RUNNING TITLE: Psychosis-Spectrum Data Sharing Initiative

Title: Enhancing psychosis-spectrum nosology through an international data sharing initiative

Anna R. Docherty¹,²,³, Eduardo Fonseca-Pedrero⁴, Martin Debbane⁵,⁶, Raymond C. K. Chan⁷,⁸, Richard J. Linscott⁹, Katherine G. Jonas¹⁰, David C. Cicero¹¹, Melissa J. Green¹², Leonard J. Simms¹³, Oliver Mason¹⁴, David Watson¹⁵, Ulrich Ettinger¹⁶, Monika Waszczuk¹⁰, Alexander Rapp¹⁷, Phillip Grant¹⁸,¹⁹, Roman Kolov¹⁰, Colin G. DeYoung²⁰, Camilo J. Ruggero²¹, Nicolas R. Eaton²², Robert F. Krueger²⁰, Christopher Patrick²³, Christopher Hopwood²⁴, F. Anthony O'NeilI²⁵, David H. Zald²⁶, Christopher C. Conway²⁷, Daniel E. Adkins¹,²,²⁸, Irwin D. Waldman²⁹, Jim van Os³⁰,³¹,³², Patrick F. Sullivan³³,³⁴, John S. Anderson¹, Andrey A. Shabalin¹, Scott R. Sponheim²⁰, Stephan F. Taylor³⁵, Rachel G. Grazioplene³⁶, Silviu A. Bacanu², T. Bernard Bigdeli²,²³, Corinna Haenschel³⁷, Dolores Malaspina³⁸, Diane C. Gooding³⁹, Kristin Nicodemus⁴⁰, Frauke Schultzze-Lutter⁴¹, Neus Barrantes-Vidal⁴²,⁴³,⁴⁴, Christine Mohr⁴⁵, William T. Carpenter⁴⁶, Alex S. Cohen⁴⁷

1. Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT, USA.
2. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA.
3. Department of Psychology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA.
4. Department of Psychiatry, Universidad de La Rioja, Oviedo, Asturias, Spain.
5. Research Department of Clinical, Educational, and Health Psychology, University College London, London, UK.
6. Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland.
7. Neuropsychology and Applied Cognitive Neuroscience Laboratory, CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.
8. Department of Psychology, Chinese Academy of Sciences, Beijing, China.
9. Department of Psychology, University of Otago, Dunedin, New Zealand.
10. Department of Psychiatry, Stony Brook School of Medicine, Stony Brook, NY, USA.
11. Department of Psychology, University of Hawaii at Manoa, Honolulu, HI, USA.
12. School of Psychiatry, University of New South Wales, Sydney, Kensington, NSW, Australia.
13. Department of Psychology, University at Buffalo, The State University of New York, Buffalo, NY, USA.
14. Department of Psychology, University of Surrey, Guildford, Surrey, UK.
15. Department of Psychology, University of Notre Dame, Notre Dame, IN, USA.
16. Department of Psychology, University of Bonn, Bonn, Germany.
17. Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany.
18. Department of Psychology, Justus-Liebig-University Giessen, Giessen, Germany.
19. Technische Hochschule Mittelhessen, University of Applied Sciences, Giessen, Germany.
20. Department of Psychology, University of Minnesota, Minneapolis, MN, USA.
21. Department of Psychology, University of North Texas, Denton, TX, USA.
22. Department of Psychology, Stony Brook University, Stony Brook, NY, USA.
23. Department of Psychology, Florida State University, Tallahassee, FL, USA.
24. Department of Psychology, University of California – Davis, Davis, CA, USA.
25. Centre for Public Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast, UK.
26. Departments of Psychology and Psychiatry, Vanderbilt University, Nashville, TN, USA.
27. Department of Psychology, College of William & Mary, Williamsburg, VA, USA.
28. Department of Sociology, University of Utah, Salt Lake City, UT, USA.
29. Department of Psychology, Emory University, Atlanta, GA, USA.
30. Department of Psychiatry and Psychology, Maastricht University Medical Centre, Maastricht, The Netherlands.
31. King’s Health Partners, Department of Psychosis Studies, Institute of Psychiatry, King’s College London, London, UK.
32. Department of Psychiatry, Brain Center Rudolf Magnus Institute, University Medical Center, Utrecht, The Netherlands.
33. Department of Psychiatry, University of North Carolina – Chapel Hill, Chapel Hill, NC, USA.
34. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
35. Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA.
36. Department of Psychology, Yale University, New Haven, CT, USA.
37. Department of Psychology, City University, London, UK.
38. Department of Psychiatry, Icahn School of Medicine, Mount Sinai, New York, NY, USA.
39. Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA.
40. Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.
41. Department of Psychiatry and Psychotherapy, Heinrich-Heine University, Dusseldorf, Germany.
42. Department of Clinical Psychology, Universitat Autònoma de Barcelona, Barcelona, Spain.
43. University of North Carolina at Greensboro, Greensboro, NC; Centre for Biomedical Research
44. Sant Pere Claver - Fundació Sanitària, Barcelona, Spain.
45. Institute of Psychology, University of Lausanne, Lausanne, Switzerland.
46. Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA.
47. Department of Psychology, Louisiana State University, Baton Rouge, LA, USA.

