www.neuinfo.org/) – a dynamic resource of Web-based neuroscience data, materials, and tools (NeuroLEX terms are actually published in NIF), and NCI Metathesaurus http://ncimeta.nci.nih.gov/ncimbrowser/) – a biomedical terminology database for translational and basic research. A detail list of these terms is presented in Table 3, and examples include “Socioeconomic Status,” “SAPS,” and “Cognitive Assessment.” For descriptions for which there are no standard ontologies, such as “Working Memory” or “Global Rating of Hallucinations,” we plan to work with NeuroLEX to arrive at standard definitions. The current version of the data dictionary can be downloaded through the data portal website, described below.

## Data sharing mechanisms

### Data Sharing Architecture: XNAT and XNAT Central

The collected MR datasets along with detailed clinical information are archived using the eXtensible Neuroimaging Archive Toolkit (XNAT), an open source data management and productivity platform for biomedical imaging research. XNAT was developed by the Neuroinformatics Research Group (NRG) at Washington University in St. Louis and the BIRN (Marcus, Olsen et al. 2007; Marcus, Wang et al. 2007). It is widely used across the world and is a core component of the emerging NIH-backed biomedical informatics backbone, including the Biomedical Informatics Research Network (BIRN) and National Alliance for Medical Image Computing (NA-MIC). XNAT includes a secure database backend and a rich web-based user interface. See paper on XNAT Central in this special issue for more details.

### XNAT Custom Schemas

XNAT relies on Extensible Markup Language (XML) schema documents to define the type of data that can be stored in the system. XML is the standard language for defining open and extensible data formats. The XML format provides a number of benefits for data

organization: it uniformly describes data and data structure, it makes data available for consistent and efficient programmatic manipulation, reuse, transmission and storage, and it simplifies data conversion to other formats. XNAT comes with a set of XML schemas that describe common data associated with neuroimaging studies. XNAT also allows for the extension of these schemas as well as the creation of custom schemas. NUSDAST contains all three types: common, extended, and custom schemas. The extended and custom schemas include: subject registration data and extended demographic and relationship information.

## Data Access

### XNAT Central

The project "NU Schizophrenia Data and Software Tool Federation using BIRN Infrastructure (NUSDAST)" is hosted on XNAT Central (visit http://central.xnat.org keyword NUSDAST or directly at http://tinyurl.com/av9h7jm)[1]. Within the NUSDAST project on the XNAT Central website, data are organized by subject following the XNAT architecture: study registration, longitudinal epochs of MR sessions. Within each epoch's MR session, scans and associated segmentations, surfaces, landmarks, and other data are listed. User download is accomplished via the Download action or the Manage Files action displayed on the XNAT web page. Data can also be retrieved via the XNAT REST API. Only users with validated accounts are able to down query and download data in NUSDAST.

### Web-based SchizConnect portal

Our recent effort to create a new and innovative way to access, query and retrieve neuroimaging data from various distributed research databases resulted in a web-based portal called SchizConnect (Wang, Alpert et al. 2014) (http://www.schizconnect.org). NUSDAST is one of the sources integrated with SchizConnect. SchizConnect communicates with NUSDAST via XNAT REST API in order to search and retrieve data based on criteria entered on SchizConnect web interface. With SchizConnet, users can get a summary of the NUSDAST data made available to them. In order to be able to retrieve NUSDAST data through SchizConnect, users are required to sign a data usage agreement (DUA). DUAs are used to restrict the use and disclosure of the available data as well as to permit publication of the research results in accordance with the applicable laws and regulations. The DUAs for SchizConnect and the current sources (COINS, NUSDAST, MCIC and FBIRN) require users to acknowledge the source of the data as well as the funding source in any publications and presentations, to protect the privacy of the subjects the data was collected from, and to keep data from being transferred to any third-party users that have not signed the DUAs related to the dataset. See paper on SchizConnect in this special issue for more details.

