PSILOGOS

Research Article

open-access

Protocol-Distribution Geometry and the Robustness of Transdiagnostic Cortical Concordance

James Wilson

ABSTRACT

Cross-disorder magnetic resonance imaging can reveal shared cortical organization, but diagnostic groups are rarely distributed uniformly across acquisition protocols. This study asks whether protocol-level recruitment overlap is sufficient to explain the reported resemblance of cortical thickness and surface-area alteration patterns among schizophrenia, bipolar disorder, major depressive disorder, and autism spectrum disorder. The analytic matrix comprised 5,549 individuals allocated across 31 scanner–protocol groups. The primary case–control audit retained 29 protocols with healthy comparators, representing 5,500 individuals; the remaining two protocol rows were retained in a sensitivity calculation. Protocol-distribution concordance analysis combined normalized recruitment entropy, effective protocol number, positive-support Gini concentration, pairwise Hellinger affinity, leave-one-protocol-out deletion, and an exact $4!$ disorder-label permutation test. Schizophrenia had the broadest recruitment support (23 protocols; effective number 14.080), whereas bipolar disorder, major depressive disorder, and autism spectrum disorder had effective protocol numbers of 5.568, 6.910, and 5.279, respectively. Pairwise recruitment affinity ranged from 0.123 for bipolar disorder–autism spectrum disorder to 0.766 for bipolar disorder–major depressive disorder. This ordering did not mirror cortical concordance. Recruitment affinity had a weak rank association with cortical-thickness similarity ($\rho = 0.257$, exact $p = 0.667$) and a near-zero association with surface-area similarity ($\rho = 0.086$, exact $p = 0.958$). Deleting one protocol at a time changed the thickness association to 0.029–0.600 and the surface-area association to 0.086–0.371, without reversing the central inference. Inclusion of both protocols lacking healthy comparators left both observed rank correlations unchanged. Protocol co-sampling therefore does not provide a sufficient account of the transdiagnostic cortical pattern. The result supports biological interpretation of the concordance while defining a concrete acquisition-allocation uncertainty that future individual-level analyses should estimate directly.

Keywords: cortical thickness; cortical surface area; multisite MRI; psychiatric disorders; Hellinger affinity; recruitment concentration; sensitivity analysis

Department of Psychological Medicine, School of Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand
j.wilson587@gmail.com
– **Published:** 15 December 2024

1. INTRODUCTION

Large collaborative neuroimaging programs have transformed psychiatric morphometry from collections of small, locally processed samples into coordinated analyses spanning scanners, institutions, and countries. Their principal advantage is not simply a larger total number of observations. They also permit disorder effects to be examined against a broader range of ages, clinical settings, hardware platforms, pulse sequences, and recruitment practices. The ENIGMA collaboration demonstrated how common processing and statistical procedures can yield stable estimates across geographically dispersed cohorts.[1] National collaborations have pursued the same goal by aligning quality control, image processing, and clinical definitions across institutions.[2] This expansion is necessary because brain-wide associations are usually modest and because estimates obtained from limited samples can vary substantially.[3] At the same time, the multisite structure introduces a second question: whether a cross-disorder resemblance reflects cortical biology, acquisition allocation, or some combination of the two.

Transdiagnostic structural studies have consistently found that conventional diagnostic boundaries do not partition cerebral morphology into wholly distinct patterns. Meta-analytic comparisons across major psychiatric conditions indicate both convergent and diagnosis-selective cortical alterations.[4] Reviews of ENIGMA findings similarly show that schizophrenia, bipolar disorder, and major depressive disorder share widespread reductions in cortical measures, although their magnitude and regional extent differ.[5] White-matter investigations also reveal cross-disorder commonality, with schizophrenia often showing the largest departure from healthy comparison groups.[6] These observations are compatible with overlapping genetic liability, partially shared developmental processes, common environmental exposures, medication effects, and nonspecific consequences of chronic illness. They also create a methodological obligation: shared anatomy should be separated from shared recruitment structure as far as the available information permits.

Disorder-specific evidence provides an important reference for this problem. Schizophrenia is associated with extensive cortical thinning and reduced surface area across large international samples.[7] Bipolar disorder shows a more restricted and generally smaller cortical-thickness alteration pattern.[8] Major depressive disorder is characterized by subtle, regionally distributed differences that depend on age and clinical composition.[9] Autism spectrum disorder shows a heterogeneous developmental profile, including age-dependent variation in cortical thickness and surface area.[10] When these diagnoses are compared, a high similarity between two cortical maps may indicate common pathophysiology, but it may also arise if the diagnostic groups were measured disproportionately on the same scanners or within the same institutions. Conversely, two disorders could have similar biology yet show attenuated map similarity if their participants were recruited through largely separate protocol sets.

Several biological findings make the cortical comparison especially consequential. Cell-type enrichment analyses have linked transdiagnostic thickness patterns to shared neurobiology across multiple psychiatric conditions.[11] Imaging-genetic syntheses suggest that common genetic liability is expressed differently across cortical and subcortical systems.[12] Large-scale cortical gradients capture axes along which diagnosis effects converge or diverge,[13] while molecular and connectomic analyses relate local cortical alterations to brain-wide organization.[14] Cortical thickness and surface area also have partly distinct genetic architectures[15] and developmental determinants.[16] A recruitment-allocation account would therefore have different implications for the two phenotypes. If thickness concordance followed protocol overlap but surface-area concordance did not, methodological structure could contribute unequally to their apparent transdiagnostic organization. If neither phenotype followed protocol overlap, simple co-sampling would be an inadequate explanation.

Prior cross-disorder analyses have compared subcortical volume, regional thickness, and surface area across developmental and psychiatric conditions.[17] Their statistical machinery usually concentrates on site-wise effect estimation, meta-analytic pooling, covariate adjustment, and multiple-comparison control. These procedures are appropriate for estimating a diagnostic contrast, yet the resulting paper-level summary rarely quantifies how diagnostic recruitment is distributed over protocols. Scanner harmonization methods can remove additive and multiplicative site effects under explicit assumptions,[18] and lifespan applications demonstrate their value in large heterogeneous collections.[19] Nevertheless, harmonization cannot make an absent diagnosis–protocol combination observable. A disorder recruited at only a few protocols has a narrower empirical acquisition support than a disorder recruited broadly, even when each site applies identical image processing.