*Corresponding author (Primary appointment): Department of Psychiatry, University of Utah School of Medicine, 501 Chipeta Way, Salt Lake City, Utah 84110, USA
tel. +1 801 213 6905, email: docherty.anna@gmail.com
Abstract

The latent structure of schizotypy and psychosis-spectrum symptoms remains poorly understood. Furthermore, molecular genetic substrates are poorly defined, largely due to the substantial resources required to collect rich phenotypic data across diverse populations. Sample sizes of phenotypic studies are often insufficient for advanced structural equation modeling approaches. In the last 50 years, efforts in both psychiatry and psychological science have moved toward 1) a dimensional model of psychopathology (e.g., the current Hierarchical Taxonomy of Psychopathology [HiTOP] initiative), 2) an integration of methods and measures across traits and units of analysis (e.g., the RDoC initiative), and 3) powerful, impactful study designs maximizing sample size to detect subtle genomic variation relating to complex traits (the Psychiatric Genomics Consortium [PGC]). These movements are important to the future study of the psychosis spectrum, and to resolving heterogeneity with respect to instrument and population. The International Consortium of Schizotypy Research (ICSR) is composed of over 40 laboratories in 12 countries, and to date, members have compiled a body of schizotypy- and psychosis-related phenotype data from over 30,000 individuals. It has become apparent that compiling data into a protected, relational database and crowdsourcing analytic and data science expertise will result in significant enhancement of current research on the structure and biological substrates of the psychosis spectrum. The authors present a data-sharing infrastructure similar to that of the PGC, and a resource-sharing infrastructure similar to that of HiTOP. This report details the rationale and benefits of the phenotypic data collective and presents an open invitation for participation.

Keywords: data sharing; schizotypy; schizotypal; psychosis; schizophrenia; phenotype; genetic; ICSR; HiTOP
Recent progress in psychiatric and psychological science underscores the need for consolidation and meta-analysis of phenotypic and molecular data, to model the latent structure of the psychosis spectrum. Support for this undertaking stems from the inadequacy of categorical diagnoses alone to reflect the apparent spectrum of psychotic disorders, quickly developing dimensional conceptualizations of psychopathology,¹ the high polygenicity of psychosis symptom dimensions² (also Bigdeli et al., unpublished data, 2018), and the selective role of rare variants in conferring risk for psychosis.³⁻⁵

Three initiatives, proceeding largely independently, have brought the field toward a critical juncture in which consolidation efforts are likely to significantly improve our understanding of severe psychopathology: the Psychiatric Genomics Consortium (PGC) has enhanced our genetic understanding of the psychosis spectrum,⁶⁻²⁹ the Hierarchical Taxonomy of Psychopathology consortium (HiTOP) has endeavored to map the latent structure of psychosis,¹,³⁰ and the NIMH Research Domain Criteria (RDoC) initiative has endeavored to develop crosswalks between multiple units of analyses (e.g., behavioral paradigms relevant to schizotypy measures).³¹⁻³⁸ Consolidation efforts are also consistent with the translational aims of the Roadmap for Mental Health Research in Europe (ROAMER) project.³⁹

As van Os et al. have recently asserted, the psychosis spectrum requires careful reconstruction.⁴⁰ The International Consortium for Schizotypy Research (ICSR) convened in March of 2017 to discuss current research and ways to improve understanding of dimensionality and discontinuity in schizotypy and risk for psychosis. The steering committee moved to collectively amass data and secured collaborations with the PGC and HiTOP to ensure informed data consolidation, reflecting strategies implemented by the PGC.⁴¹ This report details further the rationale for data sharing, the advantages it provides to collaborators, and the process by which we hope to achieve PGC-, HiTOP-, and RDoC-related aims. Broadly, the current goal of the ICSR is to create a data resource that will continue to grow and lead to discoveries which inform biology and nosology, improve assessment, and identify treatment targets.

