TECHNICAL REPORT OPEN ACCESS

The Psychosis MRI Shared Data Resource (Psy-ShareD)

Paul Allen^{1,2} | Mariana Zurita¹ | Rubaida Easmin¹ | Sara Bucci¹ | Matthew J. Kempton³ | Jack Rogers⁴ | Urvakhsh M. Mehta⁵ | Philip K. McGuire⁶ | Stephen M. Lawrie⁷ | Heather Whalley⁷ | Ary Gadelha⁸ | Graham K. Murray⁹ | Jane R. Garrison¹⁰ | Sophia Frangou^{11,12} | Rachel Upthegrove^{4,6} | Simon L. Evans¹³ | Veena Kumari¹⁴ | the Psy-ShareD Partnership

¹Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK | ²WILL Chair PSY Team, Centre LilNCog, INSERM U-1172, Lille, Haute de France, France | ³Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK | ⁴Institute for Mental Health, University of Birmingham, Birmingham, UK | ⁵Department of Psychiatry, National Institute of Mental Health and Neuro-Sciences (NIMHANS), Bangalore, India & Consciousness Studies Programme, National Institute of Advanced Studies (NIAS), Bangalore, India | ⁶Department of Psychiatry, University of Oxford, Oxford, UK | ⁷Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK | ⁸Schizophrenia Program, Department of Psychiatry, Escola Paulista de Medicina, Universidade Federal de São Paulo (PROESQ-EPM/UNIFESP), São Paulo, Brazil | ⁹Department of Psychiatry, University of Cambridge, Cambridge, UK | ¹⁰Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA | ¹²Djavad Mowafaghian Center for Brain Health, University of British Columbia, Vancouver, Canada | ¹³School of Psychology, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK | ¹⁴Department of Life Sciences, College of Health, Medicine and Life Sciences, Brunel University of London, London, UK

Correspondence: Paul Allen (paul.h.allen@kcl.ac.uk)

Received: 25 September 2024 | Revised: 30 January 2025 | Accepted: 3 February 2025

Funding: This work was supported by the United Kingdom Medical Research Council grant number MR/X010651/1. A list of funders and acknowledgements for Psy-ShareD datasets can be found at https://psyshared.com/Team.html.

ABSTRACT

Neuroimaging research in the field of schizophrenia and other psychotic disorders has sought to investigate neuroanatomical markers, relative to healthy control groups. In recent decades, a large number of structural magnetic resonance imaging (MRI) studies have been funded and undertaken, but their small sample sizes and heterogeneous methods have led to inconsistencies across findings. To tackle this, efforts have been made to combine datasets across studies and sites. While notable recent multicentre initiatives and the resulting meta- and mega-analytical outputs have progressed the field, efforts have generally been restricted to MRI scans in one or two illness stages, often overlook patient heterogeneity, and study populations have rarely been globally representative of the diversity of patients who experience psychosis. Furthermore, access to these datasets is often restricted to consortia members who can contribute data, likely from research institutions located in high-income countries. The Psychosis MRI Shared Data Resource (Psy-ShareD) is a new open access structural MRI data sharing partnership that will host pre-existing structural T1-weighted MRI data collected across multiple sites worldwide, including the Global South. MRI T1 data included in Psy-ShareD will be available in image and feature-level formats, having been harmonised using state-of-the-art approaches. All T1 data will be linked to demographic and illness-related (diagnosis, symptoms, medication status) measures, and in a number of datasets, IO and cognitive data, and medication history will also be available, allowing subgroup and dimensional analyses. Psy-ShareD will be free-to-access for all researchers. Importantly, comprehensive data catalogues, scientific support and training resources will be available to facilitate use by early career researchers and build capacity in the field. We are actively seeking new collaborators to contribute further T1 data. Collaborators will benefit in terms of authorships, as all publications arising from Psy-ShareD will include data contributors as authors.

Psy-ShareD Partnership (https://psyshared.com/Team.html).

Simon L. Evans and Veena Kumari contributed equally to this manuscript.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). Human Brain Mapping published by Wiley Periodicals LLC.

