Genetic Biomarkers of Psychiatric Disease: Discriminating Bipolar Disorder from Major Depressive Disorder with Polygenic Risk Scores

by

David T. Liebers

Submitted in Partial Fulfillment of the Requirements for the M.D. Degree with Honors in a Special Field at Harvard Medical School

February 10, 2020

Abstract.

Polygenic risk summary scores (PRSs) based on genome-wide association studies for psychiatric diseases have shown promise in discriminating cases from controls, predicting disease course, and response to treatment. The research presented here has sought to investigate whether such an approach may improve differentiation between bipolar disorder (BD) and major depressive disorder (MDD), a clinical distinction with important prognostic and therapeutic implications, and one which is often challenging based on clinical grounds alone. More specifically, this study tested whether psychiatric polygenic risk scores (PRSs) for bipolar disorder (BD) and schizophrenia (SCZ) improve clinically based classification models of BD-MDD diagnosis. The test sample in the present analysis included 843 BD and 930 MDD subjects similarly genotyped and phenotyped using the same standardized interview. The primary analysis tested the association of clinical risk factors and PRSs with diagnosis of BD versus MDD. A secondary analysis used multivariate modeling and receiver operating characteristic analysis to test the additive effect of PRSs on a baseline model with clinical features known to associate with BD-MDD status. PRSs drawn from BD (R^2=3.5%, p=4.94x10^{-12}) and SCZ (R^2 = 3.2%, p=5.71x10^{-11}) GWAS meta-analyses associated with BD-MDD diagnosis. Individuals with top decile BD PRS had a significantly increased risk for BD versus MDD compared with those in the lowest decile (OR=3.39, CI=2.19-5.25). PRSs improved the discriminatory ability of a symptom-based model (ΔC=0.021, p=1.05x10^{-4}) and a full model with symptoms and clinical features (ΔC = 0.011, p=6.48x10^{-4}). This

study demonstrates that psychiatric PRSs provide modest independent discrimination between BD and MDD cases, suggesting that PRS could ultimately have utility in subjects at the extremes of the distribution and subjects for whom clinical symptoms are poorly measured or yet to manifest.

Glossary

AUC – Area Under-the-Curve

BD – Bipolar Disorder

EHR—Electronic Health Record

GenRED – Genetics of Recurrent Early-Onset Depression

GWAS – Genome-wide Association Study

LD – Linkage disequilibrium

LDpred – Python based package that adjusts GWAS summary scores for LD.

MDD – Major Depressive Disorder

NIMH – National Institute of Mental Health

PCA – Principal Components Analysis

PRS – Polygenic Risk Score

SCZ – Schizophrenia

SNP – Single Nucleotide Polymorphism

Introduction.

Differentiating Unipolar and Bipolar Depressive Syndromes

Recurrent depressive syndromes, whether diagnostically linked to unipolar Major Depressive Disorder (MDD) or Bipolar Disorder (BD), affect up to a fifth of the world's population and are associated with substantial morbidity, increased risk for suicide, and premature mortality (1,2). The distinction between bipolar disorder (BD) and major depressive disorder (MDD) finds early roots in Emil Kraepelin's description of manic-depressive psychosis (3), and was subsequently validated by the discovery of pharmacologic agents with relative specificity for depressive and manic syndromes (4,5). Differentiating between bipolar and unipolar depression poses a diagnostic challenge. The greater burden of depressive symptoms in BD often leads to prolonged periods of misdiagnosis and ineffective treatment (6,7). While subjects with a clear history of mania are easily distinguished from those with MDD, many subjects with BD experience milder or atypical forms of (hypo)mania that may be difficult to recall or characterize, particularly within the context of routine clinical care (8). Moreover, the early manifestation of BD is often marked by depressive symptoms, with most subjects experiencing depressive episodes years before the emergence of manic symptoms (9).

Clinical features such as psychotic symptoms, anxiety, psychomotor retardation, greater overall severity, and earlier onset have been shown to associate with bipolar depression (10–14). However, the extent of these associations has varied across study populations. For example, higher rates of psychosis in the context of depression has not consistently associated with bipolar depression (15–17). Moreover, the availability of

studies that control for exposure to pharmacotherapy has been limited. Studies that include drug washout periods have demonstrated even less consistency with fewer characteristics and clinical features that reliably associate with bipolar versus unipolar depression (16,18,19). Thus, these nuanced differences are often insufficient for diagnostic purposes. Initial efforts to combine symptoms, illness features, and demographic risk factors into more comprehensive models of disease course have yielded greater discrimination, with AUC indexes ranging from 0.7-0.8, though these analyses have been limited by relatively small replication sets and heterogeneity of interviewing tools (11,12,20). Clinical features can be imprecise, poorly recalled, and only provide prognostic utility after their manifestation, thereby limiting their utility in prodromal and early onset cases, where the need for better diagnostic indices is arguably greatest (9). In this way, while anti-depressant monotherapy of BD is generally avoided due to the risk of inducing mania, the use of augmentation mood stabilizers in patients presenting with a severe first episode depressive episode is subject to significantly more inter-provider variability (21).

The Genetic Basis of Mood Disorders

The genetic basis of complex disorders such as BD and MDD is overwhelmingly polygenic, with hundreds and potentially thousands of contributing genetic loci distributed across the genome (22,23). Based on twin studies, the heritability of BD has been estimated to be greater than 70% (24–26), and the most recent GWAS meta-analysis has identified 30 loci that attain genome-wide significance. Despite individual loci having relatively modest association with disease status, consideration of all associated markers

in aggregate as polygenic risk scores (PRS) can provide improved predictive ability (24). This approach was first tested in schizophrenia where it was shown to predict case-control status in replication sets (27). The extent of the predictive potential is bounded by the overall heritability attributable to additive effects of common variation ("GWAS heritability"), which ranges from about ~10% for MDD to 20-30% for more heritable disorders such as BD or schizophrenia (22,28,29). At currently available sizes, GWAS studies are under-powered to detect the full range of heritability from common variation based on these estimates, with the proportion of variance estimates explained (R^2) ranging from ~12% for SCZ, to ~8% for BD and ~2-3% for MDD (22,28,29). Thus, at present GWAS discovery sample sizes and with currently available computational tools, PRSs have attained less than half of their predicted potential to discriminate cases from controls in test populations. While the estimated R^2 will increase with larger GWAS samples, PRSs already provide indices of genetic susceptibility, providing novel insights into nosological relationships (30) which may, in certain contexts, have potential future clinical implications.

Polygenic Risk Scores and Clinical Medicine

The use of PRSs in the psychiatry literature has spanned the investigation of disease endophenotypes, associations with cognitive performance, and success in creative professions (31–33). More recently, the field has pushed towards research with more immediate clinical relevance. Musliner *et al* (34) investigated how psychiatric PRSs associate with disease risk in the Danish population, finding that high MDD (30%), BD (5%), and SCZ (12%) PRSs associated with differential increased liability of a first

episode depression. Multi-site biobank consortia linked to electronic health records have allowed for the study of polygenic risk score associations at the level of health systems, demonstrating pleiotropic effects across a number of neurocognitive phenotypes as well as non-psychiatric comorbidities (35). Studies have linked PRS for SCZ with suboptimal response to antipsychotics in first episode psychosis (36), higher antipsychotic dose requirements (37), disease course after first psychotic episode (38) and poor response to lithium in BD (39). Although GWAS studies of medication "responders" versus "non-responders" may more directly address the question of pharmacogenetics (39,40), disease-based summary statistics have shown an ability to stratify for disease course intensity and likelihood of medication response. The conspectus of this work is that as the performance of PRSs increase, they may motivate randomized trials populated with individuals at the tails of the genetic risk spectrum for severity. The direction of association between PRS and drug response may not always be *a priori* predictable. In the study mentioned above, high SCZ PRS associated with poor response to antipsychotics. However, in a study of individuals with migraines, high PRS associated with improved response to triptans (41). Whichever the direction of association, for heritable complex diseases with severity that tracks with polygenic risk, stratifying by PRS may help capture signal that may otherwise be diluted across the full spectrum of genetic risk, and thus help match potential therapies to biomarker-defined sub-groups.

PRSs are being increasingly studied for prediction in a range of common medical disorders such heart disease (42), type II diabetes (43), obesity (44), and common cancers like breast (45) and prostate cancer (46). In these examples, clinical utility is

likely to be found in the tails of the PRS distribution, where increased risk is substantial and largely independent of clinically-based risk estimation tools, including family history, which suggests PRS could have a role in comprehensive models to inform decision making on prevention therapies and screening (47). Arguably the most successful application of polygenic risk score thus far has occurred in cardiovascular disorders, where population based samples have provided strong evidence that PRS can identify that proportion of the population at high risk for cardiovascular morbidity who are likely to benefit from implementation of risk mitigation strategies and pharmacological treatment (48). However, for most studied disorders, including BD, the link to specific clinical utility is unclear and will require further research based on prospective cohort studies and randomized trials. Ultimately, the most informative risk models are likely to integrate epidemiologic parameters, clinical data, non-genetic biomarkers, and potentially a broader index of genetic risk, summing the effects both common and rare variants.

Importantly, for PRSs to be clinically useful at the BD-MDD diagnostic decision point, they must complement clinical data. Discrimination between diagnosis group is less useful if clinical features that could be obtained from a patient interview have similar, non-orthogonal discriminative power. While genetic information has advantages over other kinds of clinical information and even other biomarkers which may fluctuate with time, PRSs should be evaluated in the context of models that integrate all available forms of clinical data.

In this study, we explored the degree to which polygenic risk scores may aid in distinguishing bipolar from unipolar depression, arguably one of the most important

clinical decisions in psychiatry, since the prognosis and optimal treatment of either disorder is likely to be different for most patients. We have specifically selected samples that were similarly genotyped and phenotyped, thereby allowing us to test the effect of PRSs individually, but also within the broader context of clinical risk factors that are typically used to help with clinical decision making.

Materials and Methods.

Subjects

The sample included BD and MDD cases from National Institute of Mental Health (NIMH) Bipolar Disorders Initiative and Genetics of Recurrent Early-Onset Depression (GenRED) collections, which were initially ascertained for genetic studies and have undergone previous case-control GWAS analyses (49,50). Both studies used the Diagnostic Interview for Genetic Studies (DIGS) and similar best estimate procedures (51). Subjects were restricted to high-confidence diagnoses of BD (more specifically, bipolar I disorder) or MDD genotyped on the Affymetrix 6.0 Microarray. All subjects were unrelated, of European ancestry, and independent of the initial training set GWAS studies. We further selected samples with non-missing data for all seven symptoms and clinical features (psychomotor retardation, incapacitation, delusions and number of mixed symptoms during worst depressive episode, length of most severe episode, number of depressive episodes, antidepressant induced “high” feeling) previously shown to discriminate BD from MDD in a larger but partially overlapping dataset (55% of the samples used in the present analysis were part of the training dataset in the previous

analysis) (11). The final sample included a total of 1773 subjects, consisting of 843 with BD and 930 with recurrent MDD. Samples were previously collected under Institutional Review Board (IRB) approved protocols (49,50) that included informed consent for genetic studies and data sharing.

Genome-wide Association Study (GWAS)

Cases with BD and MDD as described above were previously genotyped using the Affymetrix 6.0 micro-array. Raw genotypes were combined using PLINK1.9 (52) followed by sequential quality control (QC) steps removing: (1) SNPs with missing rates ≥ 5%; (2) individuals with overall missing rates ≥ 2%; (3) remaining SNPs with missing rates ≥ 2%; (4) SNPs in Hardy-Weinberg disequilibrium ($P < 0.000001$); and remaining SNPS with differential missing rates >2%. The filtered combined file underwent additional QC/harmonization steps (https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.2.7.zip) prior to imputation to the multiethnic 1000Genomes version 3 reference panel. Imputation was performed using the Michigan Imputation Server and high-quality genotypes ($r^2 > 0.6$) with a minor allele frequency of 1% or greater were extracted and converted to PLINK format for downstream analyses. Principal components analysis (PCA) was performed using PLINK1.9 by combining the filtered genotypes with the 1000G multi-ethnic panel and extracting the top 10 principal components. We excluded 5 samples that were outliers in the PCA analysis (Supplementary Figure 1).

Calculation of PRSs

PRS were calculated using recommended parameters in LD-Pred (53), which utilizes a Bayesian approach to infer the mean effect at each individual marker while accounting for linkage disequilibrium (LD) between markers. We utilized summary statistics from the largest meta-analysis GWAS of BD (22) and SCZ (28) as the training dataset sets and the European subjects from the 1000 Genomes version 3 as the linkage-disequilibrium (LD) reference genotypes. Using an LD-radius of 1775 and 1981, weighted risk scores were constructed for 5,325,407 and 5,943,410 markers in the harmonized BD and SCZ datasets, respectively. PRS were derived across nine training dataset thresholds, defined as the assumed "fraction of causal variants" (--PS=1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0001), and standardized to reflect the standard deviation as the unit of measurement. PRS thresholds were selected maximize the explained difference between BD and MDD phenotypes. For SCZ, this optimal p-value threshold (0.3) was the same as the one used in the original study validating LDPred as method of PRS calculation (53).

