JAMA Psychiatry | **Original Investigation**

Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models

Thomas Wolfers, MSc; Nhat Trung Doan, PhD; Tobias Kaufmann, PhD; Dag Alnæs, PhD; Torgeir Moberget, PhD; Ingrid Agartz, MD, PhD; Jan K. Buitelaar, MD, PhD; Torill Ueland, PhD; Ingrid Melle, MD, PhD; Barbara Franke, PhD; Ole A. Andreassen, MD, PhD; Christian F. Beckmann, PhD; Lars T. Westlye, PhD; Andre F. Marquand, PhD

IMPORTANCE Schizophrenia and bipolar disorder are severe and complex brain disorders characterized by substantial clinical and biological heterogeneity. However, case-control studies often ignore such heterogeneity through their focus on the average patient, which may be the core reason for a lack of robust biomarkers indicative of an individual's treatment response and outcome.

OBJECTIVES To investigate the degree to which case-control analyses disguise interindividual differences in brain structure among patients with schizophrenia and bipolar disorder and to map the brain alterations linked to these disorders at the level of individual patients.

DESIGN, SETTING, AND PARTICIPANTS This study used cross-sectional, T1-weighted magnetic resonance imaging data from participants recruited for the Thematically Organized Psychosis study from October 27, 2004, to October 17, 2012. Data were reanalyzed in 2017 and 2018. Patients were recruited from inpatient and outpatient clinics in the Oslo area of Norway, and healthy individuals from the same catchment area were drawn from the national population registry.

MAIN OUTCOMES AND MEASURES Interindividual differences in brain structure among patients with schizophrenia and bipolar disorder. Voxel-based morphometry maps were computed, which were used for normative modeling to map the range of interindividual differences in brain structure.

RESULTS This study included 218 patients with schizophrenia spectrum disorders (mean [SD] age, 30 [9.3] years; 126 [57.8%] male), of whom 163 had schizophrenia (mean [SD] age, 31 [8.7] years; 105 [64.4%] male) and 190 had bipolar disorder (mean [SD] age, 34 [11.3] years; 79 [41.6%] male), and 256 healthy individuals (mean [SD] age, 34 [9.5] years; 140 [54.7%] male). At the level of the individual, deviations from the normative model were frequent in both disorders but highly heterogeneous. Overlap of more than 2% among patients was observed in only a few loci, primarily in frontal, temporal, and cerebellar regions. The proportion of alterations was associated with diagnosis and cognitive and clinical characteristics within clinical groups. Patients with schizophrenia, on average, had significantly reduced gray matter in frontal regions, cerebellum, and temporal cortex. In patients with bipolar disorder, mean deviations were primarily present in cerebellar regions.

CONCLUSIONS AND RELEVANCE This study found that group-level differences disguised biological heterogeneity and interindividual differences among patients with the same diagnosis. This finding suggests that the idea of the average patient is a noninformative construct in psychiatry that falls apart when mapping abnormalities at the level of the individual patient. This study presents a workable route toward precision medicine in psychiatry.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Thomas Wolfers, MSc, Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, PO Box 9101, 6500 HB Nijmegen, the Netherlands (t.wolfers@donders.ru.nl).

Biological markers that objectively indicate someone's medical status have been identified for many diseases; for example, in oncology, these biomarkers revolutionized cancer diagnosis and treatment.[1] In addition, in psychiatry, such prospects inspired a quest for the identification of biomarkers for health and disorder,[2-4] using, for example, candidate gene approaches.[5] However, the complex etiologic and biological mechanisms of mental disorders and a fundamental reliance on symptom-based diagnoses have hindered progress. Psychiatry is now the last area of medicine in which diseases are diagnosed solely on the basis of symptoms, and biomarkers to assist treatment remain to be developed. To bring precision medicine to psychiatry, large-scale international initiatives work toward stratifying mental disorders into biologically more homogeneous subtypes based on the integration of many levels of information across multiple dimensions of functioning.[4,6]

The most significant obstacle toward finding accurate and reliable biomarkers in mental disorders is their extreme heterogeneity[7] based on current psychiatric nosology. Heterogeneity can be observed on at least 3 levels: (1) heterogeneity as a consequence of different symptom profiles that are classified under the same disorder (clinical heterogeneity), (2) heterogeneity induced by different biological predispositions converging on the same symptoms (biological heterogeneity), and (3) different environmental events that cause (or prevent) the same symptoms (environmental heterogeneity). Case-control designs, which assume that patient and control groups are distinct entities, are overwhelmingly dominant in psychiatry but are limited to detecting group differences that essentially describe an average patient. They neglect interindividual differences, which are crucial for mapping the heterogeneous disease phenotype at the level of the individual.[8] Schizophrenia and bipolar disorder are excellent examples of highly heterogeneous mental disorders.[9,10] They have been linked to multiple brain systems and neural processes, which become perturbed throughout development through complex interactions between the individual's genetic architecture and environmental stressors.[9,11] Both disorders have been linked to transdiagnostic impairments in the dopamine system.[12] However, because these conclusions are based on group-level comparisons, they provide limited information about disease mechanisms in individual patients.

This study aimed to quantify the brain structural heterogeneity in adults with schizophrenia and bipolar disorder by mapping regional brain alterations at the level of individual participants. We hypothesized that group-level differences only represent a small part of the neurobiological abnormalities that characterize these disorders and that highly individual deviations from the norm comprise the bulk of these abnormalities. To test this hypothesis, we used a normative modeling approach that maps interindividual differences in reference to the healthy range. A normative model can be understood as a statistical model that maps demographic or behavioral variables to a quantitative brain readout.[13] Similar to growth charts used in somatic medicine, in which a child's height is compared with the normative distribution for height at a particular age, a normative model can be used to characterize individuals in reference to a normative brain structure at a particular age.[13,14] Using this approach, we provide a route toward precision medicine in psychiatry in that we provide quantitative estimates of the heterogeneity among patients with schizophrenia and bipolar disorder by investigating the degree of spatial overlap in deviations from the normative model and chart the heterogeneity in alterations of brain structure across these disorders and at the level of the individual patient.

Key Points

Question Is the focus on the average patient disguising interindividual differences among patients with mental disorders?

Findings In this study of magnetic resonance imaging data from 218 patients with schizophrenia spectrum disorders and 256 healthy control individuals, mapping of interindividual differences in brain structure revealed that only a few brain loci had the same abnormalities in more than 2% of patients with the same disorder despite robust group-level differences in multiple brain regions between patients and control individuals.