Scanner allocation deserves separate attention because it can operate without a simple mean shift. A classifier may exploit scanner features when diagnostic composition is imbalanced across sites. A meta-analysis may reduce this risk by estimating contrasts within protocol, but the transdiagnostic comparison of the resulting maps can still join disorder estimates supported by different subsets of the acquisition network. Real-world multisite studies have shown that scanner differences materially affect structural MRI discrimination unless they are addressed explicitly.[20] The relevant issue is therefore not that multisite research is unreliable. Rather, the allocation matrix is part of the design and should be described quantitatively, just as sample size, age, sex, and quality-control exclusions are described.

Allocation imbalance has at least three forms that require different summaries. Support imbalance occurs when one diagnosis is absent from many protocols. Concentration imbalance occurs when a diagnosis is present broadly but most of its participants come from a small number of protocols. Pairwise imbalance occurs when two diagnoses are both well represented yet occupy different portions of the acquisition network. A protocol count, a total sample size, or a list of participating institutions cannot distinguish all three. Support size describes the first, entropy and effective protocol number describe the second, and a distributional affinity describes the third. Considering these quantities together avoids reducing a complex design to a single site count.

This distinction is particularly important for map similarity. A diagnostic case–control effect is estimated relative to HC within the protocols that contain that diagnosis. The resulting regional vector is then compared with a vector for another disorder. Even if both case–control estimates are internally valid, the pairwise comparison can involve different protocol mixtures. Shared protocol support is therefore neither a prerequisite for biological resemblance nor a guarantee of it. It is an observable design feature whose explanatory adequacy can be tested against the cortical ordering.

The protocol matrix also offers a useful intermediate level between individual images and paper-level totals. It contains no voxel or vertex information, but it preserves the acquisition allocation that is lost when only diagnostic sample sizes are reported. Such an audit can be completed when individual records are unavailable, and it can guide more demanding analyses if protocol-specific effect maps later become accessible. The method is deliberately modest in what it claims: it tests a necessary empirical pattern for a simple co-sampling explanation, not every possible scanner-related mechanism.

The present study treats the protocol-by-diagnosis count table as an analyzable design object. It asks a focused question: does similarity in protocol-level recruitment distributions track the published similarity of regional cortical alteration patterns? The question differs from estimating which diagnosis has the thinnest cortex or the smallest surface area. It evaluates whether the acquisition support itself has the same pairwise geometry as the cortical findings. Three predictions make the test falsifiable. First, a strong acquisition-allocation explanation would require disorder pairs with high recruitment affinity to show high cortical similarity. Second, disorder pairs recruited on largely separate protocol distributions should show low cortical similarity. Third, these relationships should persist when any single protocol is removed. The analysis evaluates all six disorder pairs for both cortical thickness and surface area, uses an exact label test suited to four diagnoses, and reports deletion ranges rather than relying on a single correlation.

Anatomically, thickness describes the distance between the white-matter and pial surfaces, whereas surface area reflects the extent of the cortical sheet. Their different developmental and genetic determinants make a phenotype-specific design audit necessary. The cortical rendering in Fig. 1 provides anatomical orientation only; it contains no participant measurement and no encoded result. Its purpose is to distinguish the folded cortical sheet from the protocol-allocation quantities evaluated later.

Figure 1. Digital cortical anatomy.

The visual separation between anatomy and recruitment design is substantive. Cortical concordance is a property of regional effect vectors, whereas recruitment affinity is a property of participant allocation across protocols. Similar numerical values do not make the two quantities interchangeable. Their association, rather than their absolute scale, is the object of the present test.

2. MATERIALS AND METHODS

The design has two linked levels. The first level is a protocol-by-diagnosis matrix whose rows identify scanner–protocol groups and whose columns contain healthy comparison subjects (HC), schizophrenia (SZ), bipolar disorder (BD), major depressive disorder (MDD), and autism spectrum disorder (ASD). The second level contains the six unordered diagnostic pairs and their cortical map similarities. Recruitment calculations use counts rather than individual MRI measurements. Cortical calculations use the pairwise cosine similarities of 68-region alteration vectors. This separation prevents participant counts from being treated as cortical observations.

The full matrix contains 31 protocol rows and 5,549 individuals: 3,068 HC, 1,426 SZ, 237 BD, 612 MDD, and 206 ASD. These values, the protocol labels, the two protocol rows without HC, and the cortical similarities were transcribed from the attached article.[26] The primary audit followed the case–control eligibility stated in that article and removed Hiroshima2 (42 MDD participants and no HC) and Kyushu3 (7 BD participants and no HC). The resulting matrix contained 29 protocols and 5,500 individuals, including 2,432 clinical participants. A sensitivity calculation restored the two excluded rows and recomputed all pairwise recruitment affinities.

2.1. Cortical quantities retained from coordinated morphometry

The cortical quantities were produced from T1-weighted structural MRI processed with the Desikan–Killiany parcellation[21] and FreeSurfer.[22] The coordinated procedure estimated within-protocol diagnostic effects with age and sex adjustment, expressed contrasts as unbiased standardized mean differences, pooled the protocol estimates by meta-analysis, and controlled regional multiplicity. Standardized effects are useful in this setting because they put measures from different protocols on a common scale while retaining the direction of the diagnostic contrast.[23]

For each disorder and phenotype, the 68 regional effects form a vector. The similarity between disorders a and b is the cosine of the angle between their vectors,

$$C_{ab} = \frac{\boldsymbol{d}_a^{\top}\boldsymbol{d}_b}{\|\boldsymbol{d}_a\|_2\|\boldsymbol{d}_b\|_2}, \tag{1}$$

where $\boldsymbol{d}_a$ and $\boldsymbol{d}_b$ contain regional standardized effects. Values near 1 denote similarly oriented alteration patterns, values near 0 denote little directional correspondence, and negative values denote opposite orientation. The analysis uses separate C_{ab} values for cortical thickness and surface area. It does not combine these phenotypes into one score.