Statistical Analysis

All analyses were performed in R version 3.4.2. Association of PRSs with phenotype status was performed with a generalized linear model using a logit link function. We initially fit baseline models with BD-MDD status as the dependent variable and the first ten principal components (PCs), age at interview, and sex as covariates. We subsequently added the PRS variable into that base model and compared the difference in Nagelkerke's pseudo-R^2 (ΔR^2) to estimate proportion of variance attributable to the PRS. For ease of visualization, the sample was stratified into deciles using the residuals

of the top performing BD and SCZ PRSs adjusted for first 10 PCs, age, and sex. Odds ratios (OR) for BD vs MDD status and 95% confidence intervals (CIs) were calculated for each decile in comparison with the first decile.

We performed a receiver operating characteristic (ROC) analysis to estimate classifying potential of PRS. We initially fit a logistic model with the first 10 PCs, age, and sex, and then measured the c-index. To this model we serially added (1) top performing BD and SCZ PRSs, (2) each symptom/clinical feature (psychomotor retardation, incapacitation, delusions, length of most severe episode, number of depressive episodes, presence of high feeling after antidepressant, and number of mixed symptoms), both separately and then with PRSs to measure the additive effect (3) a set of symptoms that might plausibly be ascertained from a psychiatric interview on initial presentation with a depressive episode (psychomotor retardation, delusions, and number of mixed symptoms) both as a clinical model and with PRSs, and all clinical features/symptoms taken together and PRSs.

Lastly, we tested the association of the top performing BD and SZ PRS with each of the symptoms/clinical features shown listed above, controlling for diagnosis. Logistic regression was used for binary variables (presence of psychomotor retardation, delusions, incapacitation, and high feeling after antidepressant) and linear regression for continuous variables (length of worst depression in weeks, number of episodes, and number of mixed symptoms).

Results.

Clinical and Demographic Features

Of the 1773 subjects, 843 had BD and 930 had MDD, with more females than males in both groups (66.0% and 70.0%, respectively) (Table 1). Subjects with BD were slightly older at the time of their interview (42.0 years BD versus 40.2 years MDD), had fewer years of education (14.9 BD versus 16.0 MDD), and were less likely to be married (29.8% BD versus 47.8% MDD). Consistent with our previous study (11), subjects with BD were more likely to have experienced psychomotor retardation and delusions, a shorter length depressive episode, incapacitation, a greater number of depressive episodes, high feeling after antidepressant treatment, and a greater number of mixed symptoms (Table 1).

Polygenic Association with BD vs. MDD

Given the strong co-heritability between bipolar disorder and schizophrenia (26) we performed polygenic analyses utilizing the largest meta-analyses of each disorder to provide the specific markers and weights for polygenic risk scoring. Notably, both BD and MDD samples were not part of the GWAS meta-analyses, assuring independence across the training and testing datasets. The p-value thresholds for BD (0.003) and SCZ (0.3) PRSs were chosen to maximize association with BD-MDD status (Supplementary Table 1). In our primary analysis, we found that PRS for BD was strongly associated with BD compared with MDD (ΔR^2=3.53%, p=4.94x10^{-12}), with an effect size that was approximately half of that usually seen in BD case-control studies. Similarly, PRS for SCZ was aassociated with BD versus MDD phenotype (ΔR_2=3.16%, p=5.71x10^{-11}). The

distribution of PRS among BD and MDD samples showed prominent overlap, with greater discriminative ability found at the more extreme values of PRS, as has been observed in other common complex disorders (Figure 1). Using PRSs to help classify the overall sample, we found that a baseline model with covariates and PRS produced an AUC c-statistic of 0.65 for the BD PRS and 0.64 for the SCZ PRS. For ease of interpretability, we divided the samples into PRS deciles, and found that individuals with PRS in the top decile had an OR of 3.39 (CI: 2.19-5.25) when compared to the lowest decile. Similarly, those in the top decile of SCZ PRS residuals had an OR of 3.07 (CI: 1.99-4.73) compared to the lowest decile, in terms of BD-MDD status (Figure 2).

Genetic risk and Clinical Features

We found that PRSs discriminated BD-MDD status to a degree comparable to many of the individual symptoms and clinical features tested. PRSs contributed marginal additive effects to all individual clinical features, providing evidence that prediction based on PRSs is likely to be independent of that based on clinical features (Figure 3). When combined, the clinical features showed a more significant AUC, consistent with our prior study (11). Inclusion of the PRSs to a symptom-based model and a full model with symptoms and clinical features led to statistically significant increases in discriminatory ability ($\Delta C = 0.021$, $p = 1.05 \times 10^{-4}$ and $\Delta C = 0.011$, $p = 6.48 \times 10^{-4}$, respectively). The classifying ability of PRSs in models with individual clinically-based predictors varied based on the independence of these variables. For example, PRSs provided a larger increase in discriminatory ability in a model with delusions alone ($\Delta C = 0.042$) than

incapacitation alone ($\Delta C = 0.021$) (Supplementary Table 2). The PRS association remained strongly significant even with the full multi-variate clinical model (Table 2).

Association of PRS with Symptoms / Clinical Features

Finally, in order to further explore effect of polygenic risk, we sought to formally test whether the polygenic signal was differentially associated with specific clinical features, thereby providing a more granular view of overall genetic susceptibility. Using a logistic regression model adjusted for covariates that included the main effect of diagnosis, we found a single significant association with BD PRS and incapacitation (Table 3). In contrast, the SCZ PRS was associated with longer length of depression ($p=0.0043$), but not with any of the additional risk factors, suggesting significant heterogeneity despite comparable predictive effects.

Summary, Future Work and Discussion

Although mania remains the cardinal syndrome that defines the diagnosis of BD, subjects often first present with years of depressive symptoms (54). Consequently, subjects who are eventually diagnosed with BD are often misdiagnosed on initial presentation, and may receive treatments of limited efficacy (55) that could be associated with iatrogenic harm (7). Clinical symptoms, particularly in the prodromal phase of BD, are often poorly captured and insufficiently prognostic, particularly in primary care settings (8). Patients often lack insight into the nature of their manic symptoms, which can delay diagnosis (56). Hence, there is a clear need for better prognostic tools that can help to

stratify subjects are higher risk for developing BD. In this study, we have tested whether markers of genetic risk (as indexed by PRSs) can augment traditionally used clinically-based factors, thereby helping to identify subjects with depression who may be at high risk for "converting" to bipolar disorder. We found that PRSs derived from the most recent PGC GWAS discriminate modestly between BD and MDD cases, with BD PRS demonstrating a slightly higher association than SCZ PRS. Subjects with top decile BD PRS were more than three times as likely to have BD than MDD, compared with those in the lowest BD PRS decile. This level of discrimination was similar to that found in a recent study of PRS in first episode psychosis (38). PRSs had comparable classifying ability to individual clinical features and symptoms previously shown to associate with BD, although when included in a model with all clinical features and symptoms, the increase in overall classification was more modest. The effect of PRSs appeared to be largely independent of the clinical features, suggesting that PRSs may represent orthogonal risk factors that may provide additional information beyond traditional clinical factors.

Moving Polygenic Risk Scores to the Clinic

For psychiatric disorders there is a strong need for reliable biomarkers of disease etiology and progression due to the complexity and inaccessibility of brain tissue, and an imprecise diagnostic taxonomy based exclusively on clinical phenomenology (57). While genetic data has limitations, most notably the constraints placed by heritability estimates, it may also hold advantages over clinically derived predictors, in view of the longitudinal time-course where most psychiatric disorders manifest, and polymorphic clinical manifestations that are often protean, time-limited, and poorly recalled. Unlike clinical

predictors, genetic predictors can be measured before the onset of symptoms and can be informative during the life-course. In contrast, several of the clinical features included in our model, such as number of lifetime episodes and the duration of depressive episode represents information that would not be available during a first episode or prior to the full manifestation of the disorder. Moreover, as discovery sample sizes increase, the divergence in risk between those at the tails of the PRS distribution will likely be greater. Improved computational tools that more accurately adjust for the effects of linkage disequilibrium in the generation of weights have shown promise in their improved ability to discriminate cases and controls in test sets (58). Combined, the increasing size of GWAS and improved computational, together with the inclusion of rarer variation, will push PRSs closer to their full potential as predicted by GWAS heritability estimates. These improved PRSs will likely be of greater clinical value.

Future Research

Here, I will highlight two areas of future research building on this preliminary study on the use of PRSs in discriminating mood disorder disease course.

First, the harmonization of EHR data across health systems with associated biobanks allows for the evaluation of granular patient data, longitudinally. Such datasets facilitate research on questions about how high polygenic risk scores may predict the accrual of psychiatric diagnoses, the presentation of psychiatric phenomenon, and the prescription of psychotropics over time. The cost of misdiagnosis may also be estimated.

Second, it will be important to prepare the ground for future prospective randomized trials in psychiatry that integrate genetic information. A body of research has

shown that delivery of genetic tests results in a variety of clinical settings does not have adverse consequences in terms of depression or anxiety. This had been a major concern in the era of whole-genome and whole-exome screening for research purposes (59,60). However, it is unclear whether this would hold true in patients who are at already higher risk for psychiatric disease. Individuals who present with a first depressive episode, or other psychiatric symptoms, may benefit from risk stratification with a polygenic risk score, however, they may be less inclined to choose to be tested for polygenic risk of more severe psychiatric disease due to reasons associated with historical stigma. Those that do choose to pursue testing may be less well equipped to assimilate test results than those presented with test results when they are not depressed.

In the community, physician interest and comfort with the use of PRSs for assisting with the management of comfort disease risk is still developing. Research that better delineates clinical decisions that could be nudged in one direction or another with PRSs, may help motivate clinical trials that are likely to influence physician decision making. PRSs may change a psychiatrist's threshold to augment anti-depressant therapy with lithium in patients with depression and high BD PRS. Similarly, high PRSs for severe psychiatric disease may adjust a primary care provider's threshold for referring to psychiatry or therapy in patients with first episode depression. A more thorough understanding of how physicians would respond to PRS, would increase the likelihood of defining clinically relevant thresholds for prospective studies.

Many psychiatric diseases emerge during adolescence, a period of substantial neurobiological change. If disease modifying therapies are developed based on targeting

genetic factors identified through GWAS for psychiatric disease, the timing of the delivery of such therapies will likely need to be at or before disease onset. This represents another important future role for PRSs in predicting disease and adds urgency to developing paradigms around ethical and practical concerns about the use of PRS in clinical decision-making and the design of prospective trials.

Limitations

An important strength of our study is that rather than comparing cases and healthy controls, we focus on the more clinically relevant comparison between two disorders with symptom overlap yet require relatively distinct forms of pharmacotherapy. As such, we therefore restricted our sample to subjects with either disorder who were diagnosed with identical diagnostic instruments and genotyped with the same microarray chip, in order to mitigate confounding. Additionally, our current analysis focused on the more realistic assumption that polygenic risk scores will ultimately be used in conjunction with traditional clinical risk factors, rather than studied or applied in isolation. We therefore applied polygenic risk scoring to previously identified clinical factors that had shown replicable associations in a prior comparison of BD and MDD (11). Our results therefore may have greater face validity while providing a more contextual view of where PRS may have clinical utility in future studies.

As an initial proof of concept study, there are also several important limitations. We have analyzed opportunistically samples previously collected for genetic studies that focused on the ascertainment of the more severe types of BD and on recurrent forms of MDD. Further work should test the performance of psychiatric PRS in discriminating MDD

and BD cases in more naturalistic clinical cohorts that include the full manifestation of the BD-MDD mood spectrum, particularly in the early phases of illness (61). Second, our clinical features were derived from a single cross-sectional evaluation that is subject to typical retrospective limitations, including recall bias (62). Although best-estimate procedures include additional information to supplement the interview, it is possible that certain symptoms may have been under-reported. Third, we limited our focus to clinical factors to available in our shared diagnostic instrument identified by our previous study (11), whereas a more comprehensive diagnostic model may ultimately include additional risk and state related factors that may only be reliably assessed using a prospective cohort. Lastly, our study was performed in samples of European ancestry. It is likely that these PRSs would be less predictive in non-European samples, which reflects a central limitation of integrating currently available PRS into clinical medicine, and one which has consequences for health equity (63).

Tables and Figures

Table 1. Characteristics of Bipolar Disorder and Major Depressive Disorder samples (n=1773).