Meaning These findings suggest that the idea of the average patient is a noninformative construct that falls apart when mapping interindividual differences and provide a framework toward precision medicine in psychiatry.

Methods

Participants

All participants were recruited as part of the Thematically Organized Psychosis study from October 27, 2004, to October 17, 2012. Data were reanalyzed in 2017 and 2018. Patients were recruited from inpatient and outpatient clinics in the Oslo area of Norway. Patients (aged 18-65 years) understood and spoke a Scandinavian language, had no history of severe head trauma, and had an IQ above 70. Patients were assessed by trained physicians or clinical psychologists.[15] Psychiatric diagnosis was established using the *Structured Clinical Interview for DSM-IV Axis I Disorders*.[16] Healthy individuals were randomly sampled from national registries if neither they nor their relatives had a psychiatric or alcohol or substance use disorder or any cannabis use during the past 3 months. All participants completed a neuropsychological test battery, including verbal learning and memory, processing speed, working memory, and executive functioning. Written informed consent was obtained from all participants. Magnetic resonance imaging as well as cognitive and clinical data were deidentified. The study was approved by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate. More details and a description of magnetic resonance imaging acquisition and processing are given in the eMethods in the Supplement.

Normative Modeling

We estimated a normative brain aging model by using gaussian process regression to predict regional gray and white matter volumes across the brain from age and sex using voxel-based morphometry–derived gray and white matter maps.

To supplement this analysis and to characterize volumetric differences more precisely, we also estimated normative models using Freesurfer-derived cortical thickness and pial area measures (eMethods in the Supplement). To avoid overfitting in the normative models, it is crucial to estimate performance out of sample. Therefore, we estimated the normative range for this model in healthy individuals under 10-fold cross-validation, then applied the model trained on all data to patients. Normative models were estimated using gaussian process regression,[17] yielding coherent measures of predictive confidence in addition to point estimates. Such measurement was important because we used this uncertainty measure to quantify the deviation of each patient from the group mean at each brain locus. Thus, we were able to statistically quantify deviations from the normative model with regional specificity by computing a z score for each voxel that reflected the difference between the predicted volume and the true volume normalized by the uncertainty of the prediction.[13]

We estimated mean deviations from the normative model in healthy individuals, patients with bipolar disorder, and patients with schizophrenia using Permutation Analysis of Linear Models (PALM) on the normative deviation maps,[18] which allows for permutation-based inference using t tests. PALM was used to create a map of z values for each of these groups. We established the threshold for those maps using $z = \pm 2.6$ (ie, $P < .005$) to make them comparable with individual maps of deviation, explained below. Furthermore, we determined multiple comparison–corrected, threshold-free, cluster-enhanced and modality-corrected differences between the groups. The threshold for individual deviation maps was established at $|z| > 2.6$ (ie, $P < .005$), and extreme positive and extreme negative deviations from the normative model were defined based on this threshold. We chose to use a single fixed threshold for statistical significance for each participant individually because it simplifies the comparison across individuals relative to the alternative approach of controlling the false discovery rate (FDR) separately for each participant.[13] Specifically, although FDR controls for multiple comparisons, this process is at the expense of estimating a separate threshold for each participant. Therefore, it is insensitive to an overall shift in deviations from the normative model for 1 individual. In other words, an individual with small reductions in gray matter across the entire cortex may seem to have normal findings using an FDR thresholding procedure because the overall distribution of deviation is shifted for this individual. However, we also performed the analyses again controlling the FDR at the individual participant level using the Benjamini and Hochberg procedure.[19] These analyses are reported in the eMethods in the Supplement and led to identical conclusions. All extreme deviations were combined into scores that represented the percentage of extreme positive and negative deviating voxels for each participant, expressed relative to the total number of intracerebral voxels. We tested for associations between diagnosis and those scores using a χ^2 test, corrected for multiple comparisons using the Bonferroni-Holm method.[20] To assess the extent of those extreme deviations spatially, we created individualized maps and calculated the voxelwise overlap among individuals from the same

groups. The overlap for the clinically less homogenous group of patients with schizophrenia spectrum disorder are given in the eResults in the Supplement. Finally, we tested for associations between the percentage of extremely deviating voxels and age, disease duration, and cognitive performance. We corrected for the number of regressions (n = 12) within each disorder and modality using the Bonferroni-Holm method.[20] A corrected threshold $P < .05$ was considered to be statistically significant. All analyses were performed with Python software, version 3.6 (Python Software Foundation). We performed additional checks to eliminate the possibility of nuisance effects of scanner artifacts (eg, head motion), medication, and substance abuse confounding our results (eResults in the Supplement).

Results

A total of 218 participants with schizophrenia spectrum disorders were included (mean [SD] age, 30 [9.3] years; 126 [57.8%] male), of whom 163 had pure schizophrenia (mean [SD] age, 31 [8.7] years; 105 [64.4%] male) and 190 had bipolar disorder (mean [SD] age, 34 [11.3] years; 79 [41.6%] male). The study also included 256 healthy individuals (mean [SD] age, 34 [9.5] years; 140 [54.7%] male) (eTable 1 in the Supplement).