1 | Introduction

Despite decades of research investment, schizophrenia (SCZ) spectrum disorders are still a leading cause of disability (Lopez et al. 2006; Chong et al. 2016), representing a huge economic health burden and causing untold suffering for patients and families. Given the societal and economic costs of psychotic illness and the limited efficacy of current treatment options, it is of high strategic importance that we improve our understanding of the neurobiological mechanisms that underlie SCZ spectrum disorders, to inform better detection and stratification and treatment development. Over recent decades neuroimaging research in SCZ and psychosis populations has sought to investigate neuroanatomical markers relative to healthy control (HC) groups. Numerous magnetic resonance imaging (MRI) studies investigating brain structure in SCZ and psychosis risk populations (i.e., clinical risk, familial risk and schizotypy (SZT)) have been funded in the United Kingdom and internationally, progressing our understanding of the neuroanatomical basis of SCZ and of these psychosis risk populations. The first neuroimaging anatomical investigation of patients with SCZ, using CT scans, was published in 1976 (Johnstone et al. 1976), and at the time of writing the current paper, a PubMed search using the terms 'schizophrenia MRI (volume or thickness)' shows that 3241 publications reporting the neuroanatomical basis of SCZ using MRI were published between 1986 and December 2024. Many structural MRI studies have also been published in first episode and psychosis risk populations. However, some single-site MRI studies have small sample sizes (Button et al. 2013) due to the high costs of MRI, alongside difficulties recruiting patients with SCZ and other psychotic illnesses. In particular, studies in psychosis risk groups often have sample sizes under N = 50 (Luna et al. 2022) due to difficulties in recruiting participants meeting risk criteria. While the field has advanced significantly in recent decades and now recognises the need for enhanced sample sizes, reproducibility issues persist in the literature (Marek et al. 2022) hampering our ability to develop a definitive model of the neuroanatomical basis of SCZ, and the illness's developmental trajectory. Larger samples not only increase power to detect smaller effect sizes, but also allow studies to address the issue of clinical heterogeneity within SCZ (e.g., Chand et al. 2020), which has seriously impeded efforts to uncover a 'common' underlying pathophysiology (Orsolini et al. 2022). Although recent neuroimaging work has begun to address this, through clustering and subgroup analyses (e.g., Jiang et al. 2023; Alkan and Evans 2022; Chand et al. 2020), single-site studies often lack sufficient subgroup numbers to draw reliable inferences, and the larger-scale, multi-site analyses have tended to just appraise betweengroup differences, rather than attempting to identify neuroanatomical signatures associated with the distinct symptom profiles that exist within patient populations. Increased sample sizes allows profile-specific inferences to be generated and confirmed through cross-validation, allowing better translation into potential clinical benefit.

In an attempt to address this problem, there have recently been a number of notable large, international multi-site MRI initiatives and consortia focusing on individuals at high clinical risk for psychosis or those with a diagnosis of SCZ, such as EU-GEI (Modinos et al. 2020), PRONIA (Rosen et al. 2021), PSYSCAN (Tognin et al. 2020) and PHENOM (Chand et al. 2020). Additionally, there have been meta and mega analyses of existing data by the ENIGMA consortium (ENIGMA Clinical High Risk for Psychosis Working Group 2021; Lamsma et al. 2024; van Erp et al. 2018). While these have significantly advanced the field by improving our understanding of the anatomical changes seen in SCZ and psychosis risk populations, as well as addressing power issues, these consortia studies mostly have an over-representation of samples from sites based in European, Australian and North American regions. While some initiatives (notably ENGIMA) have sought to improve their global reach, there is (particularly in recent years) an emergence of relevant research from sites based outside of these regions. It would be highly beneficial to incorporate such data (even if they are currently limited in number), as the paucity of samples from diverse ethnoracial groups (Fonseca et al. 2021) prevents conclusions as to whether patients from different backgrounds may show distinct patterns of neural markers linked to risk and aetiology (Fearon et al. 2006; Hutchinson et al. 1999; Kalra et al. 2012; Lim et al. 2011; Suhail and Cochrane 2002). Importantly, there is also a lack of open access neuroimaging data in psychosis: those that exist (e.g., http://schizconnect. org/) are severely limited in terms of size, global reach and what data and support are available to researchers. Although there are exceptions, such as ENIGMA, most consortia initiatives restrict data access to contributing sites and members (or do not publicise procedures for non-contributors to access data), making it difficult for aspiring, but poorly resourced researchers outside these consortia to conduct well-powered analyses. Thus, despite the considerable cost and effort associated with the acquisition and analyses of MRI datasets in SCZ spectrum disorder, first episode psychosis (FEP), and psychosis risk populations, clarity around the role of key brain regions and how these evolve across the disease trajectory and their link to symptoms, is still lacking (Alnæs et al. 2019; Honea et al. 2005). To address this, the ENIGMA consortium includes working groups focusing on clinical high risk (CHR) and SZT populations, and these have provided useful insights from their respective populations (e.g., ENIGMA Clinical High Risk Working Group et al. 2021; Kirschner et al. 2022). However, the field would benefit from a neuroimaging resource within which all disease phases along with corresponding clinical and cognitive profiles are adequately represented, to allow analyses across multiple disease phases. Additionally, as mentioned above, large-scale analyses have to date often ignored the diverse clinical syndromes within SCZ; it is important to consider distinct symptom profiles. Together, these issues have limited our ability to draw useful inferences regarding underlying stage-specific neuroanatomical changes and mechanisms and understand how these contribute to disease progression (Keshavan et al. 2020). Finally, many consortia outputs have relied on meta-analyses of summary data from contributing studies/centres (although some of the recent ENIGMA outputs have been based on mega analyses of pooled subject data (Lamsma et al. 2024)). While there have been some efforts to effectively harmonise feature-level data (e.g., cortical thickness and surface area measures) across participating centres, and most consortia do implement at least some harmonisation procedures, voxel level analyses have been more difficult to achieve until more recently (Si et al. 2024).

Another significant obstacle to progress in the field is the limited access to MRI datasets within SCZ and psychosis populations, which remains largely confined to a select group of researchers. In the United Kingdom, for instance, only a few centres have collected and published MRI data on SCZ populations. The study of psychosis risk cohorts is even more concentrated, with research primarily conducted at centres in London, Cambridge, Birmingham (for CHR) and Edinburgh (for familial high risk). This situation underscores the need to promote greater equity in scientific research, aligning with the UK government's 'levelling up' agenda. The current inequity has hindered the field's ability to pursue essential new research, address existing knowledge gaps and resolve inconsistencies. Additionally, early career researchers often face challenges in accessing these MRI datasets, limiting their opportunities for development and research, which in turn weakens overall research capacity in this critical area.