	BD (n=843)	MDD (n=930)
Gender, n (%)		
Female	556 (66.0)	651 (70.0)
Age at interview, mean (s.d.)	42.0 (12.3)	40.2 (11.8)
Years education, mean (s.d.)	14.9 (2.7)	16.0 (2.8)
Married, n (%)	250 (29.8)	352 (47.8)
Length in weeks, mean (s.d.)	35.8 (93.7)	81.9 (179.0)
Number of lifetime episodes, mean (s.d.)	22.1 (51.2)	8.0 (14.0)
Number of mixed symptoms, mean (s.d.)	1.3 (2.3)	0.5 (1.0)
Psychomotor retardation, number (%)	492 (58.4)	304 (32.7)
Delusions, number (%)	194 (23.0)	27 (3.0)
High after antidepressant, number (%)	357 (42.3)	105 (11.3)
Incapacitation, number (%)	672 (79.7)	382 (41.1)

Table 2: Odds Ratios of BD versus MDD Status for Clinical Features and Polygenic Risk Scores. Results of multivariate logistic regression with clinical features, symptoms and PRS as predictors of BD-MDD status (OR=odds ratios, CIL= lower bound 95% confidence interval, CIU=upper bound 95% confidence interval). Adjusted for 10 PCs, age and sex. Units for PRS is per standard deviation.

Feature	**OR**	**CIL**	**CIU**	**z-value**	**p-value**
PMR	2.45	1.92	3.13	7.20	5.99×10^{-13}
Del	5.71	3.58	9.43	7.10	1.29×10^{-12}
Mixed	1.26	1.16	1.37	5.53	3.18×10^{-8}
Episodes	1.02	1.01	1.03	5.56	2.66×10^{-8}
Weeks	0.99	0.99	1.00	-5.12	3.06×10^{-7}
High	3.70	2.76	4.97	8.76	$<2 \times 10^{-16}$
Incap	3.41	2.65	4.40	9.49	$<2 \times 10^{-16}$
BD PRS[a]	1.39	1.19	1.61	4.32	1.57×10^{-5}
SCZ PRS[a]	1.46	1.22	1.73	4.21	2.57×10^{-5}

[a]Per standard deviation from the mean.

Table 3. PRS Association with Clinical Features. Association between top performing bipolar disorder (BD) and schizophrenia (SCZ) polygenic risk scores and individual clinical features (model including diagnosis, first 10 principal components, age and sex). Top performing PRSs are defined as those calculated from p-value thresholds for BD (0.003) and SCZ (0.3), which maximized association with BD-MDD status. Results from logistic regression with Z statistics are reported for binary variables: presence of psychomotor retardation, delusions, incapacitation, and high feeling after antidepressant treatment. Results from linear regression with t-statistic are reported for continuous variables (length of worst depression in weeks, number of episodes, and number of mixed symptoms).

PRS	**SCZ**			**BD**		
	Estimate	**SE**	**P**	**Estimate**	**SE**	**P**
Length worst depression	11.264	3.942	0.004[a]	2.044	4.607	0.657
Psychomotor retardation	0.003	0.061	0.958	0.068	0.067	0.309
Number of mixed symptoms	-0.008	-0.156	0.876	-0.118	0.056	0.035
Incapacitation	0.003	0.061	0.958	0.214	0.072	0.003[a]
High after antidepressant	-0.056	0.067	0.404	-0.040	0.079	0.610
Number of episodes	-0.505	1.001	0.614	-1.122	1.167	0.336
Delusions	0.037	0.087	0.672	-0.257	0.104	0.013

[a]significant after Bonferroni correction ($p<0.007$) for multiple tests.

Figure 1. Density distribution of bipolar disorder (BD) (a) and schizophrenia (SCZ) (b) polygenic risk scores (PRSs) in BD and major depressive disorder (MDD) samples. PRS represents standardized residuals after adjustment for first ten principal components, age and sex. Dashed=MDD; solid= BD.

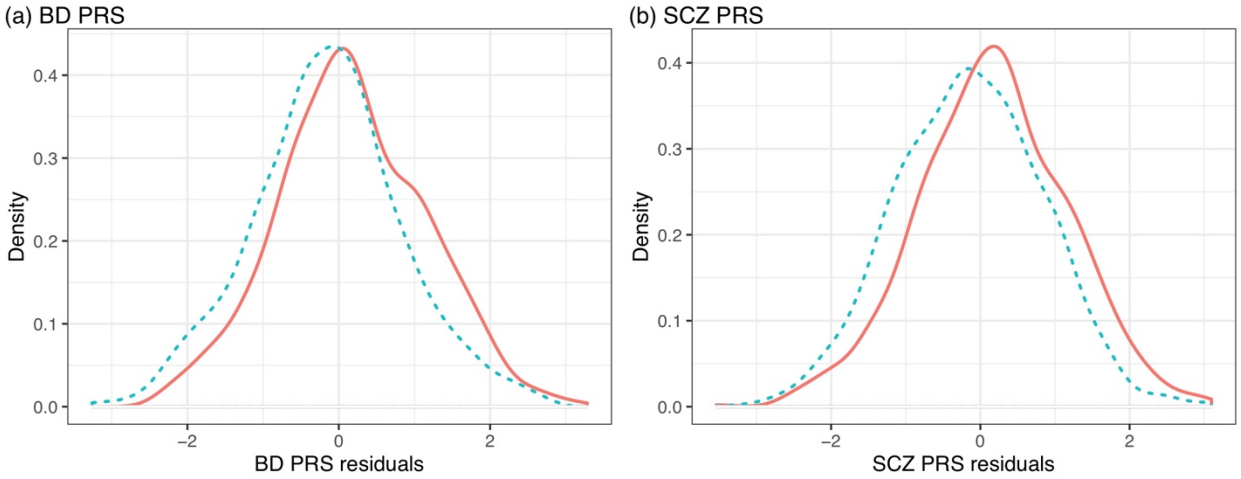


Figure 2. Odds ratios for bipolar disorder (BD) versus major depressive disorder and confidence intervals by decile for top BD and schizophrenia polygenic risk score predictors (using residuals adjusting for first 10 principal components, age and sex). Odds ratios measured against lowest decile.

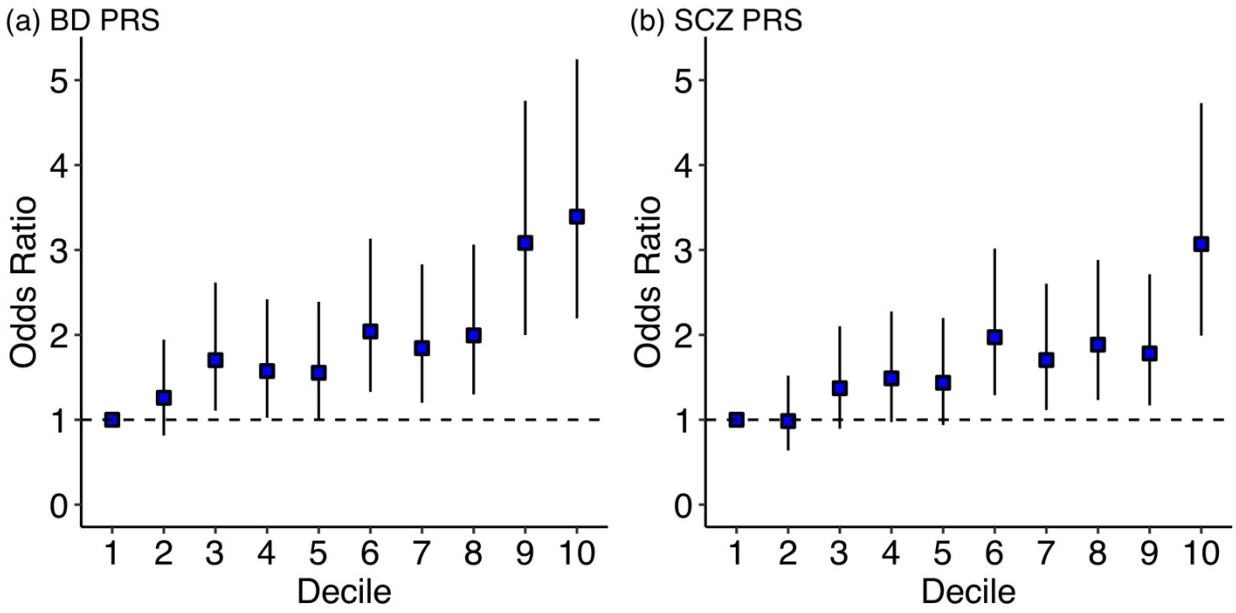
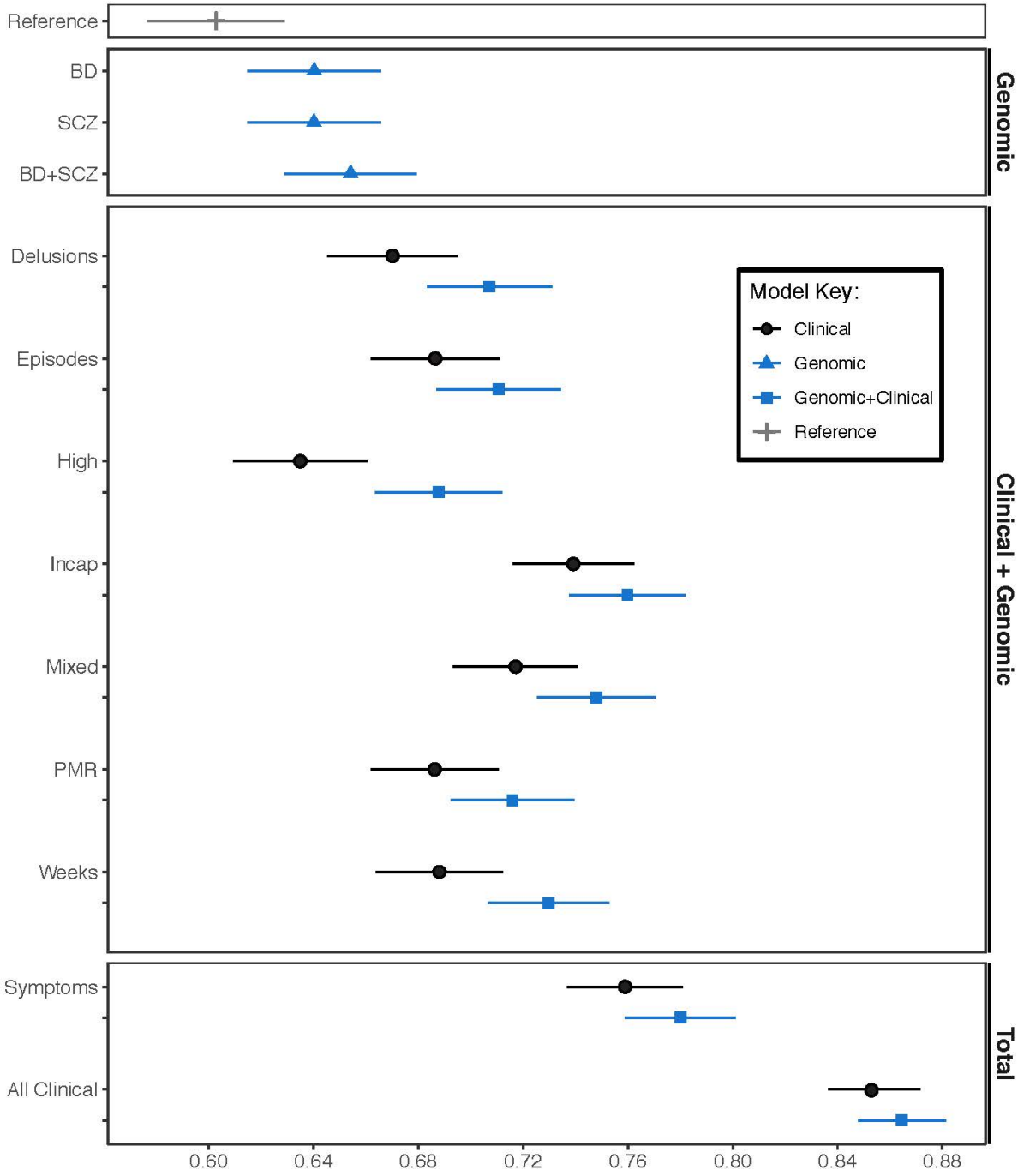


Figure 3. C-indices for models classifying bipolar disorder (BD) and major depressive disorder (MDD) showing the additive effect of genetic information to clinical models. The figure is divided into four sections. (1) Reference model including 10 principal components, age and sex. (2) Models including polygenic risk scores (PRSs) from bipolar disorder (BD) and schizophrenia (SCZ) GWAS. (3) Models with each clinical feature or symptom, individually, with (blue) and without (black) PRSs. (4) “Symptoms,” or data which could be ascertained at first interview (number of mixed symptoms, presence of delusions and psychomotor retardation), and a total model with all symptoms and clinical features, with and without PRSs. All models included 10 principal components, age and sex. High=High after antidepressant treatment; Incap=incapacitation during worst depressive episode; Mixed=Number of mixed symptoms during worst depressive episode; PMR=Psychomotor retardation during worst depressive episode; Weeks=Length of longest depressive episode.