2.2. Protocol-distribution concordance analysis

Let x_{pd} denote the number of participants with diagnosis d recruited under protocol p, and let $n_d = \sum_p x_{pd}$. The normalized recruitment distribution is

$$\pi_{pd} = \frac{x_{pd}}{n_d}. \tag{2}$$

The support size K_d is the number of protocols for which $x_{pd} > 0$. Four complementary quantities describe concentration. The maximum protocol share, $m_d = \max_p \pi_{pd}$, identifies the largest single contribution. Normalized entropy is

$$H_d^{*} = -\frac{\sum_{p:x_{pd}>0} \pi_{pd}\log(\pi_{pd})}{\log K_d}, \tag{3}$$

which equals 1 when the diagnostic sample is equally distributed over its observed support. The effective protocol number is

$$K_d^{\mathrm{eff}} = \left(\sum_p \pi_{pd}^{,2}\right)^{-1}, \tag{4}$$

the reciprocal concentration of the distribution. It can be read as the number of equally sized protocols that would produce the observed concentration. The Gini coefficient was calculated over positive counts only. This convention separates inequality among contributing protocols from the number of zero cells, which is already represented by K_d.

Pairwise recruitment similarity was measured with Hellinger affinity,

$$A_{ab} = \sum_p \sqrt{\pi_{pa}\pi_{pb}}, \tag{5}$$

where 0 indicates disjoint protocol support and 1 indicates identical normalized distributions. Affinity is bounded, symmetric, and well defined when many cells are zero. The count of co-recruiting protocols, $J_{ab} = \sum_p I(x_{pa} > 0, x_{pb} > 0)$, was also retained because two pairs can have similar affinity while differing in the number of jointly represented protocol rows.

Normalization makes the affinity insensitive to unequal diagnostic totals. This is essential here because the clinical samples range from 206 to 1,426 individuals. Without normalization, any dot product of count vectors would be dominated by SZ simply because it is larger. The square-root transformation also moderates the leverage of the largest protocol rows while retaining exact zeros for absent diagnosis–protocol combinations. Affinity therefore asks whether two diagnostic samples have similar relative allocation, not whether they have similar size.

No threshold was imposed to classify an affinity as high or low. Interpretation relied on the observed ordering and on its relationship with cortical similarities. This choice avoids treating an arbitrary cut point as a biological boundary. The deletion intervals provide additional scale: if an affinity changes sharply when one protocol is removed, its full-matrix value has limited design stability even when its absolute magnitude appears substantial.

The association between acquisition allocation and cortical concordance was estimated across the six unordered disorder pairs with Spearman's rank correlation. Separate coefficients were calculated for thickness and surface area. Rank correlation was selected because the sample consists of only six pairs, the cortical values include near-zero and negative observations, and no linear relationship was assumed. The six pairwise observations are not statistically independent because every disorder appears in three pairs. Inferential emphasis therefore rests on an exact label calculation and deletion stability rather than an asymptotic correlation test.

2.3. Exact label permutation and protocol deletion

The label calculation enumerated all $4! = 24$ assignments of the four diagnostic labels to the four recruitment distributions. For each assignment, the six affinities were reconstructed and correlated with the fixed cortical similarities. The two-sided exact probability was

$$p_{\mathrm{exact}} = \frac{1}{24}\sum_{b=1}^{24} I\left(|\rho_b| \geq |\rho_{\mathrm{obs}}|\right). \tag{6}$$

Exhaustive enumeration avoids random Monte Carlo variation. Permutation reasoning is widely used in neuroimaging when distributional approximations are doubtful.[24] Here it tests whether the observed alignment is exceptional relative to other assignments of recruitment distributions to diagnostic labels. It does not test whether the cortical maps themselves differ from zero.

Deletion stability was evaluated by omitting one of the 29 eligible protocols, renormalizing every diagnostic distribution, recomputing all six affinities, and recalculating both rank correlations. Twenty-nine deletion values were therefore obtained for each phenotype. Pair-specific minimum and maximum affinities were also recorded. A stable acquisition-allocation account would require both a substantial observed correlation and consistently similar deletion values.

2.4. Quality checks and reproducibility

Column sums were checked against the published totals before analysis. The two excluded protocol rows were verified to contain no HC. All pairwise values were generated in a fixed diagnosis order (SZ–BD, SZ–MDD, SZ–ASD, BD–MDD, BD–ASD, MDD–ASD). Computations were implemented in the included Python script using only deterministic arithmetic; the jitter in Fig. 7 affects display positions but not numerical values. The project contains the count matrix, derived comma-separated tables, analysis script, plotting script, figure files, and compiled manuscript.

Body motion can alter estimated cortical thickness, particularly in heterogeneous developmental samples.[25] The present design audit cannot reconstruct image-level motion or scanner effects, so it does not substitute for image quality control or harmonization. Instead, it quantifies a distinct property: how broadly and similarly the diagnostic groups occupy the protocol network.

3. RESULTS AND DISCUSSION

3.1. Recruitment topology across protocols

The full matrix reproduced all column totals exactly. Restriction to protocols with HC removed 49 clinical participants, leaving all 1,426 SZ participants and all 206 ASD participants, 230 of 237 BD participants, and 570 of 612 MDD participants. The bubble census in Fig. 2 displays the resulting allocation. SZ occupied 23 of 29 eligible protocols and was the only clinical group represented across most of the protocol list. BD occurred in nine protocols, MDD in 14, and ASD in eight. Several rows recruited only one clinical diagnosis alongside HC, while Tokyo5 was the sole eligible protocol containing all four diagnoses.

The census reveals why total sample size alone is insufficient for judging transdiagnostic comparability. SZ had both the largest clinical sample and the broadest support, whereas ASD had a much smaller sample concentrated in eight protocol rows. BD and MDD overlapped substantially at Hokkaido1, yet their remaining counts were distributed differently. Pairwise cortical comparisons thus join effects estimated over related but nonidentical acquisition supports.

Table 1. Recruitment concentration by diagnosis.

	n	K	m	H^*	K^{eff}	Gini
SZ	1426	23	0.151	0.906	14.080	0.421
BD	230	9	0.339	0.889	5.568	0.369
MDD	570	14	0.296	0.845	6.910	0.481
ASD	206	8	0.320	0.886	5.279	0.381

K, positive-support protocols; m, maximum protocol share; H^*, normalized entropy; K^{eff}, effective protocol number. Gini is calculated over positive protocol counts.

The concentration values in Table 1 sharpen that visual impression. SZ had a maximum protocol share of only 0.151 and an effective protocol number of 14.080. The other diagnoses had maximum shares near 0.30 or higher and effective numbers between 5.279 and 6.910. MDD had the lowest normalized entropy (0.845) and the largest positive-support Gini coefficient (0.481), indicating that its 14 contributing protocols were comparatively unequal in size. BD had fewer contributing protocols but a lower positive-support Gini because its nonzero counts were less unequal once zeros were excluded.