Therefore, the Psychosis Shared MRI Data Resource (Psy-ShareD) is important for the following key reasons: First, although the prevalence of SCZ is uniform globally (Jongsma et al. 2018), there are subtle variations in the illness characteristics, including long-term outcomes, across geographies and ethnocultural spaces (see Kalra et al. 2012 for review). Second, many brain-based investigations into the pathogenesis of SCZ have been conducted in the Global North and other developed countries, while a majority of the SCZ populace resides in the Global South (Kalra et al. 2012). Lastly, there is a growing literature on how ethnicity (encompassing genetic, linguistic, cultural and environmental factors) can impact brain structure and function (Gong et al. 2015; Strawbridge et al. 2018).

2 | Psychosis Shared MRI Data Resource (Psy-ShareD) and Methods

2.1 | Agents

For clarity going forward, here we define terms for the following agents and groups that are used in the sections below.

- The Psy-ShareD Partnership comprises Data Contributors and Team Members.
- *Data Contributors* are those agents that have contributed MRI and linked clinical, demographic and cognitive data to Psy-ShareD.
- Team Members are those agents who are co-investigators and project staff. Team Members can also be Data Contributors.
- *Data Users* are those agents that access Psy-ShareD data for analyses and publication. *Data Users* can also be *Data Contributors* and *Team Members*.

2.2 | Aims and Objectives

The *Psy-ShareD Partnership* is funded by the UK Medical Research Council (MR/X010651/1) with the remit to combine high-quality, pre-existing 'legacy' structural MRI datasets with linked clinical and cognitive data into one free-to-access resource. Psy-ShareD is suitable for a priori hypothesis testing,

exploration for novel hypothesis generation and methodological training. Specifically, the Psy-ShareD Partnership brings together pre-existing structural MRI scans in people with SCZ, FEP, CHR for psychosis, SZT and HC populations. The resource also includes MRI data from people diagnosed with mood disorders, notably major depressive disorder (MDD) and bipolar disorder (BPD). The latter populations, although not belonging to the psychosis spectrum, are diagnostic groups we plan to expand in the future (see Section 4). The resource will also contain measures of childhood trauma (CT) in both clinical and HC populations, potentially allowing categorical and correlational type analyses examining the effects of early experiences on neuroanatomy. Overall, Psy-ShareD will allow comparisons among various illness phenotypes and across illness stages, enhancing our capacity to understand both the neuroanatomical basis and their trajectory of SCZ and psychosis. A full list of available datasets can be viewed on the Psy-ShareD website (https://psysh ared.com/Data.html).

To summarise, the primary objectives of the *Psy-ShareD Partnership* are as follows:

- 1. Build a sustainable free-to-access structural MRI data repository from pre-existing MRI datasets in SCZ, FEP, CHR, SZT, BPD, MDD and HC participants, and include within the database linked demographic, clinical and IQ data. Neuropsychological, medication, treatment responsiveness and functional outcome measures will also be included where available.
- 2. Seek out existing MRI datasets from across the world, particularly those acquired in non-Western populations.
- 3. Organise and curate all Psy-ShareD datasets and catalogue these clearly via Kings College London (KCL) FigShare (https://kcl.figshare.com) and the Psy-ShareD Website (https://psyshared.com).
- 4. Undertake and publish a series of validation and proofof-concept analyses using Psy-ShareD datasets to demonstrate the reliability and feasibility of use.

2.3 | Data Transfer

Contributing centres will share raw T1 images after defacing and removing any personally identifying information. This allows Psy-ShareD to conduct rigorous harmonisation procedures based on the raw data while safeguarding participants' privacy and better accounting for variability between sites, and providing users with a higher level of reliability and power. All Psy-ShareD partner institutions that share data complete a data sharing agreement (DSA). More information about the Psy-ShareD DSA process are provided on our website (https:// psyshared.com/Datacontribution.html). Once a DSA is established with the contributing partner/institution, and prior to T1 and linked data transfer, the Psy-ShareD Team provides the transferring sites with scripts and instructions for data de-identification (defacing) and full anonymisation, in accordance with the Psy-ShareD Data Transfer Protocol (Figure 1). A script is also provided to produce unique Psy-ShareD ID codes for each MRI T1 scan transferred from the contributing site. These ID codes also link MRI and corresponding clinical,



FIGURE 1 | Schematic showing Psy-ShareD procedures for (a) data transfer and preparation steps taken by the data contributors (dark blue) prior to storage (red), (b) data processing (light blue) and harmonisation (orange) steps and results carried out by the Psy-ShareD team.

demographic and cognitive data. De-identification and anonymisation scripts are available via the Psy-ShareD website. Once data preparation steps are completed, data are transferred via BitBox (https://www.bitbox-imaging.com). BitBox (Easmin et al. 2022) was developed for the purpose of imaging and clinical data transfer within multi-site studies. BitBox is used to transfer medical imaging data from external and independent sites into the KCL neuroimaging network, securely and in accordance with GDPR regulations. All datasets contained within the Psy-ShareD resource have ethical approval for reuse and sharing. Datasets (usually older datasets) where consent for reuse cannot be traced, or that are not covered by umbrella research ethics committee approval for reuse and sharing, are not included in the resource.