Reference
BD
SCZ
BD+SCZ
Genomic
Delusions
Episodes
High
Incap
Mixed
PMR
Weeks
Model Key:
Clinical
Genomic
Genomic+Clinical
Reference
Clinical + Genomic
Symptoms
All Clinical
Total
0.60
0.64
0.68
0.72
0.76
0.80
0.84
0.88

Supplementary Table 1. ΔR^2 for model with PRS as classifier of BD-MDD (against baseline model with first 10 PCs, sex and age at interview). BD=PRS derived from Bipolar Disorder GWAS. SCZ=PRS derived from Schizophrenia GWAS.

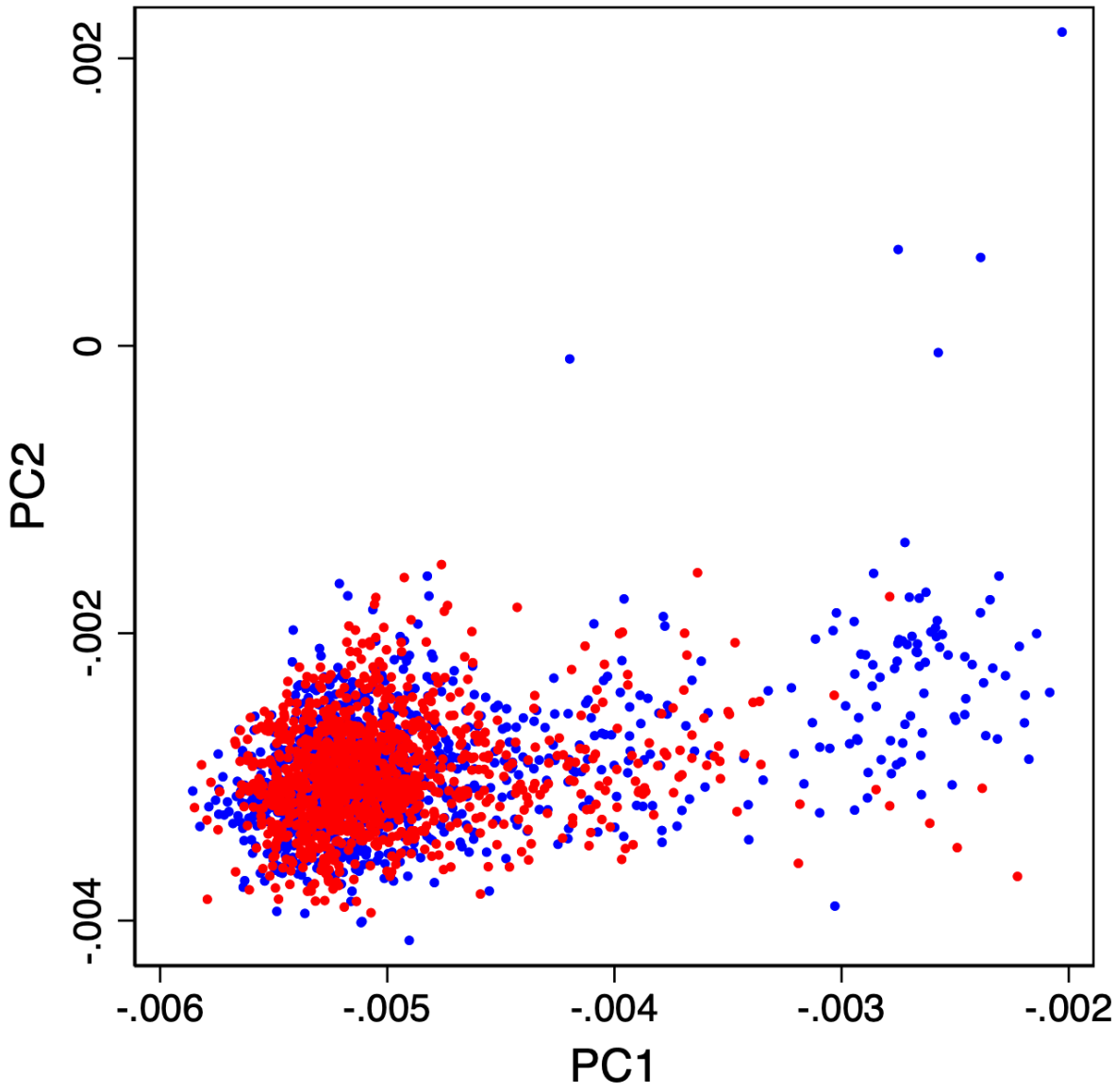
P-value Threshold	**R^2**	**ΔR^2 [a]**	**OR**	**SE**	**Z**	**P**
baseline	0.0610					
BD: 0.0001	0.0648	0.0038	1.12	0.0504	2.29	2.21×10^{-2}
BD: 0.0003	0.0617	0.0007	0.949	0.0523	-1.00	3.15×10^{-1}
BD: 0.001	0.0622	0.0012	0.934	0.0526	-1.31	1.92×10^{-1}
BD: 0.003	0.0963	0.0353	1.59	0.0672	6.91	4.94×10^{-12}
BD: 0.01	0.0942	0.0332	1.58	0.0686	6.71	1.97×10^{-11}
BD: 0.03	0.0917	0.0307	1.56	0.0685	6.46	1.08×10^{-10}
BD: 0.1	0.0907	0.0297	1.54	0.0683	6.36	2.05×10^{-10}
BD: 0.3	0.0903	0.0293	1.54	0.0683	6.31	2.73×10^{-10}
BD: 1	0.0902	0.0292	1.54	0.0682	6.30	3.01×10^{-10}
SCZ	**R2**	**ΔR^2**	**OR**	**SE**	**Z**	**P**
SCZ: 0.0001	0.0636	0.0026	1.14	0.0693	1.89	5.87×10^{-2}
SCZ: 0.0003	0.0619	0.0009	1.06	0.0488	1.13	2.59×10^{-1}
SCZ: 0.001	0.0612	0.0002	1.03	0.0504	0.546	5.85×10^{-1}
SCZ: 0.003	0.0655	0.0045	1.14	0.0540	2.49	1.28×10^{-2}
SCZ: 0.01	0.0643	0.0033	1.12	0.0541	2.15	3.16×10^{-2}
SCZ: 0.03	0.0850	0.024	1.38	0.0557	5.72	1.06×10^{-8}
SCZ: 0.1	0.0906	0.0296	1.44	0.0571	6.34	2.29×10^{-10}
SCZ: 0.3	0.0926	0.0316	1.46	0.0576	6.55	5.71×10^{-11}
SCZ: 1	0.0918	0.0308	1.45	0.0577	6.47	9.53×10^{-11}

[a] Difference baseline model and model with PRS.

Supplementary Table 2. C-index of a reference model, and models with polygenic risk scores (PRSs), clinical features, and a combination of clinical features and PRSs that associate with BD-MDD status. Symptoms include parts of history that likely could be obtained from a psychiatric history during a first episode depression (PMR + Mixed + Delusions). CIL=Lower bound of 95% confidence interval. CIU=Upper bound of 95% confidence interval.

Model	**C**	**CIL**	**CIU**	**Type**	**ΔC (with PRS)**
Reference	0.603	0.577	0.629		-
SCZ	0.640	0.615	0.666	Genomic	0.037
BD	0.640	0.615	0.666	Genomic	0.038
BD + SCZ	0.654	0.629	0.679	Genomic	0.051
Weeks	0.670	0.645	0.695	Clinical	-
Weeks + BD + SCZ	0.707	0.683	0.731	Combined	0.037
PMR	0.687	0.662	0.711	Clinical	-
PMR + BD + SCZ	0.711	0.687	0.734	Combined	0.024
Mixed	0.635	0.609	0.661	Clinical	-
Mixed + BD + SCZ	0.688	0.664	0.712	Combined	0.053
Incapacitated	0.739	0.716	0.762	Clinical	-
Incapacitation + BD + SCZ	0.760	0.738	0.782	Combined	0.021
High	0.717	0.693	0.741	Clinical	-
High + BD + SCZ	0.748	0.726	0.771	Combined	0.031
Episodes	0.686	0.662	0.711	Clinical	-
Episodes + BD + SCZ	0.716	0.693	0.739	Combined	0.030
Delusions	0.688	0.664	0.712	Clinical	-
Delusions + BD + SCZ	0.730	0.707	0.753	Combined	0.042
Symptoms	0.759	0.737	0.781	Clinical	
Symptoms + BD + SCZ	0.780	0.759	0.801	Combined	0.021
All Clinical	0.854	0.837	0.872	Clinical	-
All Clinical + BD	0.862	0.845	0.879	Combined	0.0076
All Clinical + SCZ	0.861	0.845	0.879	Combined	0.0074
All Clinical + BD + SCZ	0.865	0.848	0.881	Combined	0.011

Supplementary Figure 1. PCA demonstrating outliers that were removed prior to analysis.



References

1. Kessler RC. The Costs of Depression. Psychiatric Clinics of North America. 2012 Mar;35(1):1–14.

2. Bauer M, Andreassen OA, Geddes JR, Vedel Kessing L, Lewitzka U, Schulze TG, et al. Areas of uncertainties and unmet needs in bipolar disorders: clinical and research perspectives. The Lancet Psychiatry. 2018 Nov;5(11):930–9.

3. Leonhard K, Korff I, Schulz H. [Temperament in families with monopolar and bipolar phasic psychoses]. Psychiatr Neurol (Basel). 1962;143:416–34.

4. Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, et al. The International Society for Bipolar Disorders (ISBD) Task Force Report on Antidepressant Use in Bipolar Disorders. AJP. 2013 Nov;170(11):1249–62.

5. Altshuler LL, Post RM, Leverich GS, Mikalauskas K, Rosoff A, Ackerman L. Antidepressant-induced mania and cycle acceleration: a controversy revisited. Am J Psychiatry. 1995 Aug;152(8):1130–8.

6. Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RMA. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. Journal of Affective Disorders. 1994 Aug;31(4):281–94.

7. Baldessarini RJ, Faedda GL, Offidani E, Vázquez GH, Marangoni C, Serra G, et al. Antidepressant-associated mood-switching and transition from unipolar major depression to bipolar disorder: A review. Journal of Affective Disorders. 2013 May;148(1):129–35.

8. Das AK. Screening for Bipolar Disorder in a Primary Care Practice. JAMA. 2005 Feb 23;293(8):956.

9. Van Meter AR, Burke C, Youngstrom EA, Faedda GL, Correll CU. The Bipolar Prodrome: Meta-Analysis of Symptom Prevalence Prior to Initial or Recurrent Mood Episodes. Journal of the American Academy of Child & Adolescent Psychiatry. 2016 Jul;55(7):543–55.

10. Mitchell PB, Frankland A, Hadzi-Pavlovic D, Roberts G, Corry J, Wright A, et al. Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. Br J Psychiatry. 2011 Oct;199(4):303–9.

11. Leonpacher AK, Liebers D, Pirooznia M, Jancic D, MacKinnon DF, Mondimore FM, et al. Distinguishing bipolar from unipolar depression: the importance of clinical symptoms and illness features. Psychol Med. 2015 Aug;45(11):2437–46.

12. Gan Z, Diao F, Wei Q, Wu X, Cheng M, Guan N, et al. A predictive model for diagnosing bipolar disorder based on the clinical characteristics of major depressive episodes in Chinese population. Journal of Affective Disorders. 2011 Nov;134(1–3):119–25.

13. Frankland A, Roberts G, Holmes-Preston E, Perich T, Levy F, Lenroot R, et al. Clinical predictors of conversion to bipolar disorder in a prospective longitudinal familial high-risk sample: focus on depressive features. Psychol Med. 2018 Jul;48(10):1713–21.

14. Cuellar AK, Johnson SL, Winters R. Distinctions between bipolar and unipolar depression. Clin Psychol Rev. 2005 May;25(3):307–39.

15. Breslau N, Meltzer HY. Validity of subtyping psychotic depression: examination of phenomenology and demographic characteristics. Am J Psychiatry. 1988 Jan;145(1):35–40.

16. Beigel A, Murphy DL. Unipolar and bipolar affective illness. Differences in clinical characteristics accompanying depression. Arch Gen Psychiatry. 1971 Mar;24(3):215–20.

17. Black DW, Nasrallah A. Hallucinations and delusions in 1,715 patients with unipolar and bipolar affective disorders. Psychopathology. 1989;22(1):28–34.

18. Kupfer DJ, Weiss BL, Foster G, Detre TP, McPartland R. Psychomotor activity in affective states. Arch Gen Psychiatry. 1974 Jun;30(6):765–8.

19. Katz MM, Robins E, Croughan J, Secunda S, Swann A. Behavioural measurement and drug response characteristics of unipolar and bipolar depression. Psychol Med. 1982 Feb;12(1):25–36.

20. Schaffer A, Isometsä ET, Tondo L, Moreno DH, Sinyor M, Kessing LV, et al. Epidemiology, neurobiology and pharmacological interventions related to suicide deaths and suicide attempts in bipolar disorder: Part I of a report of the International Society for Bipolar Disorders Task Force on Suicide in Bipolar Disorder. Aust N Z J Psychiatry. 2015 Sep;49(9):785–802.