The Lorenz representation in Fig. 3 compares these distributions over all 29 eligible protocols. SZ departs least from the equality diagonal because its participants are distributed across many rows. The remaining curves remain flat through long portions of the protocol ordering, reflecting absent diagnosis–protocol combinations, and then rise rapidly among their largest contributors. This geometry matters because a high total n cannot compensate for a narrow support if the concern is transportability across acquisition settings.

These findings do not imply that the cortical estimates are invalid. The within-protocol contrast and meta-analytic pooling used in coordinated studies directly reduce scanner confounding. The allocation audit instead identifies unequal evidentiary breadth. A diagnosis represented across 23 protocols has more opportunities for protocol-specific deviations to average out than one represented across eight. This distinction is especially relevant when comparing map direction rather than testing each disorder against HC independently.

3.2. Pairwise recruitment affinity

The pairwise affinity matrix in Fig. 4 shows a structured but nonhierarchical pattern. BD–MDD had the greatest affinity at 0.766, followed by SZ–BD at 0.462, SZ–ASD at 0.452, and SZ–MDD at 0.437. MDD–ASD had an affinity of 0.222, while BD–ASD had the smallest value at 0.123. Thus, SZ was moderately aligned with every other recruitment distribution, consistent with its broad protocol coverage, whereas ASD had sharply different relationships with BD and SZ.

The matrix also demonstrates that protocol overlap is

Figure 2. Protocol recruitment census.

not equivalent to co-occurrence count. SZ and ASD co-occurred in seven protocols and reached an affinity of 0.452 because several of their largest contributions came from the same Osaka, Tokyo, and Nagoya protocols. BD and MDD also co-occurred in seven protocols but reached 0.766 because their normalized counts were more similarly concentrated, particularly at Hokkaido1 and several shared mixed-diagnosis protocols. An affinity calculation therefore captures relative allocation rather than merely counting shared rows.

The individual protocol contributions explain several less obvious entries. SZ–MDD had the largest co-protocol count, nine, yet its affinity was only 0.437 because the SZ distribution includes many additional protocols and the MDD distribution is concentrated at Hokkaido1 and selected Hiroshima protocols. SZ–ASD shared fewer total participants than SZ–MDD, but their proportional allocation across several Osaka and Tokyo rows produced a slightly larger affinity. BD–ASD had two co-recruiting protocols, Tokyo5 and Nagoya1, and the relative BD mass at other sites kept its affinity low. These details show why presence–absence overlap and proportional overlap answer different design questions.

The numerical comparison in Table 2 anticipates the main result. The pair with the greatest recruitment affinity, BD–MDD, had high thickness concordance but low surface-area concordance. MDD–ASD had relatively low recruitment affinity yet high surface-area concordance. SZ–ASD and SZ–BD had nearly equal recruitment affinities but radically different thickness similarities. These contrasts are incompatible with a monotonic acquisition-allocation explanation that applies uniformly to both cortical phenotypes.

Table 2. Pairwise recruitment and cortical concordance.

Pair	Co-protocols	Affinity	Deletion minimum	Deletion maximum	Thickness cosine	Area cosine
SZ–BD	7	0.462	0.386	0.502	0.943	0.047
SZ–MDD	9	0.437	0.356	0.474	0.959	0.945
SZ–ASD	7	0.452	0.367	0.548	−0.021	0.867
BD–MDD	7	0.766	0.658	0.817	0.934	0.119
BD–ASD	2	0.123	0.058	0.152	0.047	−0.098
MDD–ASD	4	0.222	0.169	0.269	0.002	0.811

Deletion limits are Hellinger affinities after omitting one eligible protocol. Cortical values are cosine similarities of 68-region effect vectors.

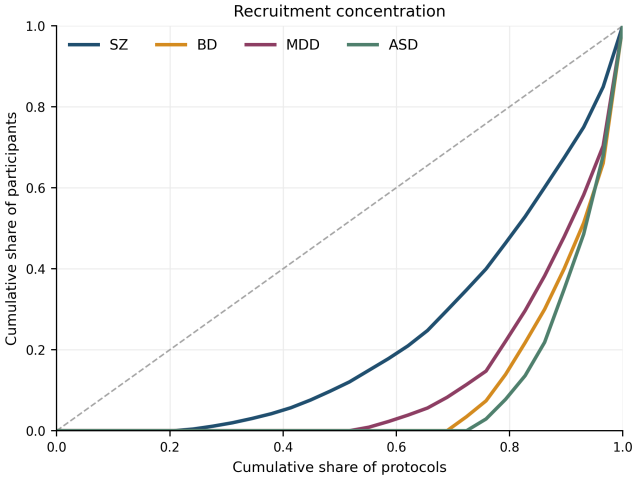

Figure 3. Recruitment concentration.

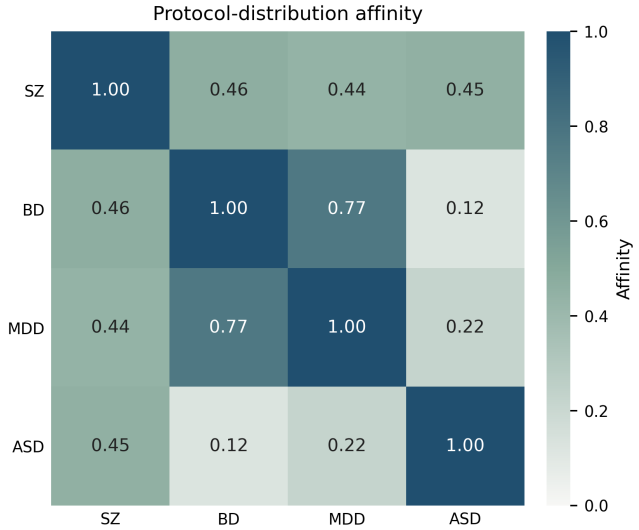

Figure 4. Pairwise protocol affinity.

3.3. Cortical thickness concordance

Thickness similarities formed two distinct groups. SZ–BD, SZ–MDD, and BD–MDD had cosine values of 0.943, 0.959, and 0.934, respectively. All three ASD pairs were close to zero: −0.021 with SZ, 0.047 with BD, and 0.002 with MDD. In Fig. 5, horizontal segments show the minimum and maximum recruitment affinity obtained by deleting one protocol, while points show the full primary matrix. The high-thickness cluster spans substantially different recruitment affinities, from 0.437 to 0.766. The near-zero thickness cluster likewise contains affinities from 0.123 to 0.452.