The significant advantage of using SchizConnect is that it allows the user to combine neuroimaging data from different databases to create a mediated dataset with related data. In our case, cognitive data, psychopathology measures based on SAPS and SANS and SNP data is stored at a different research database called REDCap. REDCap (Research Electronic

[1]Full URL: https://central.xnat.org/app/template/XDATScreen_report_xnat_projectData.vm/search_element/xnat:projectData/search_field/xnat:projectData.ID/search_value/NUDataSharing

Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.(Harris, Taylor et al. 2009). Since cognitive, SAPS/SANS and SNP data stored in the REDCap source are related to the neuroimaging data stored in NUSDAST, SchizConnect will seamlessly provide the user with all of the available data in one dataset, even though the data comes from geographically and technologically different sources.

## Discussion

In this paper, we described an instance of the Northwestern University Schizophrenia Data (NUSDAST), a static, longitudinal schizophrenia-related dataset, along with the XNAT Central hosting platform and the SchizConnect data portal used for accessing and sharing the dataset. All the clinical and longitudinal data was collected during conduction of two NIH-funded studies and all the data constituting the dataset was reviewed and approved by the investigators of these studies.

Concerning data sharing, this resource built and extended upon existing, standard schemas available for data sharing on XNAT Central (http://central.xnat.org/). Specifically, we developed additional schemas for storing demographic meta data in XNAT. This addition creates the opportunity to consistently expand and share schizophrenia research-related data[2]. We have significantly improved the way scientists are able to mine our dataset by integrating with the SchizConnect web portal for searching and downloading our data along with the accompanying longitudinal data.

Our well-described and comprehensive data on the normal controls and their siblings are valuable beyond the schizophrenia research community. For example, the SNP data, located at the REDCap resource, include ones that are related to neurodevelopment (e.g., BDNF), embryonic development and tissue repair (e.g., FGF-2), and immune response/inflammation (e.g., EGFR, IL-3). Mutations of many of these SNPs have been found to be related to cancer and neuropsychiatric disorders such as depression, anxiety and Alzheimer's disease. Therefore, with the accompanying imaging and cognitive data, our control subjects data can be of wider use, beyond schizophrenia research.

The schizophrenia research community has invested substantial resources in order to collect, manage and share increasingly larger datasets including neuroimaging data. The exploration of large, multi-modal, datasets has indeed improved our understanding of relationships among abnormalities of brain circuitry, brain function and genetic variability in schizophrenia (Kim, Manoach et al. 2009; Kim, Tura et al. 2010; Allen, Erhardt et al. 2011).

[2]Extended and new schemas created by NUSDAST can be used by other existing or future projects to describe their data. A benefit of sharing data through XNAT Central, rather than through an independent XNAT system, is that multiple projects can easily use the same schema representations and user interface components to represent common data elements. Users can then query and mine XNAT Central to locate data across the multiple projects using the shared schema.

Numerous data sharing initiatives were undertaken in order to create publicly accessible neuroimaging data collections, such as: FBIRN, MCIC, BSNIP, fMRI DataCenter (fMRIDC), the Open Access Series of Imaging Studies (OASIS). One of the main obstacles to the open- access sharing of research data is the lack of local organization and standard descriptions (Poline, Breeze et al. 2012). Different resources usually organize data in many different formats, which leads to difficulties in sharing and analyzing data. For example, the acquisition of a dataset from any given source will most likely require some programming work in order to make the dataset suitable for processing by an already utilized suite of tools. The introduction of the SchizConnect data portal implements the concept of "one-click share." The SchizConnect tool harmonizes different research databases in order to create a mediated mega-dataset with related longitudinal data. SchizConnect is a step in the direction of the creation of standardized procedures that include proliferation of open-source data sharing and of storage mechanisms. The SchizConnect portal responds to the current needs of researchers as it is a comprehensive and extensible system that accommodates different types of research and provides data-mining for common data processing functionality.

It is our hope that this effort will help overcome some of the commonly recognized technical barriers to advancing neuroimaging research (Poline, Breeze et al. 2012) by creating a research- ready dataset that meaningfully combines neuroimaging data with other relevant information. Currently, more data are being made available, such as fMRI and genome-wide scan (GWS) data. We are expanding the scope of schizophrenia neuroimaging data sharing by linking NUSDAST with FBIRN and MCIC through the development of SchizConnect, which is a powerful data mediation and integration platform that establishes a true federation of disparate, heterogeneous neuroimaging-related databases.