Normative Model

In **Figure 1**, we depict a spatial representation of the voxelwise normative model, which was characterized by a global gray matter decrease from 20 to 70 years of age, particularly in the frontal and cerebellar regions, with the largest decrease primarily in the frontal areas for both women and men. In contrast, the normative model of white matter was characterized by decreases and increases across the adult lifetime for both women and men.
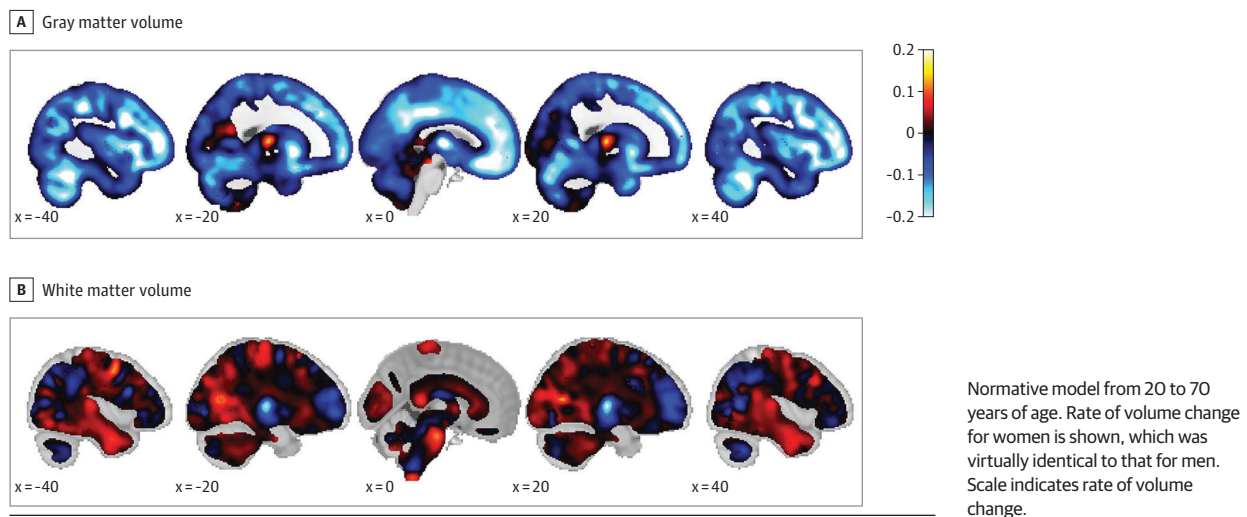
Mean Deviations Compared Across Patients and Healthy Individuals

Figure 2 shows mean group differences with respect to the normative model among healthy individuals (under cross-validation), patients with schizophrenia, and those with bipolar disorder in gray and white matter, corrected for modalities and multiple comparisons. In gray matter, patients with schizophrenia had greater mean negative deviations than did healthy individuals in frontal, temporal, and cerebellar regions; mean deviations were also more negative in patients with schizophrenia than in patients with bipolar disorder and were localized primarily in fontal brain regions. In white matter, we observed differences comparable to those described for gray matter.

Extreme Deviations Compared Across Patients and Healthy Individuals

In gray matter, patients with schizophrenia had a higher percentage of extreme negative deviations across voxels (0.9% of voxels) compared with healthy individuals (0.23% of voxels, $P < .001$; Wald $\chi^2 = 219.67$, $P < .001$, corrected $P < .001$) and patients with bipolar disorder (0.24% of voxels, $P < .001$). The percentage of extreme positive deviations across the groups

Figure 1. Characterization of the Normative Model

A | Gray matter volume

x = −40   x = −20   x = 0   x = 20   x = 40

B | White matter volume

x = −40   x = −20   x = 0   x = 20   x = 40

Normative model from 20 to 70 years of age. Rate of volume change for women is shown, which was virtually identical to that for men. Scale indicates rate of volume change.

indicated that healthy individuals differed from patients with schizophrenia and bipolar disorder (Wald χ^2 = 14.99, P = .001, corrected P = .004); this finding was associated with a larger percentage of extreme positive deviations in healthy individuals (1.08% of voxels) than in patients with bipolar disorder (0.79% of voxels, P = .001) and schizophrenia (0.78% of voxels, P = .001) (Figure 2A-C).

In white matter, patients with schizophrenia differed from healthy individuals and patients with bipolar disorder in terms of the percentage of extreme negative deviations (Wald χ^2 = 64.14, P < .001, corrected P<.004), with a larger proportion of extreme negative deviations in patients with schizophrenia (0.62% of voxels) than in healthy individuals (0.25% of voxels, P < .001) and in patients with bipolar disorder (0.41% of voxels, P = .001). In the percentage of extreme positive deviations across groups, healthy individuals differed from patients with schizophrenia and bipolar disorder (Wald χ^2 = 13.48, P = .001, corrected P<.004); a higher proportion of extreme positive deviations was seen in healthy individuals (1.14% of voxels) than in those with schizophrenia (0.83% of voxels, P = .003) and bipolar disorder (0.83% of voxels, P = .001) (Figure 2D-F).

Spatial Extent of Extreme Deviations Across Patients and Healthy Individuals

Figure 3 shows that, on average, healthy individuals did not typically deviate substantially from the normative model. Although we observed a scattered pattern of positive deviations on the overlap maps, no negative deviation was found in the mean or in the overlap maps.

Figure 4 shows that patients with schizophrenia had mean negative deviations from the normative model in frontal, superior parietal, and the cerebellum gray matter as well as positive deviations in the basal ganglia. The overlap maps for individuals with schizophrenia were dominated by extreme negative deviations in these regions. At least 2% of patients had extreme deviations in those regions. In addition, in white matter, patients with schizophrenia had widespread extreme negative deviations from the normative model, with focal

hotspots in frontal, temporal, and cerebellar regions. The pattern for all patients belonging to the schizophrenia spectrum (also including patients with schizoaffective and schizophreniform disorder) was the same as for the restricted set with schizophrenia (eFigure 1 in the Supplement).

Figure 5 shows that in patients with bipolar disorder, there were mean deviations in cerebellar, temporal, and thalamic regions. Mean effects for the deviations were not substantially affected by outliers (eFigure 6 in the Supplement). The deviations were predominantly negative, indicating that, typically, these patients had lower gray matter volume in those regions than estimated to be normative by the model. The overlap maps corresponded only marginally with this pattern; however, in the thalamic region, more than 2% of the patients had extreme negative values. We observed positive deviations in the caudate, which was supported by the overlap maps. In white matter, we observed negative deviations, in particular in the brainstem, temporal, and frontal regions. The distribution across individuals is shown using histograms of the percentage of deviating voxels in eFigure 2 in the Supplement and examples of individuals with extensive deviations in eFigure 3 and eFigure 4 in the Supplement.

Overlap maps using FDR (eFigure 5 in the Supplement) were slightly sparser but consistent with our main results. Individual extreme deviations within the different patient groups were linked to cognitive performance and disorder duration but not to age (eTable 2 in the Supplement). Furthermore, performing the analyses again using Freesurfer-derived cortical thickness and pial area measures produced nearly identical results and showed that extreme deviations were principally attributable to a thinning of the cortex in patients with schizophrenia and bipolar disorder (eFigures 7 and 8 in the Supplement).