21. Barowsky J, Schwartz TL. An Evidence-Based Approach to Augmentation and Combination Strategies for: Treatment-Resistant Depression. Psychiatry (Edgmont). 2006 Jul;3(7):42–61.

22. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019 May;51(5):793–803.

23. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Research. 2007 Sep 4;17(10):1520–8.

24. Kendler KS, Pedersen NL, Neale MC, Mathé AA. A pilot Swedish twin study of affective illness including hospital- and population-ascertained subsamples: results of model fitting. Behav Genet. 1995 May;25(3):217–32.

25. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder: AMERICAN JOURNAL OF MEDICAL GENETICS (SEMIN. MED. GENET.). Am J Med Genet. 2003 Nov 15;123C(1):48–58.

26. The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018 Jun 22;360(6395):eaap8757.

27. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature [Internet]. 2009 Jul 1 [cited 2019 Apr 20]; Available from: http://www.nature.com/doifinder/10.1038/nature08185

28. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50(3):381–9.

29. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions [Internet]. Genetics; 2018 Oct [cited 2019 Jul 28]. Available from: http://biorxiv.org/lookup/doi/10.1101/433367

30. The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018 Jun 22;360(6395):eaap8757.

31. Liebers DT, Pirooznia M, Seiffudin F, Musliner KL, Zandi PP, Goes FS. Polygenic Risk of Schizophrenia and Cognition in a Population-Based Survey of Older Adults. Schizophr Bull. 2016;42(4):984–91.

32. Ranlund S, Calafato S, Thygesen JH, Lin K, Cahn W, Crespo-Facorro B, et al. A polygenic risk score analysis of psychosis endophenotypes across brain functional, structural, and cognitive domains. Am J Med Genet. 2018 Jan;177(1):21–34.

33. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci. 2015 Jul;18(7):953–5.

34. Musliner KL, Mortensen PB, McGrath JJ, Suppli NP, Hougaard DM, Bybjerg-Grauholm J, et al. Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. JAMA Psychiatry. 2019 May 1;76(5):516.

35. Zheutlin AB, Dennis J, Karlsson Linnér R, Moscati A, Restrepo N, Straub P, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. AJP. 2019 Oct 1;176(10):846–55.

36. Zhang J-P, Robinson D, Yu J, Gallego J, Fleischhacker WW, Kahn RS, et al. Schizophrenia Polygenic Risk Score as a Predictor of Antipsychotic Efficacy in First-Episode Psychosis. American Journal of Psychiatry. 2019 Jan;176(1):21–8.

37. Hettige NC, Cole CB, Khalid S, De Luca V. Polygenic risk score prediction of antipsychotic dosage in schizophrenia. Schizophr Res. 2016 Feb;170(2–3):265–70.

38. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biological Psychiatry. 2017 Mar;81(6):470–7.

39. International Consortium on Lithium Genetics (ConLi+Gen), Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, et al. Association of Polygenic Score for Schizophrenia and HLA Antigen and Inflammation Genes With Response to Lithium in Bipolar Affective Disorder: A Genome-Wide Association Study. JAMA Psychiatry [Internet]. 2017 Nov 9 [cited 2019 Jul 28]; Available from: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/jamapsychiatry.2017.3433

40. Guo W, Machado-Vieira R, Mathew S, Murrough JW, Charney DS, Grunebaum M, et al. Exploratory genome-wide association analysis of response to ketamine and a polygenic analysis of response to scopolamine in depression. Transl Psychiatry. 2018 14;8(1):280.

41. Kogelman LJA, Esserlind A-L, Francke Christensen A, Awasthi S, Ripke S, Ingason A, et al. Migraine polygenic risk score associates with efficacy of migraine-specific drugs. Neurol Genet. 2019 Dec;5(6):e364.

42. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018 Sep;50(9):1219–24.

43. Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. Endocrine Reviews. 2019 Dec 1;40(6):1500–20.

44. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019 Apr;177(3):587-596.e9.

45. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. The American Journal of Human Genetics. 2019 Jan;104(1):21–34.

46. Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ. 2018 Jan 10;360:j5757.

47. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med [Internet]. 2019 Jan 15 [cited 2019 May 24]; Available from: http://www.nature.com/articles/s41436-018-0406-9

48. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet. 2015 Jun 6;385(9984):2264–71.

49. Smith EN, Koller DL, Panganiban C, Szelinger S, Zhang P, Badner JA, et al. Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. PLoS Genet. 2011 Jun;7(6):e1002134.

50. Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, et al. Genome-wide association study of recurrent early-onset major depressive disorder. Mol Psychiatry. 2011 Feb;16(2):193–201.

51. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry. 1994 Nov;51(11):849–59; discussion 863-864.

52. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci. 2015 Dec;4(1):7.

53. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. The American Journal of Human Genetics. 2015 Oct;97(4):576–92.

54. Bauer M, Andreassen OA, Geddes JR, Vedel Kessing L, Lewitzka U, Schulze TG, et al. Areas of uncertainties and unmet needs in bipolar disorders: clinical and research perspectives. The Lancet Psychiatry. 2018 Nov;5(11):930–9.

55. Sachs GS, Nierenberg AA, Calabrese JR, Marangell LB, Wisniewski SR, Gyulai L, et al. Effectiveness of Adjunctive Antidepressant Treatment for Bipolar Depression. N Engl J Med. 2007 Apr 26;356(17):1711–22.

56. Shen H, Zhang L, Xu C, Zhu J, Chen M, Fang Y. Analysis of Misdiagnosis of Bipolar Disorder in An Outpatient Setting. 2018;30(2):9.

57. Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting Polygenic Risk of Psychiatric Disorders. Biological Psychiatry. 2018 Dec;S000632231832119X.

58. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Dec;10(1):1776.

59. for the MedSeq Project team, Roberts JS, Robinson JO, Diamond PM, Bharadwaj A, Christensen KD, et al. Patient understanding of, satisfaction with, and perceived utility of whole-genome sequencing: findings from the MedSeq Project. Genet Med. 2018 Sep;20(9):1069–76.

60. Vassy JL, Lautenbach DM, McLaughlin HM, Kong S, Christensen KD, Krier J, et al. The MedSeq Project: a randomized trial of integrating whole genome sequencing into clinical medicine. Trials. 2014;15(1):85.

61. Smith DJ, Craddock N. Unipolar and bipolar depression: different or the same? The British Journal of Psychiatry. 2011 Oct;199(4):272–4.

62. Wells JE, Horwood LJ. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol Med. 2004 Aug;34(6):1001–11.

63. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics. 2019 Kwiecie;51(4):584–91.

Acknowledgements and Disclosures

The Bipolar Samples were genotyped under the direction of Bipolar Genome Study. The Principal Investigators and Co-Investigators were: John R Kelsoe, Tiffany A Greenwood, Caroline M Nievergelt, Rebecca McKinney, Paul D Shilling, Nicholas Schork, Erin N Smith, Cinnamon Bloss, John Nurnberger, Howard J Edenberg, Tatiana Foroud, Daniel L Koller, Elliot Gershon, Chunyu Liu, Judith A Badner, William A Scheftner, William B Lawson, Evaristus A Nwulia, Maria Hipolito, William Coryell, John Rice, William Byerley, Francis McMahon, Thomas G Schulze, David T Chen, Wade Berrettini, James B Potash, Peter P Zandi, Pamela B Mahon, Melvin G McInnis, Sebastian Zöllner, Peng Zhang, David W Craig, Szabolcs Szelinger, Thomas B Barrett. We thank Mark Chaffin for assistance in the calculation of polygenic risk scores. I would like to thank Dr. Fernando Goes for his mentorship during this project, as well as Andrea Ganna for his assistance in planning the analysis. Additionally, I would like to thank Dr. Jordan Smoller for agreeing to serve as my honors thesis mentor, and for valuable conversations in planning future work. We are grateful for the participation of all the subjects contributing to this research.

The text has been adapted from:

Liebers DT, Pirooznia M, Seiffudin F, Musliner KL, Zandi PP, Goes FS. Polygenic Risk of Schizophrenia and Cognition in a Population-Based Survey of Older Adults. Schizophr Bull. 2016;42(4):984–91.

Liebers DT, Pirooznia M, Ganna A, Bipolar Genome Study (BiGS), Goes FS (2020). Discriminating bipolar depression from major depressive disorder with polygenic risk scores. Psychological Medicine 1–8. https:// doi.org/10.1017/S003329172000015X

Leonpacher AK, Liebers D, Pirooznia M, Jancic D, MacKinnon DF, Mondimore FM, et al. Distinguishing bipolar from unipolar depression: the importance of clinical symptoms and illness features. Psychol Med. 2015 Aug;45(11):2437–46.

Attached Papers

Liebers DT, Pirooznia M, Ganna A, Bipolar Genome Study (BiGS), Goes FS (2020). Discriminating bipolar depression from major depressive disorder with polygenic risk scores. Psychological Medicine 1–8. https:// doi.org/10.1017/S003329172000015X (In Press).

Liebers DT, Pirooznia M, Seiffudin F, Musliner KL, Zandi PP, Goes FS. Polygenic Risk of Schizophrenia and Cognition in a Population-Based Survey of Older Adults. Schizophr Bull. 2016;42(4):984–91.

Leonpacher AK, Liebers D, Pirooznia M, Jancic D, MacKinnon DF, Mondimore FM, et al. Distinguishing bipolar from unipolar depression: the importance of clinical symptoms and illness features. Psychol Med. 2015 Aug;45(11):2437–46.

Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Liebers DT, Pirooznia M, Ganna A, Bipolar Genome Study (BiGS), Goes FS (2020). Discriminating bipolar depression from major depressive disorder with polygenic risk scores. *Psychological Medicine* 1–8. https://doi.org/10.1017/S003329172000015X

Received: 23 September 2019
Revised: 2 January 2020
Accepted: 14 January 2020

Key words:
Bipolar disorder; genetics; GWAS; major depressive disorder; mood disorders; polygenic risk scores

Author for correspondence:
Fernando S. Goes, E-mail: fgoes1@jhmi.edu

CAMBRIDGE UNIVERSITY PRESS

Discriminating bipolar depression from major depressive disorder with polygenic risk scores

David T. Liebers[1], Mehdi Pirooznia[2], Andrea Ganna[3,4,5,6], Bipolar Genome Study (BiGS) and Fernando S. Goes[2]

[1]Harvard Medical School, Boston, MA 02115, USA; [2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins Institute of Medicine, Baltimore, MD 21205, USA; [3]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA and [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Abstract

Background. Although accurate differentiation between bipolar disorder (BD) and unipolar major depressive disorder (MDD) has important prognostic and therapeutic implications, the distinction is often challenging based on clinical grounds alone. In this study, we tested whether psychiatric polygenic risk scores (PRSs) improve clinically based classification models of BD *v.* MDD diagnosis.

Methods. Our sample included 843 BD and 930 MDD subjects similarly genotyped and phenotyped using the same standardized interview. We performed multivariate modeling and receiver operating characteristic analysis, testing the incremental effect of PRSs on a baseline model with clinical symptoms and features known to associate with BD compared with MDD status.

Results. We found a strong association between a BD diagnosis and PRSs drawn from BD ($R^2 = 3.5\%$, $p = 4.94 \times 10^{-12}$) and schizophrenia ($R^2 = 3.2\%$, $p = 5.71 \times 10^{-11}$) genome-wide association meta-analyses. Individuals with top decile BD PRS had a significantly increased risk for BD *v.* MDD compared with those in the lowest decile (odds ratio 3.39, confidence interval 2.19–5.25). PRSs discriminated BD *v.* MDD to a degree comparable with many individual symptoms and clinical features previously shown to associate with BD. When compared with the full composite model with all symptoms and clinical features PRSs provided modestly improved discriminatory ability ($\Delta C = 0.011$, $p = 6.48 \times 10^{-4}$).

Conclusions. Our study demonstrates that psychiatric PRSs provide modest independent discrimination between BD and MDD cases, suggesting that PRS could ultimately have utility in subjects at the extremes of the distribution and/or subjects for whom clinical symptoms are poorly measured or yet to manifest.

Introduction

Recurrent depressive syndromes, whether diagnostically linked to unipolar major depressive disorder (MDD) or to bipolar disorder (BD), affect as much as a fifth of the world's population and are associated with substantial morbidity, increased risk for suicide, and premature mortality (Bauer et al., 2018; Kessler, 2012). Although the distinction between BD and MDD (Leonhard, Korff, & Schulz, 1962) has been validated by the discovery of pharmacologic agents with relative specificity for depressive and manic syndromes (Altshuler et al., 1995; Pacchiarotti et al., 2013), the greater frequency of depressive symptoms in BD often leads to prolonged periods of misdiagnosis and ineffective treatment (Baldessarini et al., 2013; Lish, Dime-Meenan, Whybrow, Price, & Hirschfeld, 1994). While subjects with a clear history of mania can be easily distinguished from those with MDD, many subjects with BD experience milder or atypical forms of (hypo)mania that may be difficult to recall or characterize, particularly within the context of routine clinical care (Das et al., 2005). Moreover, as the early manifestation of BD is often marked by depressive symptoms (Van Meter, Burke, Youngstrom, Faedda, & Correll, 2016), accurate diagnoses can be challenging to make on clinical grounds alone.