## Acknowledgments

This work was supported in part by NIH grants 1R01 MH084803, 1U01 MH097435-01A1, R01 EB009352.

## References

Allen EA, Erhardt EB, et al. A baseline for the multivariate comparison of resting-state networks. Front Syst Neurosci. 2011; 5:2. [PubMed: 21442040]

Andreasen, NC. The Scale for Assessment of Negative Symptoms (SANS). Iowa City, Iowa: The University of Iowa; 1983.

Andreasen, NC. The Scale for Assessment of Positive Symptoms (SAPS). Iowa City, Iowa: The University of Iowa; 1984.

Andreasen NC, Arndt S, et al. Symptoms of schizophrenia. Methods, meanings, and mechanisms. Arch Gen Psychiatry. 1995; 52:341–351. [PubMed: 7726714]

Barch DM, Csernansky J, et al. Working and long-term memory deficits in schizophrenia. Is there a common underlying prefrontal mechanism? Journal of Abnormal Psychology. 2002; 111:478–494. [PubMed: 12150424]

Benton, AL. Multilingual aphasia examination. Iowa City: University of Iowa; 1976.

Benton, AL., Hamsher, KDS., et al. Multilingual Aphasia Examination Third Edition (MAE). Iowa City: Psychological Assessment Resources; 1984.

Braver TS, Cohen JD, et al. A parametric study of prefrontal cortex involvement in human working memory. Neuroimage. 1997; 5(1):49–62. [PubMed: 9038284]

Buckner RL, Head D, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage. 2004; 23(2): 724–738. [PubMed: 15488422]

Claudio M, Roberto S. Using Marching Cubes on small machines. Graphical Models andImage Processing. 1994; 56:182–183.

Cobia DJ, Csernansky JG, et al. Cortical thickness in neuropsychologically near-normal schizophrenia. Schizophr Res. 2011; 133(1–3):68–76. [PubMed: 21981933]

Cornblatt BA, Risch NJ, et al. The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. Psychiatry Res. 1988; 26(2):223–238. [PubMed: 3237915]

Csernansky JG, Bardgett ME. Limbic-cortical neuronal damage and the pathophysiology of schizophrenia. Schizophr Bull. 1998; 24(2):231–248. [PubMed: 9613623]

Csernansky JG, Schindler MK, et al. Abnormalities of thalamic volume and shape in schizophrenia. Am J Psychiatry. 2004; 161(5):896–902. [PubMed: 15121656]

Csernansky JG, Wang L, et al. Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping. Am J Psychiatry. 2002; 159(12):2000–2006. [PubMed: 12450948]

Csernansky JG, Wang L, et al. Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. Neuroimage. 2004; 23(Suppl 1):S56–68. [PubMed: 15501101]

Delawalla Z, Barch DM, et al. Factors mediating cognitive deficits and psychopathology among siblings of individuals with schizophrenia. Schizophr Bull. 2006; 32(3):525–537. [PubMed: 16714471]

Desikan RS, Segonne F, 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):968–980. [PubMed: 16530430]

Duvernoy, HM. The Human Hippocampus: An Atlas of Applied Anatomy. Munich: J. F. Bergmann Verlag; 1988.

Duvernoy, HM. The Human Brain: Surface, Three-Dimensional Anatomy and MRI. Wien, New York: Springer-Verlag; 1991.

Goldman-Rakic PS. The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia. Biol Psychiatry. 1999; 46(5):650–661. [PubMed: 10472417]

Gur RC, Ragland JD, et al. Computerized neurocognitive scanning: II. The profile of schizophrenia. Neuropsychopharmacology. 2001; 25(5):777–788. [PubMed: 11682261]

Haller JW, Banerjee A, et al. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology. 1997; 202(2):504–510. [PubMed: 9015081]

Harms MP, Wang L, et al. Structural abnormalities in gyri of the prefrontal cortex in individuals with schizophrenia and their unaffected siblings. Br J Psychiatry. 2010; 196(2):150–157. [PubMed: 20118463]

Harms MP, Wang L, et al. Thalamic shape abnormalities in individuals with schizophrenia and their nonpsychotic siblings. J Neurosci. 2007; 27(50):13835–13842. [PubMed: 18077695]

Harris PA, Taylor R, et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009; 42(2):377–381. [PubMed: 18929686]

Heaton, RK., Chelune, GJ., et al. Wisconsin Card Sort-ing Test manual. Odessa, Florida: Psychological Assessment Resources, Inc; 1993.