2.4 | Data Harmonisation

The T1 MRI data in Psy-ShareD were acquired at different sites using various T1 acquisition sequences (e.g., SPGR, ADNI-GO, MPRAGE). To account for different acquisition parameters used across sites, we harmonised the data using two types of methods: (i) feature-level and (ii) image-level harmonisation. To make available feature-level data including cortical and subcortical volumes, cortical thickness and surface area data, we have used FreeSurfer version 7.3.2 (https:// surfer.nmr.mgh.harvard.edu/). We then used the neuroHarmonize package (https://github.com/rpomponio/neuroHarmo nize; Pomponio et al. 2020) to harmonise feature-level MRI data by removing unwanted variation induced by scanner differences such as differences in acquisition parameters, field strength and manufacturer, while preserving biological variability between individuals using an empirical Bayes framework (Fortin et al. 2017, 2018; Johnson et al. 2007; Pomponio et al. 2020). Since this harmonisation is sample-dependent, Psy-ShareD tools are available to aid with feature-level harmonisation. This allows users to harmonise feature-level MRI data, effectively removing unwanted variation due to scanner differences while preserving the biological variability between individuals, using an empirical Bayes framework (Fortin et al. 2017, 2018; Johnson et al. 2007; Pomponio et al. 2020). Site harmonisation with the neuroHarmonize package can be applied to all or a sub-set of the datasets in Psy-ShareD as per users' requirements. Two advantages of neuroHarmonize over other methods (e.g., using site as a covariate in statistical models) are that it improves the removal of scanner effects in datasets with small sample sizes and does not make assumptions about the neuroimaging technique being used (Radua et al. 2020).

A unique feature of Psy-ShareD is the use of image-level harmonisation to T1 data, which produces whole brain images where site-specific information is removed, therefore allowing analyses such as voxel-based morphometry (VBM; Figure 2). We are employing two distinct methods to achieve this: HACA3 (Zuo et al. 2023) and IGUANe (Roca et al. 2025). HACA3 is a deeplearning method with an 'encoder-attention-decoder' architecture (Zuo et al. 2023). In its simpler form, it takes two types of input images: a target image and an original input image to be harmonised towards the target image. It then encodes the anatomy, contrast and image artifact from the inputs. The method creates a synthetic image with the anatomical information from the original input image, but with the target image's contrast, while removing artifacts (Zuo et al. 2023). By harmonising an image to a target (Figure 2A), original contrast values from the original images are moved to values in the range of a target image (Figure 2B), while maintaining its anatomical features (Figure 2C). Full data validation work using these harmonisation processes is underway and will be reported separately.

IGUANe is based on the CycleGAN (Cycle Generative Adversarial Network) architecture (Zhu et al. 2017), which features generator-discriminator pairs. The generator produces synthetic images that closely resemble the real dataset distribution, while the discriminator distinguishes between synthetic and real images. CycleGAN incorporates a cycleconsistency constraint in its loss function to ensure that unpaired image translation and content (anatomical structure) preservation are achieved. This model is particularly beneficial for image-level harmonisation as it requires no supervision



FIGURE 2 | Example of harmonisation. (a) Single subject image that was harmonised towards a target image. (b) Voxel value frequency plot showing the values of voxels of the example image shown on (a) before and after it was harmonised towards a target image, (c) correlation plots showing the volumes of white matter (left), grey matter (centre) and cerebrospinal fluid (right) after segmenting several original images and their harmonised version (HACA3 image). Each plot also depicts the correlation coefficient and corresponding *p* value, as well as the intra-class correlation coefficient (ICC(A,1) or two-way random absolute agreement measure; McGraw and Wong 1996) between the values obtained with both images.

during the training process. Since no ground truth is required for training, the output of the second GAN (generative adversarial network) must correspond only with the input of the first GAN. This means that there is no need for acquiring scans from the same participant at each site (Cackowski et al. 2021). Given that the original study focused on a 2D deep-learning framework, we used the 3D framework of CycleGAN proposed by Roca et al. (2025), which extends the model allowing it to harmonise unseen sites. This adaptation uses PatchGAN discriminators with 3D convolutions to process entire 3D MRI scans. Scans are visually inspected, and we conduct rigorous quality checks throughout our processing pipeline to identify and address any errors. All datasets contain linked clinical, demographic and cognitive/IQ data. Where needed, harmonisation and standardisation procedures are available for these data types. The Psy-ShareD team is currently developing tools that will allow the standardisation of clinical and cognitive variables contained within the resource using ComBat-GAM algorithms similar to those used by the ENIGMA Clinical Endpoints Work Group (Kennedy et al. 2024) and conversion equations for clinical assessment tools (Grot et al. 2021). These tools can be adapted by data users as needed, for example, for standardising predictive variables, and so forth.