Clinical features such as psychotic symptoms, psychomotor retardation, greater overall severity, and earlier onset have been found to be more prevalent in subjects with bipolar depression (Frankland et al., 2018; Gan et al., 2011; Leonpacher et al., 2015; Mitchell et al., 2011). Nevertheless, these prevalence differences are generally modest and likely insufficient for diagnostic purposes. Combining symptoms with illness features into more comprehensive models of disease course can provide greater discrimination, with area under the curve (AUC) indexes ranging from 0.7 to 0.8 (Gan et al., 2011; Leonpacher et al., 2015; Schaffer et al., 2015).

However, clinical features can be imprecise, poorly recalled, and often only provide prognostic utility after their manifestation, thereby limiting their relevance in prodromal and early onset cases, where the need for better diagnostic indices is arguably greatest (Van Meter et al., 2016).

The strong genetic basis of most psychiatric disorders can, theoretically, provide diagnostic information, particularly in scenarios of clinical uncertainty. Yet, as the genetic basis of complex disorders such as BD and MDD is overwhelmingly polygenic (Stahl et al., 2019; Wray, Goddard, & Visscher, 2007), clinically relevant stratification is most likely when considering genetic markers in aggregate in the form of polygenic risk scores (PRSs) (Purcell et al., 2009; Wray et al., 2007). The extent of the genetic predictive potential is bounded by the overall heritability attributable to additive effects of common variation ['genome-wide association study (GWAS) heritability'], ranging from about 10% for MDD to 20–30% for more heritable disorders such as BD or schizophrenia (SCZ) (Howard et al., 2018; Pardiñas et al., 2018; Stahl et al., 2019). However, the proportion of variance explained (R^2) by currently available PRSs is more modest, ranging from ~12% for SCZ, to ~8% for BD and ~2–3% for MDD (Howard et al., 2018; Pardiñas et al., 2018; Stahl et al., 2019). Although the predictive ability of GWAS-derived PRS is expected to increase with larger meta-analytic sample sizes, available PRSs are nevertheless beginning to yield association with clinically relevant outcomes. Recent examples include the association of increased SCZ PRSs with a decreased response to antipsychotics in first-episode psychosis (Zhang et al., 2019), and a decreased response to lithium in BD (Amare et al., 2017).

Potentially, the most successful application of PRS thus far has occurred in cardiovascular disorders, where population-based samples have provided strong evidence that PRS can identify that proportion of the population at high risk for cardiovascular morbidity who are likely to benefit from more intensive treatment strategies and/or implementation of risk mitigation strategies (Khera et al., 2017, 2018, 2019).

In this study, we have explored the degree to which PRSs may aid in distinguishing bipolar from unipolar depression, arguably one of the most important clinical decisions in psychiatry, since the prognosis and optimal treatment of either disorder is likely to be different for most patients. We specifically selected samples that were identically genotyped and phenotyped, thereby allowing us to test the effect of PRSs individually, but also within the broader context of clinical risk factors typically used to help with clinical decision making.

Materials and methods

Subjects

The sample included BD and MDD cases from the National Institute of Mental Health (NIMH) Bipolar Disorders Initiative and Genetics of Recurrent Early-Onset Depression (GenRED) collections, respectively, which were initially ascertained for genetic studies and have undergone previous case–control GWAS analyses (Shi et al., 2011; Smith et al., 2011). These samples were originally collected under Institutional Review Board (IRB) approved protocols (Shi et al., 2011; Smith et al., 2011) that included informed consent for genetic studies and data sharing.

Both studies used the Diagnostic Interview for Genetic Studies (DIGS) and similar best estimate procedures (Nurnberger et al., 1994) based on DSM-IV criteria. Subjects were restricted to high-confidence diagnoses of BD or MDD, with diagnostic confidence levels of 3 ('confident – most but not all diagnostic criteria have been met for a typical case') or 4 ('high – typical presentation of a clear case that meets all diagnostic criteria'). We subsequently selected unrelated subjects of European ancestry who were genotyped on the Affymetrix 6.0 Microarray and independent of the initial training set GWAS meta-analyses samples. Following these criteria, there were 843 subjects broadly diagnosed as BD (808 with BD type I and 35 with schizo-affective disorder, bipolar type) and 930 with recurrent MDD. Further clinical information is provided in Table 1.

Table 1. Clinical symptoms and features across the BD and MDD samples ($n = 1773$)

	BD ($n = 843$)	MDD ($n = 930$)
Demographics		
Gender, *n* (%)		
Female	556 (66.0)	651 (70.0)
Age at interview, mean (s.d.)	42.0 (12.3)	40.2 (11.8)
Years education, mean (s.d.)	14.9 (2.7)	16.0 (2.8)
Married, *n* (%)	250 (29.8)	352 (47.8)
Age at onset	17.8 (8.8)	16.9 (5.3)
Clinical symptoms		
Number of mixed symptoms, mean (s.d.)	1.3 (2.3)	0.5 (1.0)
Psychomotor retardation, number (%)	492 (58.4)	304 (32.7)
Incapacitation, number (%)	672 (79.7)	382 (41.1)
Delusions, number (%)	194 (23.0)	27 (3.0)
Number of mixed symptoms, mean (s.d.)	1.3 (2.3)	0.5 (1.0)
Clinical features		
Duration in weeks, mean (s.d.)	35.8 (93.7)	81.9 (179.0)
Number of lifetime episodes, mean (s.d.)	22.1 (51.2)	8.0 (14.0)
High after antidepressant, number (%)	357 (42.3)	105 (11.3)

Clinical symptoms/features

In our previous study, we tested a large number of symptoms (directly inquired from the research participant) and clinical features (defined as descriptors of the available illness course) and prioritized seven symptoms/features that showed replicated evidence for association with BD *v.* MDD and that remained significant in a multivariate model (Leonpacher et al., Table 4). Building on our prior study, we have restricted our test to these four previously associated symptoms (psychomotor retardation, incapacitation, delusions, and number of mixed symptoms during worst depressive episode) and three associated clinical features (length of most severe episode, number of depressive episodes, and antidepressant-induced 'high' feeling).

We included all available subjects, including 984 (~55% of the current sample) subjects who had been previously used to select the discriminating clinical symptoms/features. Although the overlapping subjects theoretically could lead to concerns of overfitting, we note that the clinical symptoms/features were replicated in our prior studied (Leonpacher et al.) and that further sensitivity

Table 2. ORs of BD *v.* MDD status for clinical features and polygenic risk scores

Variable	OR	CI_{Lower}	CI_{Upper}	*z*-score	*p* value
Clinical symptoms					
Psychomotor retardation	2.45	1.92	3.13	7.20	5.99×10^{-13}
Incapacitation	3.41	2.65	4.40	9.49	$<2 \times 10^{-16}$
Delusions	5.71	3.58	9.43	7.10	1.29×10^{-12}
Mixed symptoms	1.26	1.16	1.37	5.53	3.18×10^{-8}
Clinical features					
Duration (weeks)	0.99	0.99	1.00	−5.12	3.06×10^{-7}
Number of episodes	1.02	1.01	1.03	5.56	2.66×10^{-8}
High after AD Rx	3.70	2.76	4.97	8.76	$<2 \times 10^{-16}$
Polygenic risk scores (PRSs)					
PRS_{BD}	1.39	1.19	1.61	4.32	1.57×10^{-5}
PRS_{SCZ}	1.46	1.22	1.73	4.21	2.57×10^{-5}

OR, odds ratio; CI_{Lower}, lower bound 95% confidence interval; CI_{Upper}, upper bound 95% confidence interval.
Multivariate logistic regression with symptoms, clinical features, and standardized PRSs as predictors of BD *v.* MDD status. The model is adjusted for age, sex, and first 10 PCs from the GWAS.

analyses in our current sample showed no differences in discriminatory ability between the overlapping and non-overlapping samples. The final sample included a total of 1773 subjects, consisting of 843 with BD and 930 with recurrent MDD.

Genome-wide association study (GWAS)

Cases with BD and MDD were genotyped using the Affymetrix 6.0 micro-array. Raw genotypes were combined using PLINK1.9 (Chang et al., 2015) followed by sequential quality control (QC) steps removing: (1) single-nucleotide polymorphisms (SNPs) with missing rates ⩾5%; (2) individuals with overall missing rates ⩾2%; (3) remaining SNPs with missing rates ⩾2%; (4) SNPs in Hardy–Weinberg disequilibrium ($p < 0.000001$); and (5) remaining SNPS with differential missing rates >2%. The filtered combined file underwent additional QC/harmonization steps (https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.2.7.zip) prior to imputation to the multiethnic 1000Genomes version 3 reference panel. Imputation was performed using the Michigan Imputation Server and high-quality genotypes ($r^2 > 0.6$) with a minor allele frequency of 1% or greater were extracted and converted to the PLINK format for downstream analyses. Principal components analysis (PCA) was performed using PLINK1.9 by combining the filtered genotypes with the 1000G multi-ethnic panel and extracting the top 10 principal components (PCs). We excluded five samples that were outliers in the PCA (online Supplementary Fig. S1).

Calculation of PRSs

PRSs were calculated using recommended parameters in LD-Pred (Vilhjálmsson et al., 2015), which uses a Bayesian algorithm to infer the mean effect at each individual marker while accounting for linkage disequilibrium (LD) between markers. We utilized summary statistics from the largest meta-analysis GWAS of BD (Stahl et al., 2019) and SCZ (Pardiñas et al., 2018) as the training datasets and the European subjects from the 1000Genomes version 3 as the LD reference genotypes. Using an LD-radius of 1775 and 1981, weighted risk scores were constructed for 5 325 407 and 5 943 410 markers in the harmonized BD and SCZ datasets, respectively. PRSs were derived across nine training dataset thresholds, defined as the assumed 'fraction of causal variants' (–PS = 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0001), and standardized by using the standard deviation as the unit of measurement. PRS thresholds were selected to maximize the explained difference between BD and MDD phenotypes. The optimal LD-Pred threshold ('fraction of causal variants') was 0.3 for SCZ and 0.003 for BD; however, other thresholds yielded broadly similar discriminatory ability (online Supplementary Table S2).

Statistical analysis

All analyses were performed in R version 3.4.2. Association of PRSs with phenotype status was performed with a generalized linear model using a logit link function. We initially fit baseline models with BD *v.* MDD status as the outcome variable and the first 10 PCs, age at interview, and sex as covariates. We subsequently added the PRS variable into that base model and compared the difference in Nagelkerke's pseudo-R^2 (ΔR^2) to estimate the proportion of variance attributable to the PRS. For better interpretability, the sample was stratified into deciles using the residuals of the top performing BD and SCZ PRSs adjusted for first 10 PCs, age, and sex. Odds ratios (ORs) for BD *v.* MDD status and 95% confidence intervals (CIs) were calculated for each decile in comparison with the first decile (Table 2).

We performed a receiver operating characteristic analysis to estimate the discriminatory potential of PRS by measuring the AUC or *C*-statistic associated with each model. We initially fit a logistic model with the first 10 PCs, age, and sex, and then measured the *C*-index. To this baseline model, we added PRSs derived from the BD and SCZ PRSs using the optimal LD-Pred threshold described above. We subsequently tested the increase in *C*-statistic from each of the four individual symptoms (psychomotor retardation, incapacitation, delusions, and number of mixed symptoms) and three clinical features (length of most severe episode, number of depressive episodes, and the presence of high feeling after antidepressant). The *C*-statistic was initially calculated separately for each individual symptom or clinical feature, and then with the

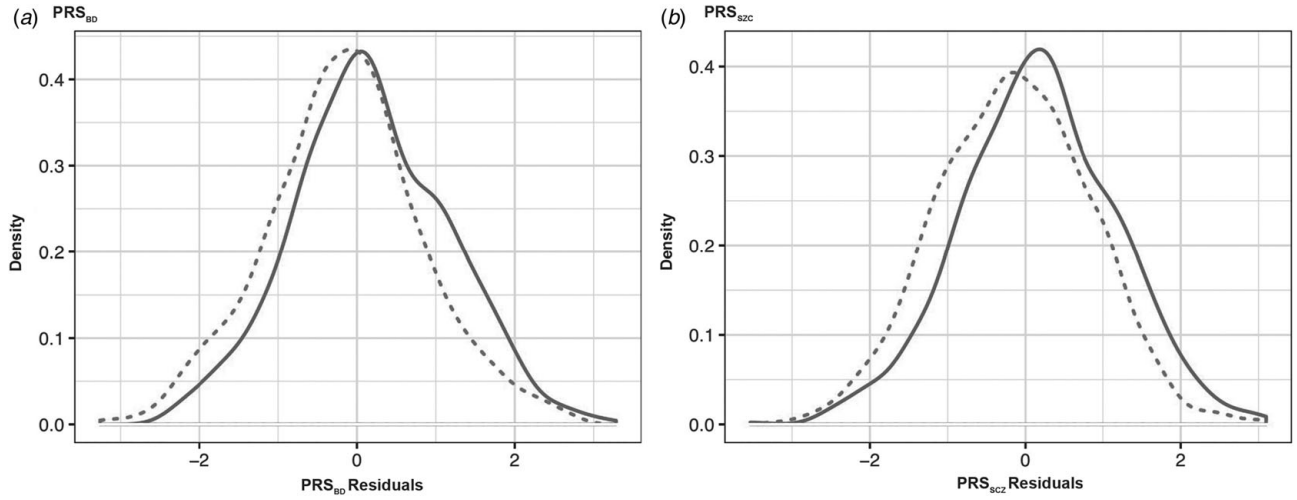


Fig. 1. Distribution of PRSs derived from (*a*) BD and (*b*) SCZ training datasets. Distribution is shown using dashed lines and solid lines for MDD and BD cases, respectively.