Jacoby LL. Deceiving the elderly: effects of accessibility bias in cued-recall performance. Cognitive Neuropsychology. 1999; 16 000-000 (001-020).

Jacoby LL, Debner JA, et al. Proactive interference, accessability bias, and process dissociations: Valid subjective reports of memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27:686–700.

Jacoby, LL., Ste-Marie, D., et al. Redefining automaticity: unconscious influences, awareness, and control Attention: selection, awareness, and control; a tribute to Donald Broadbent. A Baddeley and L Weiskrantz; New York, Clarendon: 1993. p. 261-282.

Jacoby LL, Toth JP, et al. Separating conscious and unconscious influences of memory: measuring recollection. Journal of Experimental Psychology: General. 1993; 122(2):139–154.

John JP, Wang L, et al. Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri. Psychiatry Res. 2006; 148(2–3):151–163. [PubMed: 17088050]

Joshi, S., Miller, MI., et al. The generalized dirichlet problem for mapping brain manifolds. Int'l Symp on Optical Sceince, Engineering, and Instrumentation, Vision Geometry IV, SPIE; 1995.

Joshi S, Miller MI, et al. On the geometry and shape of brain sub-manifolds. Int J PatternRecog Artificial Intell (Special Issue). 1997; 11(8):1317–1343.

Karnik-Henry MS, Wang L, et al. Medial temporal lobe structure and cognition in individuals with schizophrenia and in their non-psychotic siblings. Schizophr Res. 2012

Kim DI, Manoach DS, et al. Dysregulation of working memory and default-mode networks in schizophrenia using independent component analysis, an fBIRN and MCIC study. Hum BrainMapp. 2009; 30(11):3795–3811.

Kim MA, Tura E, et al. Working memory circuitry in schizophrenia shows widespread cortical inefficiency and compensation. Schizophr Res. 2010; 117(1):42–51. [PubMed: 20096539]

Lorensen WE, Cline HE. Marching cubes: A high resolution 3d surface construction algorithm. Computer Graphics. 1987; 21(4):163–169.

Mai, JK., Assheuer, J., et al. Atlas of the Human Brain. San Diego: Academic Press; 1997.

Mamah D, Wang L, et al. Structural analysis of the basal ganglia in schizophrenia. Schizophr Res. 2007; 89(1–3):59–71. [PubMed: 17071057]

Marcus DS, Olsen TR, et al. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics. 2007; 5(1):11–34. [PubMed: 17426351]

Marcus DS, Wang TH, et al. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci. 2007; 19(9):1498–1507. [PubMed: 17714011]

Nuechterlein KH, Barch DM, et al. Identification of separable cognitive factors in schizophrenia. Schizophr Res. 2004; 72(1):29–39. [PubMed: 15531405]

Poline JB, Breeze JL, et al. Data sharing in neuroimaging research. Front Neuroinform. 2012; 6:9. [PubMed: 22493576]

Potkin SG, Macciardi F, et al. Identifying gene regulatory networks in schizophrenia. Neuroimage. 2010; 53(3):839–847. [PubMed: 20600988]

Qiu A, Younes L, et al. Combining anatomical manifold information via diffeomorphic metric mappings for studying cortical thinning of the cingulate gyrus in schizophrenia. Neuroimage. 2007; 37(3):821–833. [PubMed: 17613251]

Rastogi-Cruz, D., Csernansky, J. Adult Psychiatry. S Guze St Louis: Mosby, Inc; 1997. Clinical rating scales.