2.5 | Data Description

All MRI T1 data are linked to anonymised demographic and clinical data as detailed in our data catalogues available via KCL FigShare (https://kcl.figshare.com). All datasets include information about participants' age and sex. Ethnoracial data, educational

level and handedness are also available in many datasets. All datasets include caseness, and several datasets also include data for illness duration/onset and medication status. Symptom severity in SCZ and FEP populations was derived from the Positive and Negative Symptom Scale (PANSS; Kay et al. 1987) and the Scale for the Assessments of Positive and Negative Symptoms (SAPS/SANS; Andreasen 1984). Cohorts of individuals at high risk for psychosis were assessed for caseness and severity using the Comprehensive Assessment for an At-Risk Mental States (CAARMS; Yung et al. 2005) and with SIPS/SOP (McGlashan et al. 2001) for some datasets. Global functioning (General Assessment of Functioning, GAF; Hall 1995) and measures for depression and anxiety symptoms are also available in several datasets. In non-clinical risk groups (SZT and CT), a range of sub-clinical measures are provided where available, such as the Oxford-Liverpool Inventory for Feelings and Experiences (Mason and Claridge 2006), Schizotypal Personality Questionnaire (Raine 1991) and Childhood Trauma Questionnaire (Bernstein et al. 2003). Cognitive and IQ data are also available in several datasets assessing intellectual function, working memory and executive function and verbal learning (see data catalogues).

3 | Psy-ShareD Data Management: Access and Storage

3.1 | Data Access and Storage

Data catalogues are available through KCL FigShare (https://kcl. figshare.com) and the Psy-ShareD website (https://psyshared. com/Data.html). Data catalogues contain information about each dataset, including caseness, sample size, data formats, demographics, clinical, medication and cognitive variables.

MRI T1 and linked clinical/demographic/cognitive data will be stored separately on the KCL Neuroimaging Network and will not be directly accessible to users. Members of the Psy-ShareD Partnership act as custodians of the data contained within the resource. Potential data users can access data contained in the Psy-ShareD database by submitting a data access request using a short data access form available on the Psy-ShareD website. The data access procedure is shown in Figure 3. The data access procedure is required to track and monitor access and to allow data contributors whose datasets are requested to provide approval.

Once a project is approved, after the data access request, data access will be managed via DataLad and GIN (G-Node Infrastructure). DataLad (https://www.datalad.org/) is an open-source data management tool that tracks data and ensures reproducibility. On the other hand, GIN (https://gin.g-node.org/) is a free and open platform for data storage and distribution, allowing data to be accessed and managed from various locations while keeping it synchronised and backed up. After the data are prepared with DataLad, it will be shared through the GIN application. This combination supports dataset versioning and provides controlled access for collaborators.

All Psy-ShareD data users will be required to adhere to the memorandum of understanding (MoU) which can be accessed via the Psy-ShareD website. The Psy-ShareD MoU provides details and instructions for Psy-ShareD *Data Users* about publication, authorship and open access data arrangements.

3.2 | Data Use and Publication Policy

Full details are available in the Psy-ShareD MoU document (see website). Briefly, we stipulate that all outputs that use Psy-ShareD datasets include data contributors as co-authors. In addition, the

Psy-ShareD Partnership will be listed on the author line of all publications using data from the Psy-ShareD resource. As such, all publications and citations resulting from Psy-ShareD will be linked to the respective data contributors. Specific publication policies (i.e., conference abstracts, posters, symposia) are listed in the Psy-ShareD MoU document.

3.3 | Tools, Training and Open Access

For partners that contribute data, instructions and tools for data preparation and transfer are available free upon request from the Psy-ShareD team. Tools for T1 image-level harmonisation using HACA3 and CycleGan and using ComBat for feature-level harmonisation will also be available via the Psy-ShareD website; all our processing protocols are fully documented and transparent. To facilitate the usage of Psy-ShareD data, the Psy-ShareD Partnership will develop and provide supporting material and workshops for potential users. These will include open access training resources for data organisation and analysis. Data sharing plans are fully in line with the UK Medical Research Council's data sharing policy and Open Science Framework principles (https://osf.io) that promote and support networking and partnership activities, enabling knowledge sharing and open access to data across institutions.

4 | Future Directions and Psy-ShareD Phase II

Psy-ShareD is a growing resource with new datasets from around the world continually sought and being added on an ongoing basis. Currently, MRI datasets are available from sites in the United Kingdom, Europe, South and Central America, India and Japan. Further, DSAs are currently in progress with sites in North America and Australia as the resource continues to expand. We also plan to expand Psy-ShareD transdiagnostically. The resource already includes MRI scans from patients diagnosed with BPD and MDD, and it is planned that MRI data



FIGURE 3 | Psy-ShareD data access procedure.

from cohorts with other DSM-5 disorders will be added to the resource in the future. It is also anticipated that, going forward, data from other MRI modalities can be added and linked to Psy-ShareD T1 data, that is, resting state functional MRI and 1H-magnetic resonance spectroscopy. In summary, future directions and objectives are as follows:

- General expansion to new sites, especially focusing on enhancing geographical and ethnoracial diversity, and continued refinement of MRI data harmonisation protocol (e.g., by computing a generalised reliability map by scanning a few identical participants on some of the most representative MRI scanners).
- Expanding beyond T1 datasets to multimodal MRI (i.e., resting state fMRI, MRS).
- Increasing the range of diagnoses for future clinical samples.
- Inclusion of genetic and omics data (inflammatory markers etc.) where possible.
- Development of clinical tools, using Psy-ShareD data.