**Rationale for ICSR Data Sharing**

There are several important reasons to amass phenotypic data on the psychosis spectrum. There is some consensus that the current concept of “schizophrenia,” described by diagnostic guidelines and later reified, confines research to a constantly changing “construct that does not exist.”⁴² Research on
schizotypy, i.e., the latent diathesis for psychosis and psychosis-spectrum disorders, and schizotypal signs and symptoms has addressed some of the problems of ‘reification’ by characterizing cognitive and emotional facets of these symptoms across populations, and by comparing categorical high-risk states with symptoms in non-clinical, healthy populations. But because categorical conceptualizations of high-risk states and psychometrically-identified schizotypy can be similarly ‘reified’, symptom dimensions should be empirically evidenced and mapped more comprehensively across a broad network of phenotypes, with careful consideration of the differences between phenomena and symptoms as well as assessments. Due to the heterogeneous nature of assessments, limited sampling and insufficient statistical power to conduct appropriate structural equation modeling, this issue must be addressed with mass collaboration.

1. **Individuals identified by current psychometric approaches appear to represent a small fraction of a heterogeneous spectrum phenotype.**

Psychometrically-identified high-risk groups, based on arbitrary cut-offs, have restricted research to the narrow view of psychotic experiences and/or extreme anhedonia (e.g., ‘Ultra-High Risk’), despite evidence from genomic research that such individuals are a small sample of the psychosis spectrum. Subthreshold psychosis spectrum symptoms should be redefined and supplemented to improve prediction of actual onset of psychosis in the general population.

The structure of psychosis and related sequelae, within a hierarchical model such as HiTOP, remains relatively undefined compared with other dimensional components of personality and internalizing/externalizing disorders. Addressing this concern requires enhanced quantitative approaches to refining what we consider to be schizotypal traits, ideally involving network, longitudinal growth curve, and machine learning approaches—all of which are impossible with the currently limited availability of psychosis-spectrum phenotypic data. This also requires careful attention to the constructs involved and their conceptualization.

2. **Psychiatric traits and symptoms are genetically complex, and light phenotyping is inherent to large genomic efforts**
Using very large samples collected for genomic mega-analysis, we experience the drawbacks of necessarily lighter phenotyping—dramatically abbreviated scales, or even the use of a single item—and a reduced diversity of clinical and behavioral data. It has become apparent that very small numbers of items are needed to economically test genomic relationships in large epidemiological studies, and the PGC working groups are interested in careful psychometric validation of such items. With many different datasets, measures, methods, and populations compiled for side-by-side comparison, the field is better able to identify effective items using methods such as CFA and IRT. Moreover, items may be tailored to culture or clinical population (case, pedigree, college student). One deliverable advance stemming from this effort is developmental and testing support from involved PIs for a web application that may be used in large-scale epidemiological studies. Not only will replicable findings on the validity and utility of items be useful to the PGC, but these efforts will in turn inform phenotypic measurement.

3. **Schizotypy may easily vary by genomic profile, and rare genomic features could isolate key symptom dimensions**

Genetic subtyping of psychosis-spectrum disorders is highly desirable if it can lead to more accurate classification, early prediction, and effective pharmacological interventions. We observe in genomic psychosis-spectrum research that 1) traits are highly polygenic, and yet 2) specific rare variants result in psychosis despite low genome-wide polygenic risk for schizophrenia.