Reitan, RM., Wolfson, D. The Halstead-Reitan neuropsychological test battery: theory andclinical interpretation. Tucson, Arizona: Neuropsychology Press; 1985.

Ripke S, Sanders AR, et al. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011; 43(10):969–976. [PubMed: 21926974]

Smith MJ, Wang L, et al. Thalamic morphology in schizophrenia and schizoaffective disorder. J Psychiatr Res. 2011; 45(3):378–385. [PubMed: 20797731]

Venkatesan R, Haacke E. Role of high resolution in magnetic resonance (MR) imaging: Applications for MR angiography, intracranial T1-weighted imaging, and image interpolation. Int J Imaging Syst Technol. 1997; 8:529–543.

Wang, L., Alpert, KI., et al. SchizConnect: Large-Scale Schizophrenia Neuroimaging Data Integration and Sharing. Annual Meeting of the American College of Neuropsychopharmacology (ACNP); Phoenix, Arizona. 2014.

Wang L, Hosakere M, et al. Abnormalities of cingulate gyrus neuroanatomy in schizophrenia. Schizophr Res. 2007; 93(1–3):66–78. [PubMed: 17433626]

Wang L, Joshi SC, et al. Statistical analysis of hippocampal asymmetry in schizophrenia. Neuroimage. 2001; 14(3):531–545. [PubMed: 11506528]

Wang L, Kogan A, et al. Northwestern University Schizophrenia Data and Software Tool (NUSDAST). Front Neuroinform. 2013; 7:25. [PubMed: 24223551]

Wang L, Lee DY, et al. Validity of large-deformation high dimensional brain mapping of the basal ganglia in adults with Tourette syndrome. Psychiatry Res. 2007; 154(2):181–190. [PubMed: 17289354]

Wang L, Mamah D, et al. Progressive deformation of deep brain nuclei and hippocampal- amygdala formation in schizophrenia. Biol Psychiatry. 2008; 64(12):1060–1068. [PubMed: 18814865]

Wechsler, D. Wechsler Adult Intelligence Scale. San Antonio: The Psychological Corporation; 1997.

Weinberger DR, Berman KF, et al. Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. Am J Psychiatry. 1992; 149(7):890–897. [PubMed: 1609867]

**Table 1**

Subject Characteristics at Baseline.

| | Schizophrenia Subjects | Control Subjects | Schizophrenia Siblings | Control Siblings |
|---|---|---|---|---|
| N | 171 | 170 | 44 | 66 |
| Age at baseline (yr) | 33.8(12.5 [17–63]) | 31.4(13.8 [13–67]) | N/A | N/A |
| Gender (Male/Female) | 114/57 | 86/84 | 21/23 | 16/50 |
| Race (Caucasian/African-American/Other) | 90/78/3 | 61/105/2 | 17/27/0 | 16/50/0 |
| Global SAPS Score | 11.1 (12.7 [0–81]) | 0.06(0.3 [0–4]) | 0.5 (1.3 [0–7]) | N/A |
| Global SANS Score | 9.6 (10.7 [0–62]) | 0.04(1.7 [0–19]) | 2.3 (4.8 [0–38]) | N/A |

**Table 2**

MR scan parameters.

| Sequence | Protocol parameters |
|---|---|
| 3D turbo-FLASH | TR=20 ms, TE=5.4 ms, flip=30°, ACQ=1, 256×256 matrix, 1×l mm in-plane resolution, 180 slices, slice thickness 1 mm, 13:30 min scan time |
| 3D MPRAGE (2–4 repeats) | TR=9.7 ms, TE=4 ms, flip=10°, ACQ=1, 256×256 matrix, 1×l mm in-plane resolution, 128 slices, slice thickness 1.25 mm, 5:36 min scan time each |

## Table 3

Table of available ontology for data dictionary terms.