We are exploring methodological approaches that will allow Psy-ShareD data to be used for clinical applications, for example, clustering and neurobiological stratification (Lalousis et al. 2022) and normative modelling of brain morphometry (Haas et al. 2024) approaches. Finally, we will link our data catalogues to DATAMIND (https://datamind.org.uk) and we are exploring the possibility of also linking to HDRUK (https://www.hdruk.ac.uk).

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

P.A. has been funded by FrieslandCampina, G.K.M. consults for Leso Digital Health. Sophia Frangou (S.F.) is a deputy editor of Human Brain Mapping and a co-author of this article. To minimise bias, S.F. was excluded from all editorial decision-making related to the acceptance of this article for publication.

Data Availability Statement

The data that support the findings of this study are openly available in Psy_ShareD at https://psyshared.com/Home.html.

References

Alkan, E., and S. L. Evans. 2022. "Clustering of Cognitive Subtypes in Schizophrenia Patients and Their Siblings: Relationship With Regional Brain Volumes." *Schizophrenia* 8, no. 1: 50.

Alnæs, D., T. Kaufmann, D. van der Meer, et al. 2019. "Brain Heterogeneity in Schizophrenia and Its Association With Polygenic Risk." *JAMA Psychiatry* 76, no. 7: 739–748. https://doi.org/10.1001/jamapsychiatry.2019.0257.

Andreasen, N. C. 1984. Scale for the Assessment of Positive Symptoms. University of Iowa. https://doi.org/10.1037/t48377-000.

Bernstein, D. P., J. A. Stein, M. D. Newcomb, et al. 2003. "Development and Validation of a Brief Screening Version of the Childhood Trauma Questionnaire." Child Abuse & Neglect 27, no. 2: 169–190. https://doi. org/10.1016/s0145-2134(02)00541-0.

Button, K. S., J. P. Ioannidis, C. Mokrysz, et al. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews. Neuroscience* 14, no. 5: 365–376. https://doi.org/10.1038/nrn3475.

Cackowski, S., E. L. Barbier, M. Dojat, and T. Christen. 2021. "comBat Versus cycleGAN for Multi-Center MR Images Harmonization." https:// openreview.net/forum?id=cbJD-wMlJK0.

Chand, G. B., D. B. Dwyer, G. Erus, et al. 2020. "Two Distinct Neuroanatomical Subtypes of Schizophrenia Revealed Using Machine Learning." *Brain: A Journal of Neurology* 143, no. 3: 1027–1038. https://doi.org/10.1093/brain/awaa025.

Chong, H. Y., S. L. Teoh, D. B. Wu, S. Kotirum, C. F. Chiou, and N. Chaiyakunapruk. 2016. "Global Economic Burden of Schizophrenia: A Systematic Review." *Neuropsychiatric Disease and Treatment* 16, no. 12: 357–373.

Easmin, R., G. Nordio, A. Giacomel, F. Turkheimer, S. Williams, and M. Veronese. 2022. "Bitbox: A Cloud-Based Data Sharing Solution for Medical Images. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society." *Annual International Conference* 2022: 2712–2715. https://doi.org/10.1109/EMBC48229.2022.9871689.

ENIGMA Clinical High Risk for Psychosis Working Group. 2021. "Association of Structural Magnetic Resonance Imaging Measures With Psychosis Onset in Individuals at Clinical High Risk for Developing Psychosis: An ENIGMA Working Group Mega-Analysis." *JAMA Psychiatry* 78, no. 7: 753–766. https://doi.org/10.1001/jamapsychiatry. 2021.0638.

Fearon, P., J. B. Kirkbride, C. Morgan, et al. 2006. "Incidence of Schizophrenia and Other Psychoses in Ethnic Minority Groups: Results From the MRC AESOP Study." *Psychological Medicine* 36, no. 11: 1541–1550. https://doi.org/10.1017/S0033291706008774.

Fonseca, L., B. F. Sena, N. Crossley, et al. 2021. "Diversity Matters: Opportunities in the Study of the Genetics of Psychotic Disorders in Low- and Middle-Income Countries in Latin America." *Brazilian Journal of Psychiatry* 43, no. 6: 631–637. https://doi.org/10.1590/1516-4446-2020-1240.

Fortin, J.-P., N. Cullen, Y. I. Sheline, et al. 2018. "Harmonization of Cortical Thickness Measurements Across Scanners and Sites." *NeuroImage* 167: 104–120. https://doi.org/10.1016/j.neuroimage.2017.11.024.

Fortin, J.-P., D. Parker, B. Tunç, et al. 2017. "Harmonization of Multi-Site Diffusion Tensor Imaging Data." *NeuroImage* 161: 149–170. https:// doi.org/10.1016/j.neuroimage.2017.08.047.

Gong, Q., P. Dazzan, C. Scarpazza, et al. 2015. "A Neuroanatomical Signature for Schizophrenia Across Different Ethnic Groups." *Schizophrenia Bulletin* 41, no. 6: 1266–1275. https://doi.org/10.1093/schbul/sbv109.

Grot, S., C. É. Giguère, S. Smine, et al. 2021. "Converting Scores Between the PANSS and SAPS/SANS Beyond the Positive/Negative Dichotomy." *Psychiatry Research* 305: 114199.