The observed rank correlation between recruitment affinity and thickness similarity was $\rho = 0.257$. Its exact two-sided label probability was 0.667, so the observed ordering was common among the 24 possible assignments of recruitment distributions to diagnoses. The coefficient is positive because BD–MDD ranks highly on both quantities, but other pairs break the ordering. SZ–MDD has the greatest thickness similarity despite only moderate recruitment affinity, while SZ–ASD has similar recruitment affinity but virtually no thickness concordance.

This result bears directly on the research question. Cosampling may still contribute to individual estimates, especially if unmodeled scanner effects remain, but the protocol count geometry does not reproduce the transdiagnostic thickness geometry. The sharp diagnostic separation involving ASD occurs despite moderate SZ–ASD recruitment affinity. Conversely, the shared SZ–BD–MDD thickness pattern persists across pairs whose recruitment affinities differ by more than 0.3. Protocol overlap alone is therefore insufficient to generate the observed grouping.

The biological interpretation remains plausible because thickness is influenced by cellular, myelination, pruning, and developmental processes that may be shared among mood and psychotic disorders. Genetic studies also indicate common liability among SZ, BD, and MDD and a partly distinct relationship with ASD. Such interpretation should remain probabilistic: a map-level cosine does not identify a particular cell type, developmental window, or causal pathway. Nevertheless, the present design test removes one simple alternative account by showing that the pairwise recruitment ordering is different.

3.4. Cortical surface-area concordance

Surface-area similarities followed another diagnostic grouping. SZ–MDD, SZ–ASD, and MDD–ASD had high cosine values of 0.945, 0.867, and 0.811. Pairs involving BD were low: 0.047 with SZ, 0.119 with MDD, and −0.098 with ASD. The ordered display in Fig. 6 uses point color for recruitment affinity and point size for the number

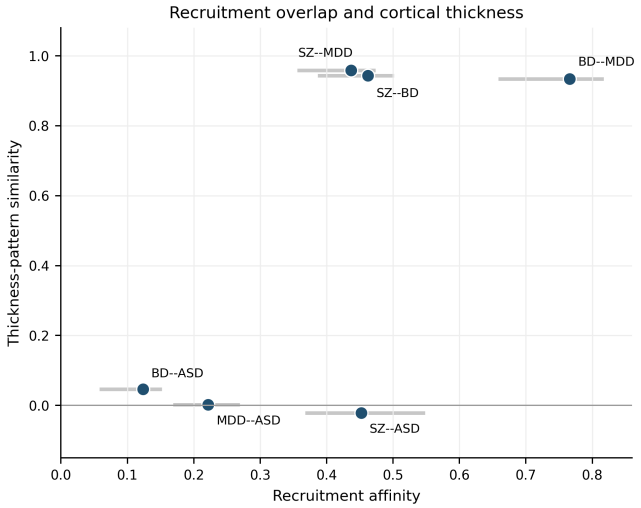

Figure 5. Thickness concordance.

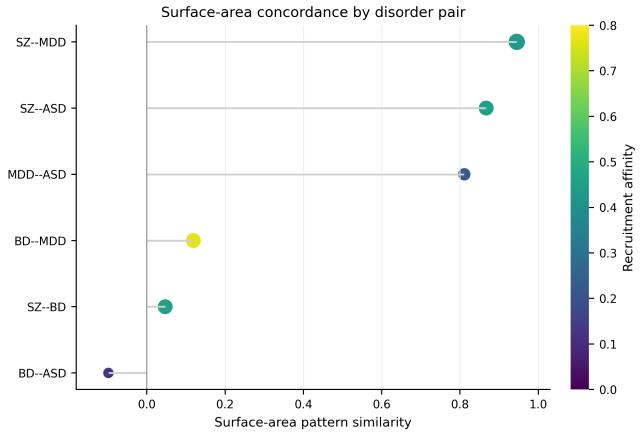

Figure 6. Surface-area concordance.

of co-recruiting protocols. High surface-area similarities occur across moderate and low affinities, whereas the highest affinity belongs to BD–MDD, whose surface-area similarity is only 0.119.

The rank association was $\rho = 0.086$ with an exact probability of 0.958. This near-zero coefficient is not a consequence of one isolated pair. The three high surface-area similarities include MDD–ASD, whose affinity is only 0.222, and SZ–ASD, whose affinity is 0.452. Meanwhile, BD–MDD combines the greatest affinity with a low cortical similarity. The lack of alignment is therefore structural across the six-pair ordering.

Thickness and surface area should not be expected to share identical transdiagnostic organization. Their genetic determinants are partly independent, and developmental work associates surface expansion with progenitor dynamics while thickness reflects later cellular and intracortical processes. The contrasting diagnostic clusters may therefore represent different biological axes. The acquisition audit strengthens that interpretation only in a limited sense: it shows that the clusters cannot be read directly from protocol co-sampling. It does not establish which molecular or developmental mechanism produced them.

Prior cross-disorder imaging has shown that morphological commonality can differ between cortical, subcortical, and white-matter measures. COCORO studies illustrate how coordinated samples can replicate and broaden international findings.[27] Work on subcortical asymmetry,[28] cognitive subgroups,[29] memory-related subcortical structure,[30] and cognitive or social associations[31] further demonstrates that a diagnosis-level mean effect is only one layer of neurobiological variation. Symptom-related pallidal differences[32] and frontal white-matter associations with social function[33] likewise caution against assigning a single anatomical meaning to a cross-disorder cosine.

Functional relevance also varies across tissue measures. White-matter microstructure has been related to work hours,[34] perceptual organization and working memory,[35] and longitudinally relevant cognitive decline.[36] A structural vulnerability index can separate clinical from subclinical depression and relate anatomy to cognition.[37] These results suggest a productive next step: pair the acquisition-allocation audit with individual-level transdiagnostic dimensions and outcomes. Such analyses would determine whether a cortical similarity that is robust to protocol allocation is also clinically coherent.