Discussion

We mapped the biological heterogeneity of schizophrenia and bipolar disorder in reference to normative brain aging across the adult lifespan. We found that in patients with schizophrenia, the

Figure 2. Characterization of Mean Group-Level Deviations From the Normative Model
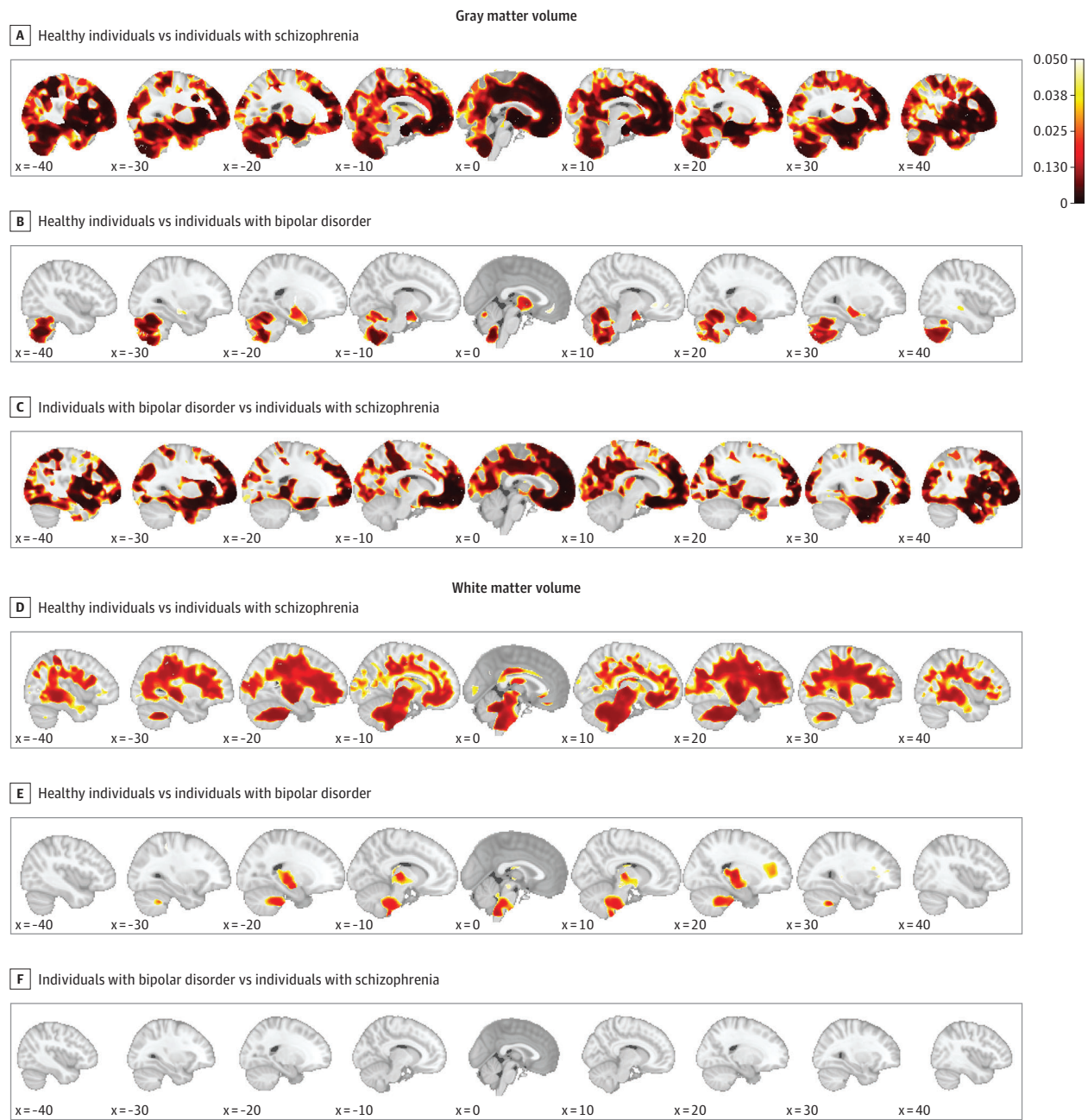
The mean differences were corrected for modalities and multiple comparisons. A-C, In gray matter, healthy individuals had stronger mean negative deviations than individuals with schizophrenia, especially in the frontal, temporal, and cerebellar regions; furthermore, individuals with bipolar disorder had stronger mean negative deviations than healthy in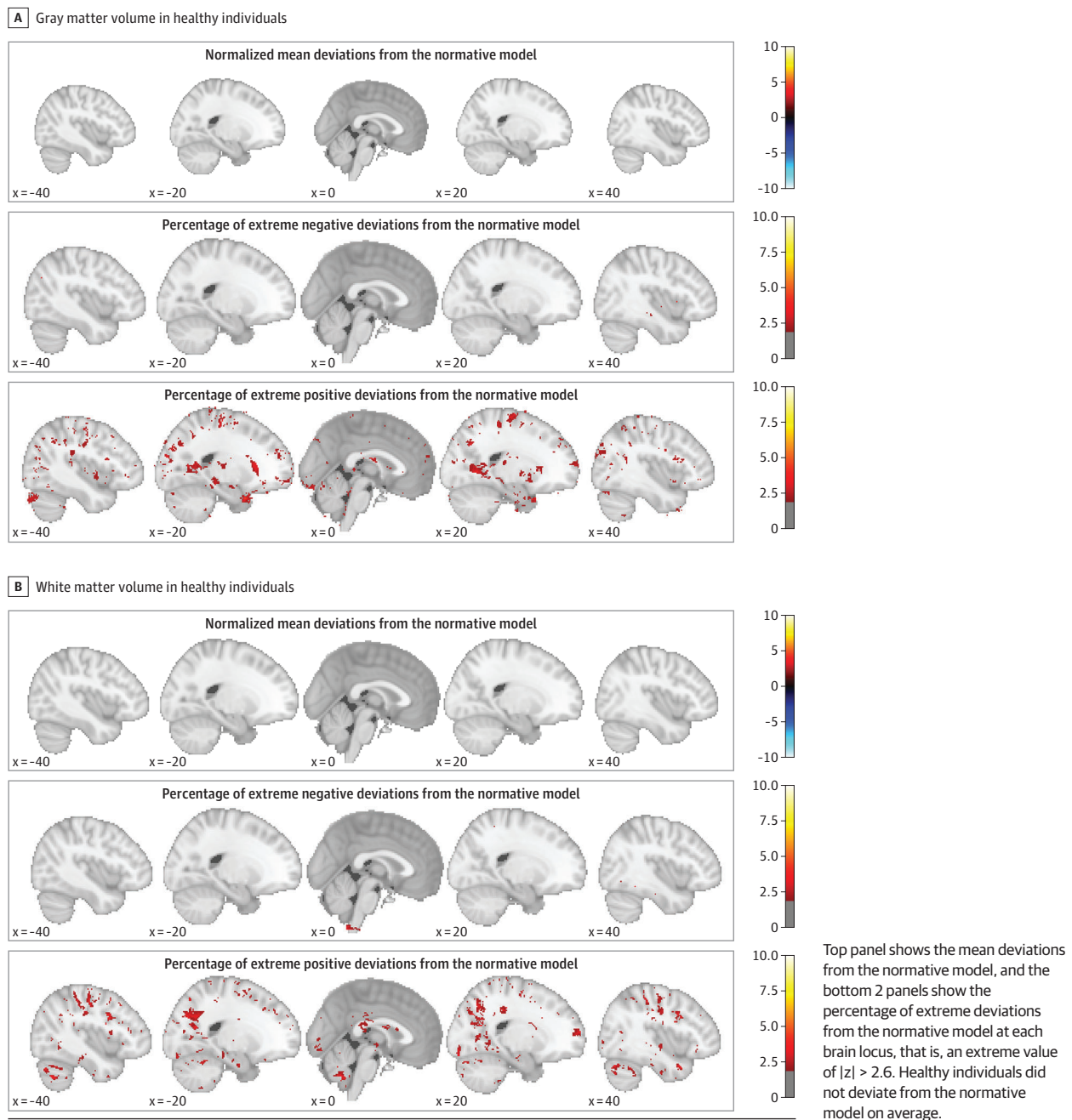dividuals in the cerebellum. Patients with bipolar disorder had weaker mean negative deviations than patients with schizophrenia in the frontal and temporal brain regions but not in the cerebellum. D-F, In white matter, the differences were comparable to those observed in gray matter. Healthy individuals had no regions with significant deviations in either gray or white matter. Scale indicates corrected P values.