BD PRSs to measure the respective additive effect of PRS. Subsequently, we performed the above analyses serially for: all clinical symptoms combined, all clinical features combined, and a composite of all symptoms and clinical features.

Lastly, we tested the association of the top performing BD and SZ PRS with each of the symptoms/clinical features listed above, controlling for diagnosis. Logistic regression was used for binary variables (the presence of psychomotor retardation, delusions, incapacitation, and high feeling after antidepressant) and linear regression for continuous variables (length of worst depression in weeks, number of episodes, and number of mixed symptoms).

Results

Clinical and demographic features

Of the 1773 subjects, 843 had BD and 930 had MDD, with more females than males in both groups (66.0% and 70.0%, respectively) (Table 1). Subjects with BD were slightly older at the time of their interview (42.0 years BD *v.* 40.2 years MDD), had fewer years of education (14.9 BD *v.* 16.0 MDD), and were less likely to be married (29.8% BD *v.* 47.8% MDD). Consistent with our previous study (Leonpacher et al., 2015), subjects with BD were more likely to have experienced psychomotor retardation, incapacitation, delusions, and mixed symptoms during the their more severe depressive episodes. In terms of clinical features, subjects with BD has a shorter length of depression during their most severe episode, but a greater number of depressive episodes, and a higher prevalence of feeling 'high after antidepressant treatment' (Table 1).

Polygenic association with BD v. *MDD*

Given the strong co-heritability between BD and SCZ (The Consortium et al., 2018), we performed polygenic analyses utilizing the largest meta-analyses of each disorder for polygenic risk scoring, while ensuring that the BD and MDD samples were independent of the training set samples. In our primary analysis, we found that PRS for BD was strongly associated with BD compared with MDD ($\Delta R^2 = 3.53\%$, $p = 4.94 \times 10^{-12}$), with an effect size approximately half of that usually seen in BD case–control studies. Similarly, PRS for SCZ was also similarly associated with BD *v.* MDD phenotype ($\Delta R^2 = 3.16\%$, $p = 5.71 \times 10^{-11}$). While the distribution of PRS among BD and MDD samples showed prominent overlap, greater discriminative ability was present at the more extreme values of PRS, as has been observed in other common complex disorders (Fig. 1).

Using PRSs to help classify the overall sample, we found that a baseline model with covariates and PRS produced an AUC for the BD and SCZ PRS, respectively and *C*-statistic of 0.65 for the combined PRS (online Supplementary Table S1). For ease of interpretability, we divided the samples into PRS deciles, finding that individuals with PRS in the top decile had an OR of 3.39 (CI 2.19–5.25) when compared with the lowest decile. Similarly, those in the top decile of SCZ PRS residuals had an OR of 3.07 (CI 1.99–4.73) compared with the lowest decile, in terms of BD–MDD status (Fig. 2).

Genetic risk and clinical features

In a multivariate model adjusted for covariates, we found significant associations the BD and SCZ PRSs as well as of the seven clinical symptoms/features tested (Table 1). As shown in Fig. 3, we found that the combined PRS contributed significant additive effects to all individual clinical features, providing evidence that prediction based on PRSs is likely to be independent of that based on clinical features (data shown in online Supplementary Table S2). The classifying ability of PRSs in models with individual clinically based predictors varied based on the specific variable. For example, the combined PRSs provided a larger increase in discriminatory ability in a model with delusions alone ($\Delta C = 0.042$) than incapacitation alone ($\Delta C = 0.021$).

Polygenic risk was also found to increase discrimination when compared with the combined symptoms and combined clinical features categories. As the combined symptoms category represents a set of symptoms that could be plausibly ascertained on initial presentation, these results suggest that PRS could provide additional predictive ability early in the illness course before the longitudinal clinical features have had time to become establish themselves. When both symptoms and clinical features were combined, the AUC increase to 0.855, which was consistent with our

Fig. 2 - B/W online, B/W in print

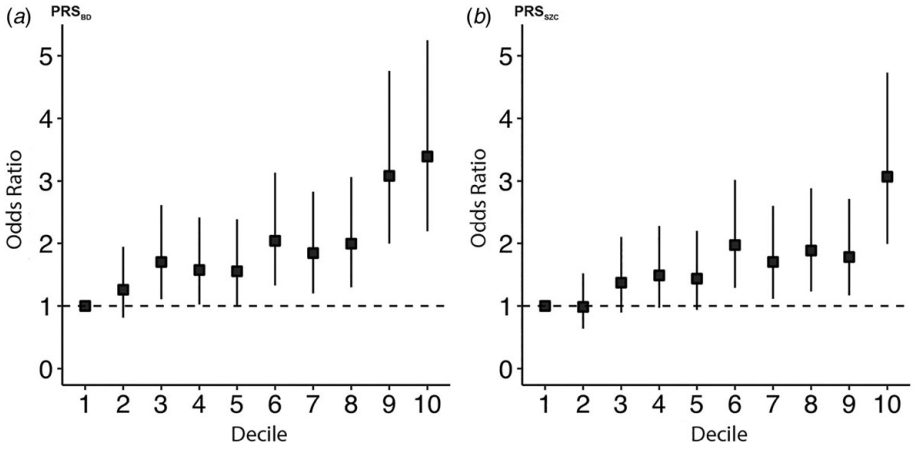


Fig. 2. ORs for association with BD compared with MDD binned by polygenic risk deciles using residuals after adjusting for first 10 PCs, age, and sex. ORs represent comparisons against the lowest decile.

prior study (Leonpacher et al., 2015). However, even in the full multi-variate clinical model, the effect of the combined PRS remained significant and led to modest but statistically significant increases in discriminatory ability ($\Delta C = 0.011$, $p = 6.48 \times 10^{-4}$).

Association of PRS with symptoms and clinical features

Finally, in order to explore the specific effect of polygenic risk, we sought to test whether the polygenic signal was differentially associated with individual clinical symptoms or features, thereby providing a more granular view of the underlying genetic susceptibility. Using a logistic regression model adjusted for covariates that included the main effect of diagnosis, we found a single significant association with BD PRS and incapacitation (Table 3). In contrast, the SCZ PRS was associated with longer length of depression ($p = 0.0043$), but not with any of the additional risk factors, suggesting that the BD and SCZ PRS have heterogeneous effects despite comparable overall predictive effects.

Discussion

While mania remains the cardinal syndrome that defines the diagnosis of BD, subjects often initially present with years of depressive symptoms (Bauer et al., 2018) that can result in misdiagnosis on initial presentation, potentially leading to the initiation of treatments of limited efficacy (Sachs et al., 2007) that may be associated with iatrogenic harm (Baldessarini et al., 2013). Even when present, relevant clinical symptoms are under-recognized or poorly measured, particularly in primary care settings (Das et al., 2005). Hence, there is a clear need for better prognostic tools that can help identify subjects at high risk for developing BD.

In this study, we have tested whether markers of genetic susceptibility for BD or SCZ (as indexed by PRSs) can augment traditionally used clinically based factors, in helping to identify subjects presenting with depressive symptoms who may be at high risk for 'converting' to BD. We found that PRSs derived from the most recent PGC GWAS discriminate modestly between BD and MDD cases, with BD PRS demonstrating a slightly higher association than SCZ PRS. Subjects with top decile BD PRS were more than three times as likely to have BD than MDD, compared with those in the lowest BD PRS decile, a level of discrimination that is similar to that found in a recent study of PRS in first-episode psychosis (Vassos et al., 2017). In addition, PRSs had comparable classifying ability to individual clinical features and/or symptoms previously shown to associate with BD, although when included in a model with all clinical features and symptoms, the increase in overall classification was modest. Nevertheless, the effect of PRSs appeared to be largely independent of the clinical features, suggesting that PRSs may represent orthogonal risk factors that may provide incremental information beyond traditional clinical factors.

In other medical disorders, PRSs are increasingly being studied for their potential use in the prediction of common disorders such as heart disease (Khera et al., 2018), type-II diabetes (Udler, McCarthy, Florez, & Mahajan, 2019), obesity (Khera et al., 2019), and cancers such as breast (Mavaddat et al., 2019) and prostate cancer (Seibert et al., 2018). In such examples, clinical utility is most likely to be found in the extremes of the PRS distribution, where the degree of increased risk approximates those found in our study. PRS for these disorders has also been found to be largely independent of clinically based risk estimation tools, including family history, suggesting that PRS could have an independent and incremental role in comprehensive models to inform decision making on prevention therapies and screening (Lee et al., 2019). Indeed, in some disorders such as coronary artery disease, PRS has not only been associated with increased risk, but also with a greater likelihood of benefit from pharmacological treatment (Mega et al., 2015). However, for most studied disorders, including BD, the link to specific clinical utility is currently unclear and will require further study based on prospective cohort studies and randomized trials. Ultimately, the most informative risk models are likely to integrate epidemiologic parameters, clinical data, non-genetic biomarkers, and potentially a broader index of genetic risk, summing the effects both common and rare variants.

While genetic data have important limitations, including the constraints placed by heritability estimates, it may also hold advantages over more traditional clinically derived predictors. Unlike most clinical predictors, genetic predictors can be measured before the onset of symptoms and can be informative throughout the life-course of the illness. In contrast, several of the clinical features included in our model, such as number of lifetime episodes, or the duration of depressive episode represents information that would not be available during a first episode or prior to the full manifestation of the disorder. Moreover, clinical symptoms are often experienced in a time-limited manner, leading to poor recall and suboptimal inter-rater reliability (Regier et al., 2013).

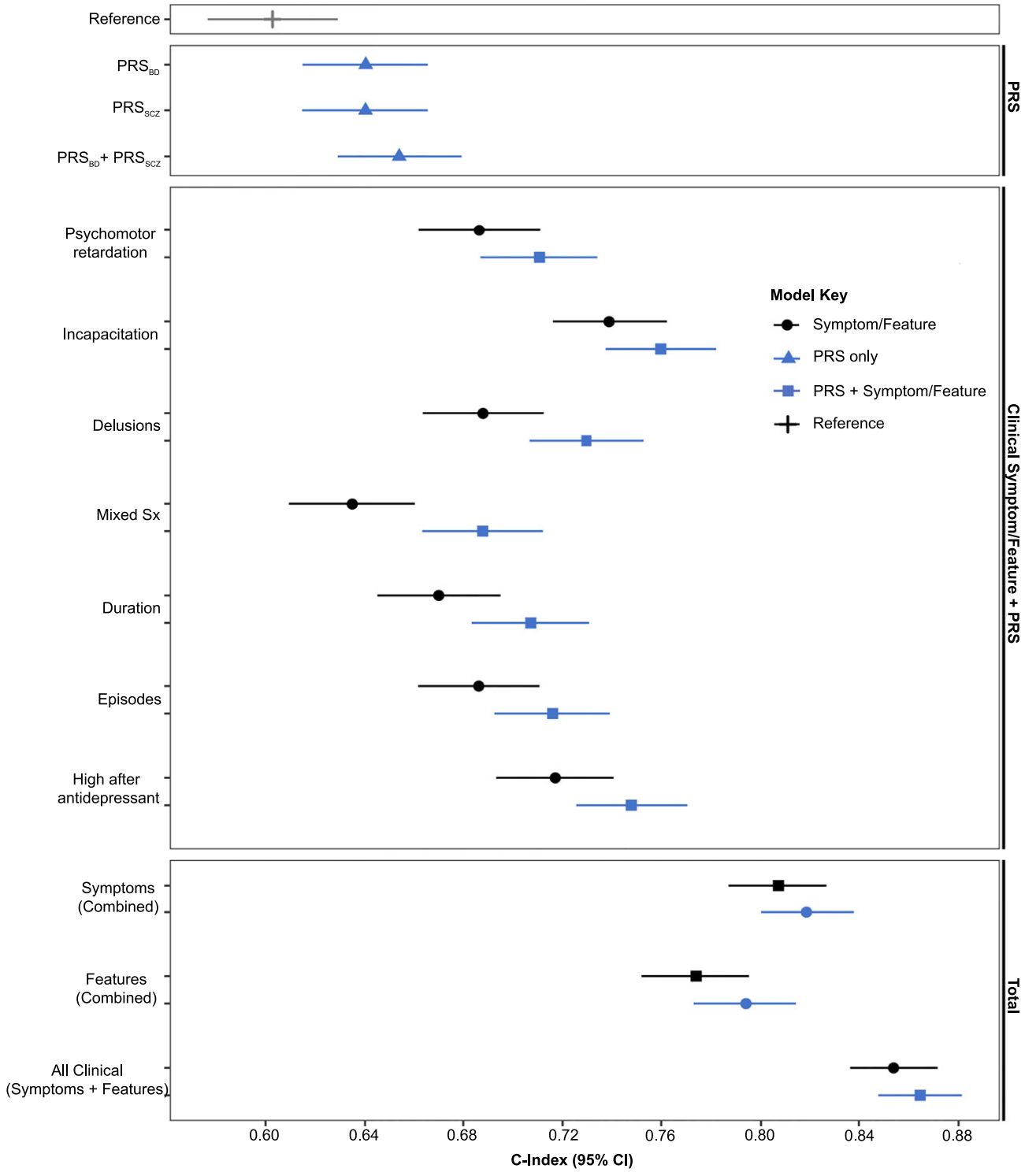


Fig. 3 - Colour online, Colour in print

Fig. 3. C-indexes (AUC) for classification of BD *v.* MDD. The figure shows the incremental effects of including individual symptoms or clinical features with and without the additional influence of PRSs. The reference models represent the effect of baseline covariates (age, sex, and 10 PCs). Additional clinical predictors are categorized as depressive symptoms [psychomotor retardation (PMR), incapacitation, delusions, and number of mixed symptoms during worst depressive episode] or associated clinical features (length of most severe episode, number of depressive episodes, and antidepressant-induced 'high' feeling).