Genetic subtyping of complex psychiatric traits has been slow to develop, but is moving forward in autism spectrum disorder research, where many probands in dense pedigrees inherit rare variants which disproportionally affect cognitive ability. It is possible that probands with rare mutations will have symptoms similar to those with more typical genetic profiles, but this is not a certainty, and the degree to which probands are atypical is facilitated by modeling the genetic and phenotypic heterogeneity of individuals with autism spectrum disorders. The same can be said for schizophrenia. The genome is increasingly examined as a dimensional measure (rather than inspected for genome-wide significant hits) to characterize schizophrenia risk, and this is leading to findings relevant to classification. These methods, although advancing, remain hindered by over-reliance on categorical diagnosis. We can facilitate informed decisions about clinically meaningful differences within the psychosis spectrum after
proper symptom data consolidation efforts across multiple research groups, integrating family history, personality, and developmental data.

4. **Dimensional data are statistically powerful, and some categorical measures can harmonize with dimensional measures.**

It is now established that schizotypal symptom expression, as currently measured, is detectable across the general population, in biological relatives, and in probands, and can be better characterized using a quantitative dimensional approach than a dichotomous distinction made with arbitrary cut points.\(^{58,67}\) To date, the ‘Ultra High Risk’ characterization has had moderate clinical utility, and ‘schizotypy’ categorical distinctions have been useful insofar as measures have leveraged several phenotypes at once to gather additional evidence of dimensionality across healthy and clinical populations.\(^{68-81}\) Again, assumptions that the endophenotype is any less complex or heterogeneous than schizophrenia itself should be avoided,\(^{82}\) and can lead to premature attempts at parsing symptom factors. Reviews of the literature that integrate studies using dimensional measures have proven to be a promising start,\(^{58,83-87}\) but with harmonization of measures across large numbers of samples, and proper assessment of measurement invariance, we can begin to build statistically powerful models of causation and genome/environment interactions.

In psychiatry, the movement toward a dimensional framework for defining and diagnosing psychiatric illness is well established and appears to be a more reliable and comprehensive model than the DSM-5 and ICD-10 categorical framework.\(^1,88\) However, psychosis-spectrum conditions prove problematic in the typical internalizing/externalizing dimensional framework, leading researchers to propose and test a distinct psychosis dimension.\(^{30,89,90}\)

There is a need for clarity regarding how this psychosis dimension ought to be structured. One camp suggests a return to the model which preceded Kraepelin’s classification of 2 types of psychosis.\(^{91}\) This would create a single, unifying dimension of psychosis which encompasses both of Kraepelin’s distinctions,\(^{92}\) and indeed there is ample evidence to suggest that the conditions of psychosis share common genetic and environmental factors.\(^{93}\) However, a singular psychosis dimension may also fail to capture the complexity of a given condition. Multiple studies have found that a 5-dimensional structure, including dimensions labeled *internalizing, disinhibited externalizing, antagonistic externalizing,*
detachment, and thought disorder, better harmonizes with existing categorical diagnoses.\textsuperscript{94-96} Yet other researches have sought to blend the parsimony of a single dimension with the nuance of a 5-dimensional structure using a bi-factor model in which a general psychosis dimension is assessed first, and then used as a guide for assessment by the 5 specific symptom dimensions.\textsuperscript{97,98} The efforts of ICSR will further inform these findings and evaluate current proposed models.

5. The clinic: Current high-risk classification approaches are not sufficient to understand risk

Field evidence has yet to justify a DSM risk syndrome category, with only 11% meeting criteria for UHR classification developing psychosis in one study,\textsuperscript{99} and 39% in another.\textsuperscript{100} Inclusion of an attenuated psychosis syndrome into the full text of DSM-5.1 is still being debated.\textsuperscript{101} The best psychometrically-guided predictors in non-clinical populations, outside of family history, are extreme scores on symptom surveys, and even they do not predict actual psychosis at high rates. Negative schizotypy studies find an impressive 24% of students with (categorical) extreme social anhedonia develop schizophrenia-spectrum psychopathology, but not typically psychosis. In general, schizotypy measures in predicting psychosis have been underwhelming (for review of this literature, see \textsuperscript{85}). Family and molecular genetic studies have provided evidence that dimensionally-measured negative symptoms may hold predictive utility\textsuperscript{55, 80, 102, 103, 68, 104, 105} but again, phenotyping in genomic studies thus far has been light, or samples too small, to adequately examine the genomic architecture of psychosis-related symptom dimensions.