frontal regions, the cerebellum, and the temporal cortex usually (ie, in the average patient) had reduced cortical volume compared with a healthy lifespan trajectory. For the average patient with bipolar disorder, this pattern was less pronounced and primarily present in cerebellar regions. This finding is in line with earlier, well-powered group comparison studies[21-25] that reported small to medium effects. Of more

importance, we found that these mean deviations masked extreme interindividual differences. Only a few brain loci had extreme deviations in more than 2% of the patients.

In this study, patients with schizophrenia and bipolar disorder differed extremely on an individual level; the lack of substantial overlap among patients in terms of extreme deviations from the normative model is evidence of the high degree of biologi-

Figure 3. Characterization of Extreme Deviations From the Normative Model in Healthy Control Individuals

A Gray matter volume in healthy individuals

B White matter volume in healthy individuals

Top panel shows the mean deviations from the normative model, and the bottom 2 panels show the percentage of extreme deviations from the normative model at each brain locus, that is, an extreme value of |z| > 2.6. Healthy individuals did not deviate from the normative model on average.
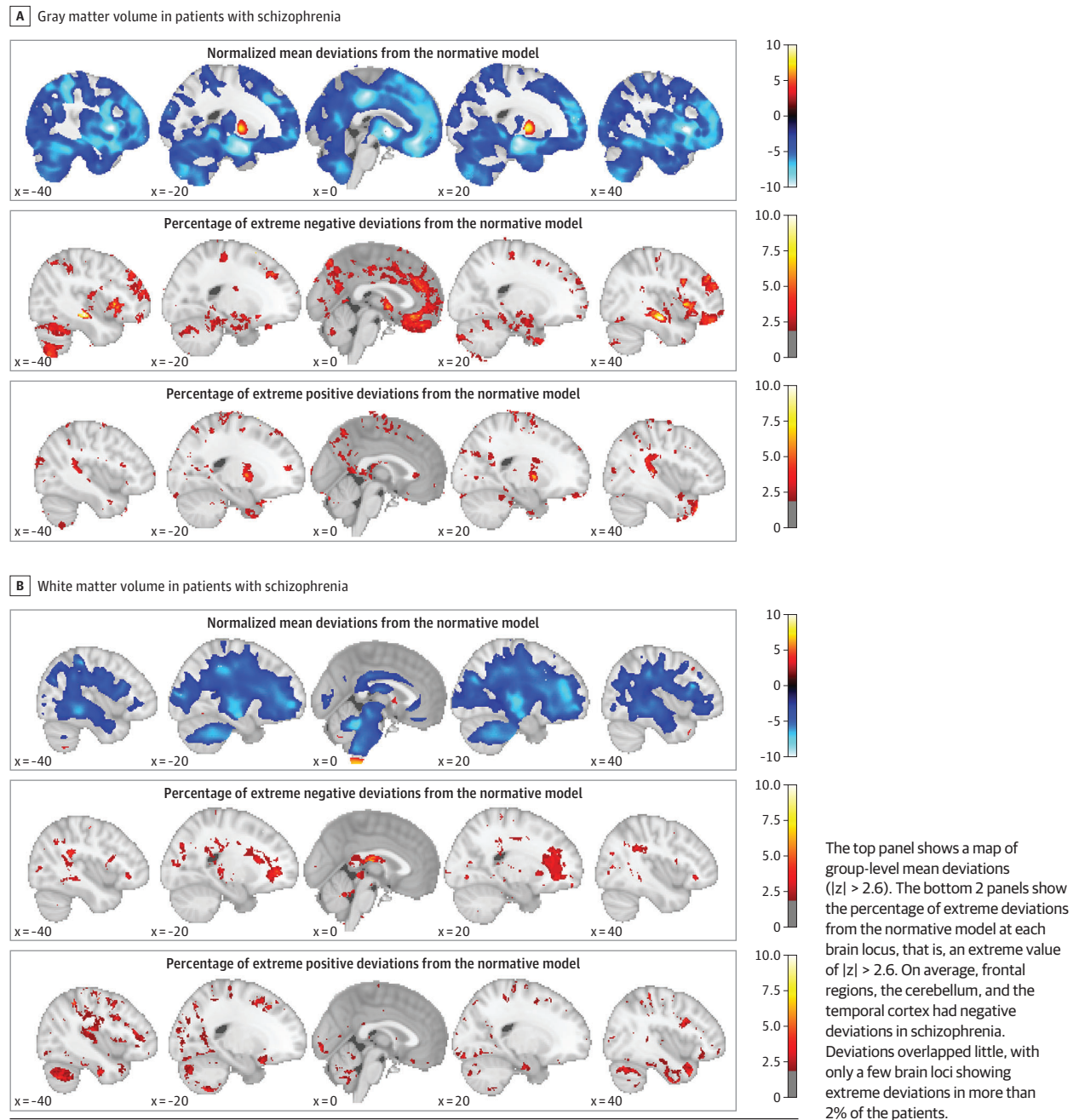
cal heterogeneity of both disorders. This finding is in line with the notion that mental disorders are complex, with little sharing of causal brain structural defect, genetic variants, or environmental stressors. Schizophrenia was conceptualized as a polygenic disorder half a century ago,[26] consistent with published identification of genetic risk factors.[27] Together with our current results on neuroimaging-based evidence of heterogeneity, these findings corroborate that the categorization of mental health disorders, as defined using current diagnostic manuals, does not conform with biology[3,4,28]; such work also emphasizes the need to develop tools for clinical stratification and characterization