Haas, S. S., R. Ge, I. Agartz, et al. 2024. "Normative Modeling of Brain Morphometry in Clinical High-Risk for Psychosis." *JAMA Psychiatry* 81, no. 6: 77–88. https://doi.org/10.1001/jamapsychiatry.2023.3850.

Hall, R. C. 1995. "Global Assessment of Functioning. A Modified Scale." *Psychosomatics* 36, no. 3: 267–275. https://doi.org/10.1016/S0033 -3182(95)71666-8.

Honea, R., T. J. Crow, D. Passingham, and C. E. Mackay. 2005. "Regional Deficits in Brain Volume in Schizophrenia: A Meta-Analysis of Voxel-Based Morphometry Studies." *American Journal of Psychiatry* 162, no. 12: 2233–2245. https://doi.org/10.1176/appi.ajp.162.12.2233.

Hutchinson, G., N. Takei, P. Sham, I. Harvey, and R. M. Murray. 1999. "Factor Analysis of Symptoms in Schizophrenia: Differences Between White and Caribbean Patients in Camberwell." *Psychological Medicine* 29, no. 3: 607–612. https://doi.org/10.1017/S0033291799008430.

Jiang, Y., J. Wang, E. Zhou, et al. 2023. "Neuroimaging Biomarkers Define Neurophysiological Subtypes With Distinct Trajectories in Schizophrenia." *Nature Mental Health* 1: 186–199.

Johnson, W. E., C. Li, and A. Rabinovic. 2007. "Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods." *Biostatistics* 8, no. 1: 118–127. https://doi.org/10.1093/biostatistics/ kxj037.

Johnstone, E., C. D. Frith, T. J. Crow, J. Husband, and L. Kreel. 1976. "Cerebral Ventricular Size and Cognitive Impairment in Chronic Schizophrenia." *Lancet* 308, no. 7992: 924–926. https://doi.org/10.1016/ S0140-6736(76)90890-4.

Jongsma, H. E., C. Gayer-Anderson, A. Lasalvia, et al. 2018. "Treated Incidence of Psychotic Disorders in the Multinational EU-GEI Study." *JAMA Psychiatry* 75, no. 1: 36–46. https://doi.org/10.1001/jamapsychi atry.2017.3554.

Kalra, G., D. Bhugra, and N. Shah. 2012. "Cultural Aspects of Schizophrenia." *International Review of Psychiatry* 24, no. 5: 441–449. https://doi.org/10.3109/09540261.2012.708649.

Kay, S. R., A. Fiszbein, and L. A. Opler. 1987. "The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia." *Schizophrenia Bulletin* 13, no. 2: 261–276. https://doi.org/10.1093/schbul/13.2.261.

Kennedy, E., S. Vadlamani, H. M. Lindsey, et al. 2024. "ENIGMA Clinical Endpoints Working Group. Bridging Big Data in the ENIGMA Consortium to Combine Non-Equivalent Cognitive Measures." *Scientific Reports* 14, no. 1: 24289.

Keshavan, M. S., G. Collin, S. Guimond, S. Kelly, K. M. Prasad, and P. Lizano. 2020. "Neuroimaging in Schizophrenia." *Neuroimaging Clinics of North America* 30, no. 1: 73–83. https://doi.org/10.1016/j.nic.2019. 09.007.

Kirschner, M., B. Hodzic-Santor, M. Antoniades, et al. 2022. "Cortical and Subcortical Neuroanatomical Signatures of Schizotypy in 3004 Individuals Assessed in a Worldwide ENIGMA Study." *Molecular Psychiatry* 27, no. 2: 1167–1176.

Lalousis, P. A., L. Schmaal, S. J. Wood, et al. 2022. "Neurobiologically Based Stratification of Recent-Onset Depression and Psychosis: Identification of Two Distinct Transdiagnostic Phenotypes." *Biological Psychiatry* 92, no. 7: 552–562. https://doi.org/10.1016/j.biopsych.2022. 03.021.

Lamsma, J., A. Raine, S. M. Kia, et al. 2024. "Structural Brain Abnormalities and Aggressive Behaviour in Schizophrenia: Mega-Analysis of Data From 2095 Patients and 2861 Healthy Controls via the ENIGMA Consortium." *medRxiv*: The Preprint Server for Health Sciences. 2024.02.04.24302268. https://doi.org/10.1101/2024.02.04. 24302268.

Lim, C. S., M. Subramaniam, L. Y. Poon, S. A. Chong, and S. Verma. 2011. "Cross-Ethnic Differences in Severity of Symptomatology of Individuals With First-Episode Schizophrenia Spectrum Disorder." *Early Intervention in Psychiatry* 5: 242–248.

Lopez, A. D., C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray, eds. 2006. *Global Burden of Disease and Risk Factors*. International Bank for Reconstruction and Development/The World Bank. http:// www.ncbi.nlm.nih.gov/books/NBK11812/.

Luna, L. P., J. Radua, L. Fortea, et al. 2022. "A Systematic Review and Meta-Analysis of Structural and Functional Brain Alterations in Individuals With Genetic and Clinical High-Risk or Psychosis and Bipolar Disorder." *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 117: 110540.