The take-home message of this research is 1) family history predicts general psychopathology, 2) subthreshold symptoms of a psychosis-spectrum disorder predict symptoms of the disorder, 3) there is little diagnostic specificity with regard to prediction, and 4) the more prevalent the disorder, the greater role environment plays in etiology. Given these points, perpetuating the literature on clinically-measured high risk without refining the phenotype is unlikely to improve research on early intervention.

One function of the ICSR can be to better operationalize phenotypes in accordance with dimensional models, and to improve understanding of the relation of schizotypy to other psychosis spectrum phenomenology, e.g., of UHR and frank psychosis. Language, social behavior, emotional expression, beliefs, perceptual experience and emotional response can all be reduced to behavioral function, and these can be normed using large international samples. Relatedly, an advantage of the effort is to explore
the role of culture on illness expression and phenotype. The ICSR is well poised to accomplish this, given that it is truly international and intercultural.

The NIMH, PGC, HiTOP, and ICSR are moving toward a framework that is more compatible with dimensional biological risk, and the UHR/FHR research community is encouraged to collaborate with these efforts to improve clinical outcomes. Importantly, the addition of dimensional measures is meant to enhance our understanding of the latent structure of psychosis and psychosis risk, but not mire us in assumptions about diagnosis or dimension.

**An Open Invitation**

Three primary steps of this initiative are illustrated in Figure 1. Data distributions are examined and large matrices of data used to develop empirically-driven covariance structures. Models will implicate facets of schizotypy and psychosis-related symptoms relevant to specific intervention targets, and will be assessed relative to genomic findings. These data will include both clinical and psychometric measures. With larger sample sizes, CFA, SEM, IRT, machine learning and other relevant methods may be implemented to validate models of latent structure.

The ICSR requires a massive, collective effort to obtain and maintain data to facilitate efficient analysis and replication. This initiative is facilitated in part by Anna Docherty's cluster farm and by the Utah Center for Genomic Discovery (UCGD), a University of Utah initiative to integrate patient genome information into health care and develop tools for genome interpretation. Together this accounts for 2.5 PB of disc storage. There is a core team of six on-site data scientists and analysts, including authors Anna Docherty, John Anderson, Andrey Shabalin, and Daniel Adkins. Computing space, resources for proband and extended pedigree data collection, and an active undergraduate data collection are available to collaborative PIs. The authors of this publication themselves share broad skill sets and actively seek collaboration with other interested PIs.

"Schizotypy" is a complex, multidimensional construct, whose dimensions differ widely both in the degree and specificity with which they reflect genetic liability to schizophrenia. Thus to date, our consolidated data come from multiple community, risk, family, and case populations. 30 of our collaborators share clinical and/or cognitive data from first-degree biological relatives, 19 share college
student data, and 21 share community data. Most have also collected data in cases, with measures that can be harmonized and meta-analyzed. The ICSR has amassed approximately 40,000 samples with clinical phenotype data including items from schizotypy and schizotypal personality assessments. We hope to double samples over the following year and plan to apply for additional external funding. These data and measures will be characterized in a first publication with all collaborators.

A portal for participation is housed on the ICSR website, at srconsortium.org. PIs or institutions may use this website to contact the data sharing facilitators. Participation includes invitation to collaborate on analyses and publication of scientific findings. Projects will be managed in the same way data analytics are handled within PGC, such that individuals and research teams submit proposals to analyze data. One example analysis by the ICSR and PGC may be focused on item refinement for future genomic research using smartphone applications. Another study will pilot items in a genetic study of undergraduate samples.

We draw attention to the precedent of the PGC to effectively organize and collaborate on a large scale. This is done with adherence to common principles described in the inaugural publication. Participating members will use identical guidelines and are in active consultation with the PGC Director Patrick Sullivan to promote external review of the collaborative.

With the introduction of spectrum phenotypes in DSM5, the field is better positioned to synthesize research to refine signs and symptoms. Impactful psychosis research in the present day reflects collective efforts to understand 1) the genetic architecture of psychosis, and 2) the latent structure of the psychosis spectrum. These efforts can be symbiotic and benefit both genomic and phenotypic research.
Cited References


3. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science.* 2008;320(5875):539-543.


49. Linscott RJ, Morton SE. The latent taxonicity of schizotypy in biological siblings of probands with schizophrenia. *Schizophr Bull.* 2017. doi: 10.1093/schbul/sbx143