spanning conventional diagnostic boundaries, such as in the Research Domain Criteria.[4]

A previous study[29] that used a classic case-control design did not identify biological signatures for schizophrenia or bipolar disorder informative enough for individualized estimations. In practice, the discriminative capability of candidate biological signatures is most commonly studied using multivariate pattern classifiers that integrate a large number of features in a single model. However, in patients with schizophrenia and bipolar disorder, these discriminative patterns do not identify the diagnostic categories with an accuracy that

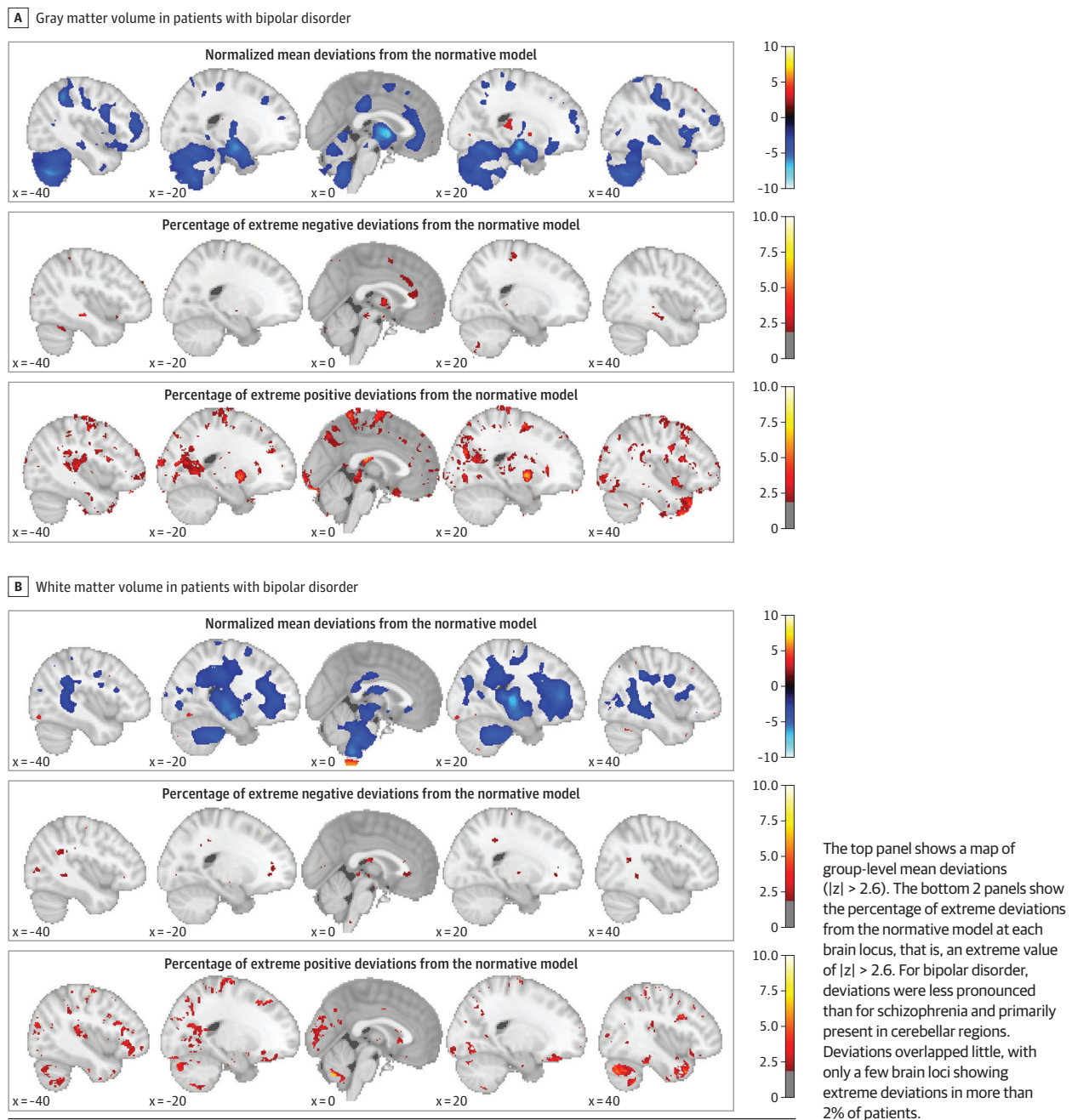Figure 4. Characterization of Extreme Deviations From the Normative Model in Patients With Schizophrenia

A Gray matter volume in patients with schizophrenia

Normalized mean deviations from the normative model

Percentage of extreme negative deviations from the normative model

Percentage of extreme positive deviations from the normative model

B White matter volume in patients with schizophrenia

Normalized mean deviations from the normative model

Percentage of extreme negative deviations from the normative model

Percentage of extreme positive deviations from the normative model

The top panel shows a map of group-level mean deviations (|z| > 2.6). The bottom 2 panels show the percentage of extreme deviations from the normative model at each brain locus, that is, an extreme value of |z| > 2.6. On average, frontal regions, the cerebellum, and the temporal cortex had negative deviations in schizophrenia. Deviations overlapped little, with only a few brain loci showing extreme deviations in more than 2% of the patients.

can be considered to be clinically useful, especially in large samples.[29-31] The present results suggest that this outcome is possibly a result of collapsing individual patients with different biological signatures into a single diagnostic group. Furthermore, even though the biological stratification of mental disorders may be useful,[3,4,6,32] our results suggest that potentially emerging biological strata are likely smaller than previously anticipated. Interindividual differences are vast, and we estimate that those differences will not easily boil down to reliable and robust biological subtypes of mental disorders.