3.5. Deletion and full-matrix sensitivity

Pair-specific deletion intervals were narrow enough to preserve the broad affinity ordering, although protocols with large mixed-diagnosis samples exerted visible influence. SZ–BD ranged from 0.386 to 0.502, SZ–MDD from 0.356 to 0.474, and SZ–ASD from 0.367 to 0.548. BD–MDD remained the largest affinity under every deletion, ranging from 0.658 to 0.817. BD–ASD remained the smallest or near-smallest, ranging from 0.058 to 0.152. No single protocol converted the recruitment geometry into the cortical geometry.

At the association level, the thickness coefficient ranged from 0.029 to 0.600 and the surface-area coefficient from 0.086 to 0.371. The individual deletion values in Fig. 7 are discrete because Spearman correlation over six observations has a limited set of attainable ranks. The open diamonds mark the full-matrix coefficients. Although omission of some protocols increased the thickness correlation, none of the deletion values supplied consistent evidence of a strong relationship across both phenotypes.

Restoring Hiroshima2 and Kyushu3 changed the six affinities to 0.455, 0.421, 0.452, 0.728, 0.122, and 0.214 in the fixed pair order. Both rank correlations remained exactly unchanged at 0.257 for thickness and 0.086 for surface area. This invariance occurs because the added BD-only and MDD-only rows altered magnitudes modestly without changing the rank ordering of the six affinities. The primary inference therefore does not depend on whether participants from protocols lacking HC are included in the design audit.

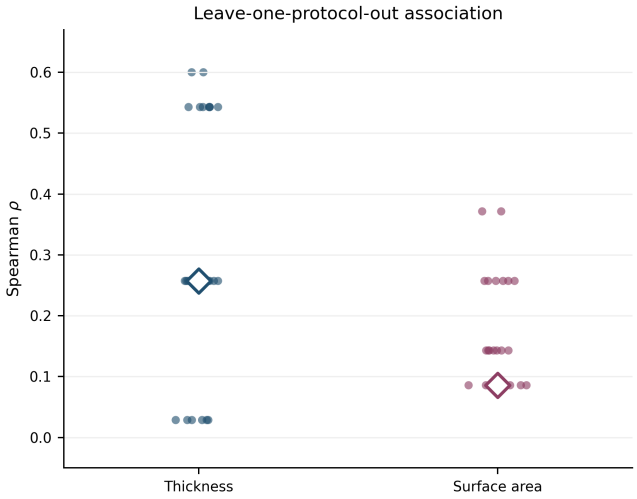

Figure 7. Deletion robustness.

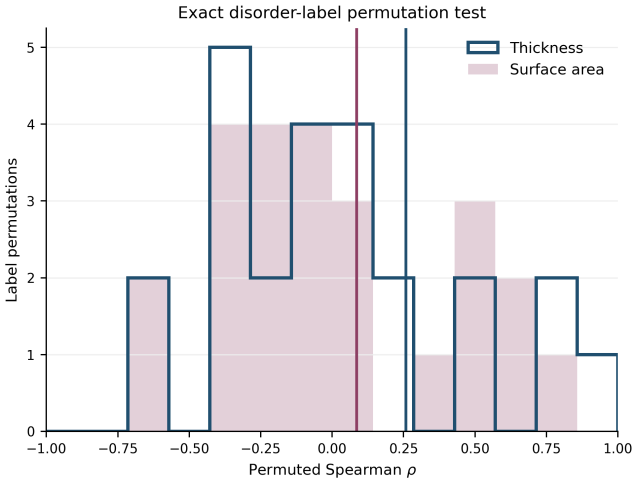

Figure 8. Exact label permutations.

The exact permutation distributions in Fig. 8 place the observed coefficients within the body of their attainable values. Neither observed line lies in an extreme tail. With only four disorder labels, the test has coarse resolution and should not be interpreted as proving absence of any acquisition effect. Its value is logical rather than asymptotic: the observed assignment is not unusually aligned compared with the complete set of diagnostic relabelings.

The deletion and permutation calculations answer different questions. Deletion asks whether a particular protocol drives the observed coefficient; label permutation asks whether the diagnostic assignment is exceptional among all assignments. Their agreement is important. The first shows that the conclusion is not attributable to one protocol row, and the second shows that the modest correlations are not rare under relabeling.

3.6. Interpretive implications

The principal finding is a mismatch between recruitment geometry and cortical geometry. For thickness, SZ, BD, and MDD form a highly concordant cortical group even though their pairwise recruitment affinities range from moderate to high. ASD has near-zero thickness concordance with all three, despite an SZ–ASD recruitment affinity comparable to SZ–BD and SZ–MDD. For surface area, SZ, MDD, and ASD form the concordant group, including the low-affinity MDD–ASD pair, while BD–MDD has the highest recruitment affinity and low cortical concordance. A single acquisition-allocation mechanism cannot parsimoniously explain both patterns.

The pairwise contrasts provide a stronger argument than either correlation alone. SZ–BD and SZ–ASD are close in recruitment affinity, differing by only 0.010, yet their thickness similarities differ by 0.964. BD–MDD has an affinity 0.329 higher than SZ–MDD, but its surface-area similarity is 0.826 lower. MDD–ASD has an affinity of only 0.222 and nevertheless a surface-area similarity of 0.811. These are not minor departures from a smooth trend; they exchange the pairwise ordering at clinically meaningful parts of the comparison.

The phenotype contrast is equally revealing. Recruitment affinity is identical for a disorder pair regardless of which cortical measure is analyzed, because the participants occupy the same protocol matrix. Cortical similarity, however, changes greatly between thickness and area for SZ–BD, SZ–ASD, BD–MDD, and MDD–ASD. An allocation explanation that does not distinguish cortical phenotypes would predict a more consistent pairwise ranking. The observed phenotype dependence instead points toward tissue-specific or measurement-specific processes.

Magnitude remains separate from direction. High cosine similarity does not mean that two disorders have equally severe alteration, and the broadest disorder effect can coexist with a common spatial pattern. The source quantities show that SZ generally has larger negative cortical effects than the other diagnoses, while the cosine analysis describes whether regional deviations point in similar directions. The design audit is therefore most appropriately interpreted as a test of spatial ordering, not a ranking of disease burden.

This mismatch should increase confidence in the descriptive transdiagnostic result, but it is not evidence that scanner effects are absent. Protocol counts reveal where participants were measured, not how each scanner altered regional values. Two protocols with equal participant proportions could differ substantially in field strength, coil, sequence, software, or quality control. Conversely, distinct protocols could yield comparable regional estimates after within-protocol contrast estimation. The analysis therefore evaluates sufficiency: whether the published cortical ordering can be explained by protocol co-sampling alone. The answer is no.