Marek, S., B. Tervo-Clemmens, F. J. Calabro, et al. 2022. "Reproducible Brain-Wide Association Studies Require Thousands of Individuals." *Nature* 603, no. 7902: 654–660. https://doi.org/10.1038/s41586-022-04492-9.

Mason, O., and G. Claridge. 2006. "The Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE): Further Description and Extended Norms." *Schizophrenia Research* 82, no. 2: 203–211. https://doi.org/10. 1016/j.schres.2005.12.845.

McGlashan, T. H., T. J. Miller, S. W. Woods, R. E. Hoffman, L. Davidson, and A Scale for the Assessment of Prodromal Symptoms and States. 2001. *Early Intervention in Psychotic Disorders*, edited by T. J. Miller, S. A. Mednick, T. T. McGlashan, J. Liberger, and J. O. Johannessen, 135– 149. Kluwer Academic Publishers.

McGraw, K. O., and S. P. Wong. 1996. "Forming Inferences About Some Intraclass Correlation Coefficients." *Psychological Methods* 1, no. 1: 30–46.

Modinos, G., M. J. Kempton, S. Tognin, et al. 2020. "Association of Adverse Outcomes With Emotion Processing and its Neural Substrate in Individuals at Clinical High Risk for Psychosis." *JAMA Psychiatry* 77, no. 2: 190–200. https://doi.org/10.1001/jamapsychiatry.2019.3501.

Orsolini, L., S. Pompili, and U. Volpe. 2022. "Schizophrenia: A Narrative Review of Etiopathogenetic, Diagnostic and Treatment Aspects." *Journal of Clinical Medicine* 11, no. 17: 5040.

Pomponio, R., G. Erus, M. Habes, et al. 2020. "Harmonization of Large MRI Datasets for the Analysis of Brain Imaging Patterns Throughout the Lifespan." *NeuroImage* 208: 116450.

Radua, J., E. Vieta, R. Shinohara, et al. 2020. "Increased Power by Harmonizing Structural MRI Site Differences With the ComBat Batch Adjustment Method in ENIGMA." *NeuroImage* 218: 116956. https://doi.org/10.1016/j.neuroimage.2020.116956.

Raine, A. 1991. "The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria." *Schizophrenia Bulletin* 17, no. 4: 555–564. https://doi.org/10.1093/schbul/17.4.555.

Roca, V., G. Kuchcinski, J. P. Pruvo, D. Manouvriez, and R. Lopes. 2025. "IGUANE: A 3D Generalizable CycleGAN for Multicenter Harmonization of Brain MR Images." *Medical Image Analysis* 99: 103388. https://doi.org/10.1016/j.media.2024.103388.

Rosen, M., L. T. Betz, F. Schultze-Lutter, et al. 2021. "Towards Clinical Application of Prediction Models for Transition to Psychosis: A Systematic Review and External Validation Study in the PRONIA Sample." *Neuroscience and Biobehavioral Reviews* 125: 478–492. https://doi.org/10.1016/j.neubiorev.2021.02.032.

Si, S., A. Bi, Z. Yu, et al. 2024. "Mapping Gray and White Matter Volume Abnormalities in Early-Onset Psychosis: An ENIGMA Multicenter Voxel-Based Morphometry Study." *Molecular Psychiatry* 29, no. 2: 496–504.

Strawbridge, R. J., J. Ward, L. M. Lyall, et al. 2018. "Genetics of Self-Reported Risk-Taking Behaviour, Trans-Ethnic Consistency and Relevance to Brain Gene Expression." *Translational Psychiatry* 8, no. 1: 178. https://doi.org/10.1038/s41398-018-0236-1.

Suhail, K., and R. Cochrane. 2002. "Effect of Culture and Environment on the Phenomenology of Delusions and Hallucinations." *International Journal of Social Psychiatry* 48, no. 2: 126–138. https://doi.org/10.1177/ 002076402128783181.

Tognin, S., H. H. van Hell, K. Merritt, et al. 2020. "Towards Precision Medicine in Psychosis: Benefits and Challenges of Multimodal Multicenter Studies-PSYSCAN: Translating Neuroimaging Findings From Research Into Clinical Practice." *Schizophrenia Bulletin* 46, no. 2: 432–441. https://doi.org/10.1093/schbul/sbz067.

van Erp, T. G. M., E. Walton, D. P. Hibar, et al. 2018. "Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium." *Biological Psychiatry* 84, no. 9: 644–654. https://doi.org/10.1016/j.biopsych.2018.04.023. Yung, A. R., A. R. Yung, H. Pan Yuen, et al. 2005. "Mapping the Onset of Psychosis: The Comprehensive Assessment of At-Risk Mental States." *Australian and New Zealand Journal of Psychiatry* 39, no. 11–12: 964–971. https://doi.org/10.1080/j.1440-1614.2005.01714.x.

Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros. 2017. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." In *Proceedings of the IEEE International Conference on Computer Vision (ICCV).* 2242–2251. https://doi.org/10.1109/ICCV.2017.244.

Zuo, L., Y. Liu, Y. Xue, et al. 2023. "HACA3: A Unified Approach for Multi-Site MR Image Harmonization." *Computerized Medical Imaging and Graphics* 109: 102285. https://doi.org/10.1016/j.compmedimag. 2023.102285.