Published studies[33-35] that compared the estimated age with the true age of a patient have suggested that schizophrenia is

characterized by accelerated brain aging in contrast to bipolar disorder. The present results are in line with those general findings; however, we also found deviations from the normative pattern in patients with bipolar disorder. Of more importance, we found that high interindividual differences were a hallmark for both disorders in gray and white matter; however, the mapping of differences at the level of the individual patient is not possible using the brain age approach because it relies exclusively on group comparisons. A point of consideration in this context is that the individual deviations that we observed may be functionally related (eg, concentrated in functionally related brain areas). If so, our results suggest that brain networks are unlikely to be affected in

Figure 5. Characterization of Extreme Deviations From the Normative Model in Patients With Bipolar Disorder

A Gray matter volume in patients with bipolar disorder

Normalized mean deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

Percentage of extreme negative deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

Percentage of extreme positive deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

B White matter volume in patients with bipolar disorder

Normalized mean deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

Percentage of extreme negative deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

Percentage of extreme positive deviations from the normative model

x = -40   x = -20   x = 0   x = 20   x = 40

The top panel shows a map of group-level mean deviations (|z| > 2.6). The bottom 2 panels show the percentage of extreme deviations from the normative model at each brain locus, that is, an extreme value of |z| > 2.6. For bipolar disorder, deviations were less pronounced than for schizophrenia and primarily present in cerebellar regions. Deviations overlapped little, with only a few brain loci showing extreme deviations in more than 2% of patients.

the same way across patients, but different structural abnormalities may impair the working of a specific functional network via different mechanisms,[36,37] converging on similar symptoms. Although testing this hypothesis was beyond the scope of this study, we are planning work to combine functional and structural measures to better chart the nature of abnormalities in schizophrenia and bipolar disorder and aim to map the multimodal heterogeneity of both disorders.[38-40] The present work is the start of a research line that aims to systematically map the heterogeneity of mental disorders across biological readouts in the spirit of precision medicine. A logical next step is to apply clustering algorithms to the deviations from normative models to find sub-

types, similar to other subtyping approaches.[41,42] However, this step is best performed in larger samples and requires extensive validation to ensure that clusters are present in the data and are clinically relevant, although normative modeling can also be used even if there are no clearly defined subtypes in the data.[13,14]

Limitations
A limitation of this study is that it does not permit strong inferences about the degree to which certain confounding variables may have influenced our findings. We did not find evidence of confounding effects of medication but found minor associations between overall image quality and substance abuse with nega-

tive deviations in gray matter for schizophrenia (eTables 3-5 in the Supplement). However, the measures that we used to assess this effect are crude, and our design does not permit inferences about the direction of causality (eg, patients with greater severity may move more in the scanner and be more prone to drug abuse). These issues should be more fully addressed in future studies.

Conclusions

Our results have important implications for case-control designs in (neuroimaging-based) psychiatric research because we found little overlap among individual patients with the same disorder. This finding agrees with the notion that severe mental disorders are complex, with highly polygenic and multifactorial causes; the findings also provide a step toward a systematic mapping of the heterogeneity of these disorders. Although the shift from group-level psychiatry toward precision medicine has only just started, on the basis of these results, it appears that appropriate ways to incorporate interindividual differences may determine the success of the transition. Our results suggest that a full understanding of the biological features underlying schizophrenia and bipolar disorder may not be achieved by studying the average patient but by mapping patients' individual pathophysiologic signatures.

Author Affiliations: Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands (Wolfers, Franke); Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands (Wolfers, Beckmann, Marquand); Norwegian Centre for Mental Disorders Research, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway (Doan, Kaufmann, Alnæs, Moberget, Agartz, Ueland, Melle, Andreassen, Westlye); Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway (Agartz, Melle); Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, the Netherlands (Buitelaar, Beckmann, Marquand); Karakter Child and Adolescent Psychiatry University Centre, Radboud University Medical Centre, Nijmegen, the Netherlands (Buitelaar); Department of Psychology, University of Oslo, Oslo, Norway (Ueland, Westlye); Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands (Franke); Centre for Functional MRI of the Brain, University of Oxford, Oxford, United Kingdom (Beckmann); Department of Neuroimaging, Institute of Psychiatry, King's College London, London, United Kingdom (Marquand).

Author Contributions: Drs Westlye and Marquand contributed equally to this work. Mr Wolfers had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Wolfers, Beckmann, Westlye, Marquand.
Acquisition, analysis, or interpretation of data: Wolfers, Doan, Kaufmann, Alnæs, Moberget, Agartz, Buitelaar, Ueland, Melle, Franke, Andreassen, Westlye, Marquand.
Drafting of the manuscript: Wolfers, Melle, Franke, Westlye, Marquand.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Wolfers, Beckmann, Marquand.

Obtained funding: Agartz, Buitelaar, Melle, Franke, Andreassen, Westlye, Marquand.
Administrative, technical, or material support: Agartz, Ueland, Andreassen, Beckmann.
Supervision: Buitelaar, Franke, Andreassen, Beckmann, Westlye, Marquand.