Clinical heterogeneity remains another source of uncertainty. Medication exposure, duration of illness, symptom severity, smoking, alcohol use, and other substance exposures can affect brain structure. Alcohol use has been associated with thinner cortex and ventricular differences across severe mental illness and controls.[38]

Smoking is associated with cortical thinning in population samples[39] and with cingulate and insular differences in severe mental illness.[40] Cocaine exposure has also been linked to cortical-thickness abnormalities.[41] If these variables differ across diagnosis–protocol cells, they could shape cortical maps without producing a simple relationship with recruitment affinity.

The audit also distinguishes broad support from equal support. SZ has the highest entropy and effective protocol number, yet its positive-support Gini remains 0.421. MDD occupies more protocols than BD or ASD but has the greatest inequality among its contributing rows. Researchers should therefore report at least three allocation descriptors: support size, concentration within support, and pairwise distribution affinity. Any one quantity alone misses part of the design.

Reporting these descriptors changes how readers can evaluate transportability. A broad support with moderate concentration suggests that an effect has been observed across varied acquisition conditions, even if several large protocols dominate precision. A narrow support with low within-support inequality indicates internally balanced recruitment over only a few conditions; it should not be described as broadly distributed. Pairwise affinity then clarifies whether two diagnoses draw evidence from similar conditions. Together, the values offer a compact design portrait without disclosing participant-level records.

The measures can also guide the order of sensitivity work. Pairs with low affinity deserve an intersecting-protocol analysis first because their maps are supported by different acquisition mixtures. Pairs with high affinity but concentrated recruitment deserve deletion analysis because one mixed-diagnosis protocol may dominate both distributions. Diagnoses with a low effective protocol number require cautious generalization even when their total participant count is adequate. These priorities follow directly from the allocation matrix and do not depend on seeing a favorable or unfavorable cortical result.

For future multisite analyses, protocol-distribution concordance can be calculated before image modeling. A low-affinity disorder pair can then be flagged for matched-protocol analysis, inverse allocation weighting, hierarchical partial pooling, or transportability assessment. A high-affinity pair does not require correction merely because it is high; instead, it indicates that acquisition support is similar and that direct map comparison has a stronger design basis. These uses are diagnostic rather than corrective. They help determine which sensitivity analyses are necessary.

Individual-level extension would require protocol-specific regional effect vectors. One useful model would estimate each disorder map under partial pooling, then compare pairwise similarity after projecting out protocol-associated components. Another would restrict every disorder pair to the intersection of eligible protocols and compare that result with the full pooled map. A third would estimate leave-one-site-out rather than leave-one-protocol-out stability where multiple scanner protocols occur at the same institution. These analyses cannot be recovered from counts and paper-level cosines, but the present results identify precisely why they are worth conducting.

4. CONCLUSION

The research question was whether protocol-level recruitment overlap is sufficient to explain transdiagnostic cortical map similarity across SZ, BD, MDD, and ASD. The evidence answers that question in the negative. Pairwise recruitment affinity varied widely, yet its ordering had only a weak association with thickness concordance and an almost null association with surface-area concordance. The highest-affinity pair did not show high surface-area similarity, and several pairs with modest or low affinity showed strong cortical concordance. Exact disorder-label permutations placed both observed correlations well within ordinary attainable values, deletion of individual protocols did not overturn the inference, and restoration of all 5,549 participants left the correlations unchanged.

The cortical grouping therefore cannot be read as a simple imprint of diagnostic co-sampling across scanner protocols. This conclusion supports continued biological investigation of the distinct thickness and surface-area patterns, while also showing that recruitment allocation should be quantified in multisite cross-disorder research. The included method provides a reproducible audit from information that is often already present in cohort tables. Its strongest use is to identify disorder pairs that require matched-protocol or individual-level sensitivity analyses before causal interpretation.

FUNDING

No external funding was received for this work.

COMPETING INTERESTS

The author declares no competing interests.

BIBLIOGRAFIA / REFERENCES

1. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, et al. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry*. 2020;10:100.
2. Onitsuka T, Hirano Y, Nemoto K, Hashimoto N, Kushima I, Koshiyama D, et al. Trends in big data analyses by multicenter collaborative translational research in psychiatry. *Psychiatry Clin Neurosci*. 2022;76:1–14.
3. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603:654–660.