| Ontology<br><br>Term | NeuroLex Name & ID<br>NIF Standard Ontology ID<br>NCI Methathesaurus ID<br>Cognitive Atlas |
|---|---|
| Gender | NeuroLex: Gender assessment, birnlex_3026;<br>NIF: nif_inv:birnlex_3026;<br>NCI: C44177 |
| Race | NeuroLex: Race assessment, birnlex_3040;<br>NIF: nif_inv:birnlex_3040;<br>NCI: C17049 |
| Group | NeuroLex: Control role, birnlex_11017[3] |
| Ethnicity | NeuroLex: Ethnicity assessment, birnlex_3015;<br>NIF: nif_inv:birnlex_3015;<br>NCI: C16564 |
| Marital Status | NeuroLex: Marital status assessment, birnlex_3031;<br>NIF: nif_inv:birnlex_3031;<br>NCI: C25188 |
| Type of Housing/Living Arrangement | NeuroLex: Living arrangement assessment, birnlex_3030;<br>NIF: nif_inv:birnlex_3030;<br>NCI: C94852 |
| Number of Siblings | NCI: C102469 |
| Number of Children | NeuroLex: Offspring cardinality assessment, birnlex_3035;<br>NIF: nif_inv:birnlex_3035 |
| Current Occupation/Job Title | NeuroLex: Occupation assessment, birnlex_3036[4];<br>NIF: nif_inv:birnlex_3036;<br>NCI: C25193 |
| Principle Occupation | NeuroLex: Occupation assessment, birnlex_3036[4];<br>NIF: nif_inv:birnlex_3036;<br>NCI: C25193 |
| Level of Education | NeuroLex: Education assessment, birnlex_3014[5];<br>NIF: nif_inv:birnlex_3014 |
| Years of Schooling | NeuroLex: Education assessment, birnlex_3014[5];<br>NIF: nif_inv:birnlex_3014;<br>NCI: C17953 |
| Father's Level of Education | NeuroLex: Father's education, birnlex_3021[6];<br>NIF: nif_inv:birnlex_3021 |
| Father's Years of Schooling | NeuroLex: Father's education, birnlex_3021[6];<br>NIF: nif_inv:birnlex_3021 |
| Mother's Level of Education | NeuroLex: Mother's education, birnlex_3023[7];<br>NIF: nif_inv:birnlex_3023 |
| Mother's Years of Schooling | NeuroLex: Mother's education, birnlex_3023[7];<br>NIF: nif_inv:birnlex_3023 |
| Socioeconomic Status | NeuroLex: Socio-Economic status, birnlex_3048<br>NIF: nif_inv:birnlex_3048<br>NCI: C17468 |
| Handedness | NCI: (CUI) C0023114 |
| Edinburg Handedness Assessment | NeuroLex: Edinburg handedness assessment, birnlex_3013;<br>NIF: nif_inv:birnlex_3013 |
| SAPS | NeuroLex: Scale for the Assessment of Positive Symptoms, birnlex_3045[8]; |

| Ontology<br><br>Term | NeuroLex Name & ID<br>NIF Standard Ontology ID<br>NCI Methathesaurus ID<br>Cognitive Atlas |
|---|---|
| | NIF: nif_inv:birnlex_3045 |
| SANS | NeuroLex: Scale for the Assessment of Negative Symptoms, birnlex_3041[9];<br>NIF: nif_inv:birnlex_3041 |
| Cognitive Assessment | NeuroLex: birnlex_2021;<br>NIF: nif_inv:birnlex_2021;<br>NCI: C0870300 |
| Crystallized Intelligence | CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d09f64 |
| Working Memory | CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0b5a7 |
| Episodic Memory | CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0a1f4 |
| Executive Function | CogAtlas: http://cognitiveatlas.org/concept/id/trm_4a3fd79d0a252 |

[3]Currently a proxy class to be replaced by its OBI (The Ontology for Biomedical Investigations) equivalent.

[4]NeuroLex Occupation assessment includes work specialties as defined by duties and required skills as well as principal activity that a person does to earn money.

[5]NeuroLex Educational assessment includes Level of Education and Years of Schooling.

[6]NeuroLex Father's Education assessment includes Father's Level of Education and Father's Years of Schooling.

[7]NeuroLex Mother's Education assessment includes Mother's Level of Education and Mother's Years of Schooling.

[8]NeuroLex SAPS describes a type of assessment without any defined attributes.

[9]NeuroLex SANS describes a type of assessment without any defined attributes.