REFERENCES

1. Kalia M. Biomarkers for personalized oncology: recent advances and future challenges. *Metabolism*. 2015;64(3)(suppl 1):S16-S21. doi:10.1016/j.metabol .2014.10.027

2. Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. *Nat Med*. 2016;22(11):1248-1255. doi:10.1038/nm.4190

3. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012; 17(12):1174-1179. doi:10.1038/mp.2012.105

4. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748-751. doi:10.1176/appi .ajp.2010.09091379

5. Zhu M, Zhao S. Candidate gene identification approach: progress and challenges. *Int J Biol Sci*. 2007;3(7):420-427.

6. Schumann G, Binder EB, Holte A, et al. Stratified medicine for mental disorders. *Eur Neuropsychopharmacol*. 2014;24(1):5-50. doi:10.1016 /j.euroneuro.2013.09.010

7. Buchsbaum MS, Rieder RO. Biological heterogeneity and psychiatric research. *Arch Gen Psychiatry*. 1979;36(36):1163-1169. doi:10.1001 /archpsyc.1979.01780110017001

8. Foulkes L, Blakemore SJ. Studying individual differences in human adolescent brain development. *Nat Neurosci*. 2018;21(3):315-323. doi:10.1038/s41593-018-0078-4

9. Meyer-Lindenberg A. From maps to mechanisms through neuroimaging of schizophrenia. *Nature*. 2010;468(7321):194-202. doi:10.1038/nature09569

10. Angst J, Gerber-Werder R, Zuberbühler HU, Gamma A. Is bipolar I disorder heterogeneous? *Eur Arch Psychiatry Clin Neurosci*. 2004;254(2):82-91. doi:10.1007/s00406-004-0501-6

11. Andreazza AC, Young LT. The neurobiology of bipolar disorder: identifying targets for specific agents and synergies for combination treatment. *Int J Neuropsychopharmacol*. 2014;17(7):1039-1052. doi:10.1017/S1461145713000096

12. Jauhar S, Nour MM, Veronese M, et al. A test of the transdiagnostic dopamine hypothesis of psychosis using positron emission tomographic imaging in bipolar affective disorder and

schizophrenia. *JAMA Psychiatry*. 2017;74(12):1206-1213. doi:10.1001/jamapsychiatry.2017.2943

13. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biol Psychiatry*. 2016;80(7):552-561. doi:10.1016/j.biopsych.2015.12.023

14. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2016;1(5):433-447. doi:10.1016/j.bpsc.2016.04.002

15. Simonsen C, Sundet K, Vaskinn A, et al. Neurocognitive dysfunction in bipolar and schizophrenia spectrum disorders depends on history of psychosis rather than diagnostic group. *Schizophr Bull*. 2011;37(1):73-83. doi:10.1093/schbul/sbp034

16. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Washington, DC: American Psychiatric Press; 1997.

17. Rasmussen CE, Williams CKI. Model selection and adaptation of hyperparameters. In: *Gaussian Processes for Machine Learning*. Boston, MA: MIT Press; 2006:105-128.

18. Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM. Multi-level block permutation. *Neuroimage*. 2015;123:253-268. doi:10.1016/j.neuroimage.2015.05.092

19. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57(1):289-300. doi:10.2307/2346101

20. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65-70. doi:10.2307/4615733

21. Hibar DP, Adams HHH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun*. 2017;8:13624. doi:10.1038/ncomms13624

22. Hibar DP, Westlye LT, Doan NT, et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry*. 2018;23(4):932-942. doi:10.1038/mp.2017.73

23. van Erp TGM, Hibar DP, Rasmussen JM, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry*. 2016;21(4):547-553. doi:10.1038/mp.2015.63

24. Moberget T, Doan NT, Alnæs D, et al. Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol Psychiatry*. 2017;23:1-9. doi:10.1038/mp.2017.106

25. Amann BL, Canales-Rodríguez EJ, Madre M, et al. Brain structural changes in schizoaffective disorder compared to schizophrenia and bipolar disorder. *Acta Psychiatr Scand*. 2016;133(1):23-33. doi:10.1111/acps.12440

26. Gottesman II, Shields J. A polygenic theory of schizophrenia. *Proc Natl Acad Sci U S A*. 1967;58(1):199-205. doi:10.1073/pnas.58.1.199

27. Ripke S, Neale BM, Corvin A, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

28. Fried EI. Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward. *Front Psychol*. 2015;6(MAR):309. doi:10.3389/fpsyg.2015.00309

29. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav Rev*. 2015;57:328-349. doi:10.1016/j.neubiorev.2015.08.001

30. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars [published online June 24, 2017]. *Neuroimage*. doi:10.1016/j.neuroimage.2017.06.061

31. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 2017;20(3):365-377. doi:10.1038/nn.4478

32. Weinberger DR, Goldberg TE. RDoCs redux. *World Psychiatry*. 2014;13(1):36-38. doi:10.1002/wps.20096

33. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated

brain aging in schizophrenia: a longitudinal pattern recognition study. *Am J Psychiatry*. 2016;173(6):607-616. doi:10.1176/appi.ajp.2015.15070922

34. Nenadić I, Dietzek M, Langbein K, Sauer H, Gaser C. BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Res Neuroimaging*. 2017;266:86-89. doi:10.1016/j.pscychresns.2017.05.006

35. Koutsouleris N, Davatzikos C, Borgwardt S, et al. Accelerated brain aging in schizophrenia and beyond: a neuroanatomical marker of psychiatric disorders. *Schizophr Bull*. 2014;40(5):1140-1153. doi:10.1093/schbul/sbt142

36. Honey CJ, Kötter R, Breakspear M, Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci U S A*. 2007;104(24):10240-10245. doi:10.1073/pnas.0701519104

37. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*. 2009;10(3):186-198. doi:10.1038/nrn2575

38. Groves AR, Beckmann CF, Smith SM, Woolrich MW. Linked independent component analysis for multimodal data fusion. *Neuroimage*. 2011;54(3):2198-2217. doi:10.1016/j.neuroimage.2010.09.073

39. Doan NT, Kaufmann T, Bettella F, et al. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. *Neuroimage Clin*. 2017;15:719-731. doi:10.1016/j.nicl.2017.06.014

40. Wolfers T, Arenas AL, Onnink AMH, et al. Refinement by integration: aggregated effects of multimodal imaging markers on adult ADHD. *J Psychiatry Neurosci*. 2017;42(6):386-394. doi:10.1503/jpn.160240

41. Clementz BA, Sweeney JA, Hamm JP, et al. Identification of distinct psychosis biotypes using brain-based biomarkers. *Am J Psychiatry*. 2016;173(4):373-384. doi:10.1176/appi.ajp.2015.14091200

42. Ivleva EI, Clementz BA, Dutcher AM, et al. Brain structure biomarkers in the psychosis biotypes: findings from the bipolar-schizophrenia network for intermediate phenotypes. *Biol Psychiatry*. 2017;82(1):26-39. doi:10.1016/j.biopsych.2016.08.030