4. Opel N, Goltermann J, Hermesdorf M, Berger K, Baune BT, Dannlowski U. Cross-disorder analysis of brain structural abnormalities in six major psychiatric disorders: a secondary analysis of mega- and meta-analytical findings from the ENIGMA consortium. *Biol Psychiatry*. 2020;88:678–686.
5. Cheon EJ, Bearden CE, Sun D, Ching CRK, Andreassen OA, Schmaal L, et al. Cross-disorder comparisons of brain structure in schizophrenia, bipolar disorder, major depressive disorder, and 22q11.2 deletion syndrome: a review of ENIGMA findings. *Psychiatry Clin Neurosci*. 2022;76:140–161.
6. Koshiyama D, Fukunaga M, Okada N, Morita K, Nemoto K, Usui K, et al. White matter microstructural alterations across four major psychiatric disorders: mega-analysis study in 2937 individuals. *Mol Psychiatry*. 2020;25:883–895.
7. van Erp TGM, Walton E, Hibar DP, Schmaal L, Jiang W, Glahn DC, et al. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis Consortium. *Biol Psychiatry*. 2018;84:644–654.
8. Hibar DP, Westlye LT, Doan NT, Jahanshad N, Cheung JW, Ching CRK, et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry*. 2018;23:932–942.
9. Schmaal L, Hibar DP, S"amann PG, Hall GB, Baune BT, Jahanshad N, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry*. 2017;22:900–909.
10. van Rooij D, Anagnostou E, Arango C, Auzias G, Behrmann M, Busatto GF, et al. Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan. *Am J Psychiatry*. 2018;175:359–369.
11. Writing Committee for the Attention-Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, Bipolar Disorder, Major Depressive Disorder, Obsessive-Compulsive Disorder, and Schizophrenia ENIGMA Working Groups, et al. Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. *JAMA Psychiatry*. 2021;78:47–63.
12. Radonjić NV, Hess JL, Rovira P, Andreassen O, Buitelaar JK, Ching CRK, et al. Structural brain imaging studies offer clues about the effects of the shared genetic etiology among neuropsychiatric disorders. *Mol Psychiatry*. 2021;26:2101–2110.
13. Park BY, Kebets V, Larivière S, Hettwer MD, Paquola C, van Rooij D, et al. Multiscale neural gradients reflect transdiagnostic effects of major psychiatric conditions on cortical morphology. *Commun Biol*. 2022;5:1024.
14. Hansen JY, Shafiei G, Vogel JW, Smart K, Bearden CE, Hoogman M, et al. Local molecular and global connectomic contributions to cross-disorder cortical abnormalities. *Nat Commun*. 2022;13:4682.
15. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP, et al. The genetic architecture of the human cerebral cortex. *Science*. 2020;367:eaay6690.
16. Silbereis JC, Pochareddy S, Zhu Y, Li M, Sestan N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron*. 2016;89:248–268.
17. Boedhoe PSW, van Rooij D, Hoogman M, Twisk JWR, Schmaal L, Abe Y, et al. Subcortical brain volume, regional cortical thickness, and cortical surface area across disorders: findings from the ENIGMA ADHD, ASD, and OCD working groups. *Am J Psychiatry*. 2020;177:834–843.
18. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*. 2018;167:104–120.
19. Pomponio R, Erus G, Habes M, Doshi J, Srinivasan D, Mamourian E, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*. 2020;208:116450.
20. Nemoto K, Shimokawa T, Fukunaga M, Yamashita F, Tamura M, Yamamori H, et al. Differentiation of schizophrenia using structural MRI with consideration of scanner differences: a real-world multisite study. *Psychiatry Clin Neurosci*. 2020;74:56–63.
21. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006;31:968–980.
22. Fischl B. FreeSurfer. *NeuroImage*. 2012;62:774–781.
23. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc*. 2007;82:591–605.
24. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. *NeuroImage*. 2014;92:381–397.
25. Bedford SA, Park MTM, Devenyi GA, Tullo S, Germann J, Patel R, et al. Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. *Mol Psychiatry*. 2020;25:614–628.
26. Matsumoto J, Fukunaga M, Miura K, Nemoto K, Okada N, Hashimoto N, et al. Cerebral cortical structural alteration patterns across four major psychiatric disorders in 5549 individuals. *Mol Psychiatry*. 2023;28:4915–4923. doi:10.1038/s41380-023-02224-7.
27. Koshiyama D, Miura K, Nemoto K, Okada N, Matsumoto J, Fukunaga M, et al. Neuroimaging studies within Cognitive Genetics Collaborative Research Organization aiming to replicate and extend works of ENIGMA. *Hum Brain Mapp*. 2022;43:182–193.
28. Okada N, Fukunaga M, Yamashita F, Koshiyama D,

Yamamori H, Ohi K, et al. Abnormal asymmetries in subcortical brain volume in schizophrenia. *Mol Psychiatry*. 2016;21:1460–1466.
29. Yasuda Y, Okada N, Nemoto K, Fukunaga M, Yamamori H, Ohi K, et al. Brain morphological and functional features in cognitive subgroups of schizophrenia. *Psychiatry Clin Neurosci*. 2020;74:191–203.
30. Koshiyama D, Fukunaga M, Okada N, Yamashita F, Yamamori H, Yasuda Y, et al. Subcortical association with memory performance in schizophrenia: a structural magnetic resonance imaging study. *Transl Psychiatry*. 2018;8:20.
31. Koshiyama D, Fukunaga M, Okada N, Yamashita F, Yamamori H, Yasuda Y, et al. Role of subcortical structures on cognitive and social function in schizophrenia. *Sci Rep*. 2018;8:1183.
32. Ito S, Miura K, Miyayama M, Matsumoto J, Fukunaga M, Ishimaru K, et al. Association between globus pallidus volume and positive symptoms in schizophrenia. *Psychiatry Clin Neurosci*. 2022;76:602–603.
33. Koshiyama D, Fukunaga M, Okada N, Morita K, Nemoto K, Yamashita F, et al. Role of frontal white matter and corpus callosum on social function in schizophrenia. *Schizophr Res*. 2018;202:180–187.
34. Matsumoto J, Fukunaga M, Miura K, Nemoto K, Koshiyama D, Okada N, et al. Relationship between white matter microstructure and work hours. *Neurosci Lett*. 2020;740:135428.
35. Koshiyama D, Fukunaga M, Okada N, Morita K, Nemoto K, Yamashita F, et al. Association between the superior longitudinal fasciculus and perceptual organization and working memory: a diffusion tensor imaging study. *Neurosci Lett*. 2020;738:135349.
36. Matsumoto J, Miura K, Fukunaga M, Nemoto K, Koshiyama D, Okada N, et al. Association study between white matter microstructure and intelligence decline in schizophrenia. *Clin EEG Neurosci*. 2021. doi:10.1177/15500594211063314.
37. Kochunov P, Ma Y, Hatch KS, Schmaal L, Jahanshad N, Thompson PM, et al. Separating clinical and subclinical depression by big data structural vulnerability index and its impact on cognition: ENIGMA dot product. *Pac Symp Biocomput*. 2022;27:133–143.
38. Lange EH, Nerland S, Jørgensen KN, Mørch-Johnsen L, Nesvåg R, Hartberg CB, et al. Alcohol use is associated with thinner cerebral cortex and larger ventricles in schizophrenia, bipolar disorder and healthy controls. *Psychol Med*. 2017;47:655–668.
39. Karama S, Ducharme S, Corley J, Chouinard-Decorte F, Starr JM, Wardlaw JM, et al. Cigarette smoking and thinning of the brain's cortex. *Mol Psychiatry*. 2015;20:778–785.
40. Jørgensen KN, Skjærvø I, Mørch-Johnsen L, Haukvik UK, Lange EH, Melle I, et al. Cigarette smoking is associated with thinner cingulate and insular cortices in patients with severe mental illness. *J Psychiatry Neurosci*. 2015;40:241–249.
41. Makris N, Gasic GP, Kennedy DN, Hodge SM, Kaiser JR, Lee MJ, et al. Cortical thickness abnormalities in cocaine addiction: a reflection of both drug use and a pre-existing disposition to drug abuse? *Neuron*. 2008;60:174–188.