An important strength of our study is that rather than comparing cases and healthy controls, we focus on the more clinically relevant comparison between two disorders with prominent symptom overlap that nevertheless require relatively distinct forms of pharmacotherapy. As such, we therefore restricted our analyses to subjects with either disorder who were diagnosed with identical diagnostic instruments and genotyped with the same microarray chip, in order to minimize the potential for confounding. Additionally, our current analysis focused on the arguably more realistic assumption that PRSs will be used in

Table 3. Polygenic association with individual symptoms and clinical features

PRS	SCZ			BD		
	β estimate	S.E.	p value	β estimate	S.E.	p value
Psychomotor retardation	0.003	0.061	0.958	0.068	0.067	0.309
Incapacitation	0.003	0.061	0.958	0.214	0.072	0.003[a]
Delusions	0.037	0.087	0.672	−0.257	0.104	0.013
Number of mixed symptoms	−0.008	−0.156	0.876	−0.118	0.056	0.035
Duration (weeks)	11.264	3.942	0.004[a]	2.044	4.607	0.657
Number of episodes	−0.505	1.001	0.614	−1.122	1.167	0.336
High after antidepressant	−0.056	0.067	0.404	−0.040	0.079	0.610

Results show the association between individual clinical symptoms or features and PRSs derived for BD and SCZ. Binary variables were tested with logistic regression models, whereas continuous variables were tested using linear models. All models are adjusted for diagnosis, the first 10 PCs, age, and sex.
[a]Significant after Bonferroni correction ($p < 0.007$) for multiple tests.

conjunction with traditional clinical risk factors, rather than studied or applied in isolation. We therefore applied polygenic risk scoring to previously identified clinical factors that had shown replicable associations in a prior comparison of BD and MDD (Leonpacher et al., 2015). Our results therefore may have greater face validity while providing a more contextual view of where PRS may have clinical utility in future studies.

As an initial proof of concept study, there are also several important limitations. We have analyzed opportunistically samples previously collected for genetic studies that focused on the ascertainment of the more severe types of BD and on recurrent forms of MDD. Further work should test the performance of psychiatric PRS in discriminating MDD and BD cases in more naturalistic clinical cohorts that include the full manifestation of the BD *v.* MDD mood spectrum, particularly in the early phases of illness (Smith & Craddock, 2011). Second, our clinical features were derived from a single cross-sectional evaluation that is subject to typical retrospective limitations, including recall bias (Wells & Horwood, 2004). Although best-estimate procedures include additional diagnostic information to supplement the interview, it is possible that certain symptoms may have been under-reported. Third, we limited our focus to clinical factors available in the diagnostic instrument used in our previous study (Leonpacher et al., 2015), whereas a more comprehensive diagnostic model may ultimately include additional risk and state-related factors more reliably assessed using a prospective cohort. Fourth, in the current study we have only tested PRSs derived from BD and SCZ studies, reasoning that other prognostically relevant phenotypes such as major depression or broadly defined anxiety remained underpowered for polygenic prediction. Future studies will no doubt be benefitted by including more powerful training data from diagnostic and cross-disorder phenotypes in order sets to help clarify how best to integrate genetic and clinical information in clinical decision-making algorithms. Lastly, our study was performed exclusively in samples of European ancestry with PRSs similarly derived in GWAS meta-analyses of subjects of European ancestry. These PRSs are known to be less predictive in non-European samples, reflecting a central limitation of integrating currently available PRS into clinical medicine, and one which has important consequences for health equity (Martin et al., 2019).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S003329172000015X

Acknowledgements. The Bipolar Samples were genotyped under the direction of Bipolar Genome Study. The Principal Investigators and Co-Investigators were: John R. Kelsoe, Tiffany A. Greenwood, Caroline M. Nievergelt, Rebecca McKinney, Paul D. Shilling, Nicholas Schork, Erin N. Smith, Cinnamon Bloss, John Nurnberger, Howard J. Edenberg, Tatiana Foroud, Daniel L. Koller, Elliot Gershon, Chunyu Liu, Judith A. Badner, William A. Scheftner, William B. Lawson, Evaristus A. Nwulia, Maria Hipolito, William Coryell, John Rice, William Byerley, Francis McMahon, Thomas G. Schulze, David T. Chen, Wade Berrettini, James B. Potash, Peter P. Zandi, Pamela B. Mahon, Melvin G. McInnis, Sebastian Zöllner, Peng Zhang, David W. Craig, Szabolcs Szelinger, and Thomas B. Barrett. We thank Mark Chaffin for assistance in the calculation of polygenic risk scores. We are grateful for the participation of all the subjects contributing to this research.

Financial support. This study was supported in part by Brain & Behavior Research Foundation YI23970 to FSG. FSG has received research support from Janssen Pharmaceuticals.

References

Altshuler, L. L., Post, R. M., Leverich, G. S., Mikalauskas, K., Rosoff, A., & Ackerman, L. (1995). Antidepressant-induced mania and cycle acceleration: A controversy revisited. *The American Journal of Psychiatry*, *152*, 1130–1138.

Baldessarini, R. J., Faedda, G. L., Offidani, E., Vázquez, G. H., Marangoni, C., Serra, G., & Tondo, L. (2013). Antidepressant-associated mood-switching and transition from unipolar major depression to bipolar disorder: A review. *Journal of Affective Disorders*, *148*, 129–135.

Bauer, M., Andreassen, O. A., Geddes, J. R., Vedel Kessing, L., Lewitzka, U., Schulze, T. G., & Vieta, E. (2018). Areas of uncertainties and unmet needs in bipolar disorders: Clinical and research perspectives. *The Lancet Psychiatry*, *5*, 930–939.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.

Das, A. K., Olfson, M., Gameroff, M. J., Pilowsky, D. J., Blanco, C., Feder, A., … Weissman, M. M. (2005). Screening for bipolar disorder in a primary care practice. *Journal of the American Medical Association*, *293*, 956–963.

Frankland, A., Roberts, G., Holmes-Preston, E., Perich, T., Levy, F., Lenroot, R., … Mitchell, P. B. (2018). Clinical predictors of conversion to bipolar disorder in a prospective longitudinal familial high-risk sample: Focus on depressive features. *Psychological Medicine*, *48*, 1713–1721.

Gan, Z., Diao, F., Wei, Q., Wu, X., Cheng, M., Guan, N., … Zhang, J. (2011). A predictive model for diagnosing bipolar disorder based on the clinical characteristics of major depressive episodes in Chinese population. *Journal of Affective Disorders*, *134*, 119–125.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., … McIntosh, A. M. (2018). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*, *22*, 343–352.

International Consortium on Lithium Genetics (ConLi+Gen), Amare, A. T., Schubert, K. O., Hou, L., Clark, S. R., Papiol, S., … Baune, B. T. (2018). Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: A genome-wide association study. *JAMA Psychiatry*, 2018, *75*, 65–74.

Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics of North America*, *35*, 1–14.

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., … Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*, 1219–1224.

Khera, A. V., Chaffin, M., Wade, K. H., Zahid, S., Brancale, J., Xia, R., … Kathiresan, S. (2019). Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell*, *177*, 587–596.e9.

Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., … Antoniou, A. C. (2019). BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine*, *21*, 1708–1718.

Leonhard, K., Korff, I., & Schulz, H. (1962). Temperament in families with monopolar and bipolar phasic psychoses. *Psychiatria et Neurologia*, *143*, 416–434.

Leonpacher, A. K., Liebers, D., Pirooznia, M., Jancic, D., MacKinnon, D. F., Mondimore, F. M., … Goes, F. S. (2015). Distinguishing bipolar from unipolar depression: The importance of clinical symptoms and illness features. *Psychological Medicine*, *45*, 2437–2446.

Lish, J. D., Dime-Meenan, S., Whybrow, P. C., Price, R. A., & Hirschfeld, R. M. A. (1994). The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders*, *31*, 281–294.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*, 584–591.

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., … Easton, D. F. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *The American Journal of Human Genetics*, *104*, 21–34.

Mega, J. L., Stitziel, N. O., Smith, J. G., Chasman, D. I., Caulfield, M., Devlin, J. J., … Sabatine, M. S. (2015). Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: An analysis of primary and secondary prevention trials. *Lancet*, *385*, 2264–2271.

Mitchell, P. B., Frankland, A., Hadzi-Pavlovic, D., Roberts, G., Corry, J., Wright, A., … Breakspear, M. (2011). Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. *British Journal of Psychiatry*, *199*, 303–309.

Nurnberger, J. I., Blehar, M. C., Kaufmann, C. A., York-Cooler, C., Simpson, S. G., Harkavy-Friedman, J., … Reich, T. (1994). Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Archives of General Psychiatry*, *51*, 849–859.

Pacchiarotti, I., Bond, D. J., Baldessarini, R. J., Nolen, W. A., Grunze, H., Licht, R. W., … Vieta, E. (2013). The International Society for Bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. *American Journal of Psychiatry*, *170*, 1249–1262.

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., … CRESTAR Consortium. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*, 381–389.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., … Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 748–752.

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, Part II: Test–retest reliability of selected categorical diagnoses. *The American Journal of Psychiatry*, *170*, 59–70.

Sachs, G. S., Nierenberg, A. A., Calabrese, J. R., Marangell, L. B., Wisniewski, S. R., Gyulai, L., … Thase, M. E. (2007). Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine*, *356*, 1711–1722.

Schaffer, A., Isometsä, E. T., Tondo, L., Moreno, D. H., Sinyor, M., Kessing, L. V., … Yatham, L. (2015). Epidemiology, neurobiology and pharmacological interventions related to suicide deaths and suicide attempts in bipolar disorder: Part I of a report of the International Society for Bipolar Disorders Task Force on Suicide in Bipolar Disorder. *Australian & New Zealand Journal of Psychiatry*, *49*, 785–802.

Seibert, T. M., Fan, C. C., Wang, Y., Zuber, V., Karunamuni, R., Parsons, J. K., … Practical Consortium. (2018). Polygenic hazard score to guide screening for aggressive prostate cancer: Development and validation in large scale cohorts. *BMJ*, *360*, j5757.

Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., … Levinson, D. F. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular Psychiatry*, *16*, 193–201.

Smith, D. J., & Craddock, N. (2011). Unipolar and bipolar depression: Different or the same? *The British Journal of Psychiatry*, *199*, 272–274.

Smith, E. N., Koller, D. L., Panganiban, C., Szelinger, S., Zhang, P., Badner, J. A., … Kelsoe, J. R. (2011). Genome-wide association of bipolar disorder suggests an enrichment of replicable associations in regions near genes. *PLoS Genetics*, *7*, e1002134.

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., … Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, *51*, 793–803.

Consortium, T. B., Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science*, *360*, eaap8757.

Udler, M. S., McCarthy, M. I., Florez, J. C., & Mahajan, A. (2019). Genetic risk scores for diabetes diagnosis and precision medicine. *Endocrine Reviews*, *40*, 1500–1520.

Van Meter, A. R., Burke, C., Youngstrom, E. A., Faedda, G. L., & Correll, C. U. (2016). The bipolar prodrome: Meta-analysis of symptom prevalence prior to initial or recurrent mood episodes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*, 543–555.

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., … Breen, G. (2017). An examination of polygenic score risk prediction in individuals with first-episode psychosis. *Biological Psychiatry*, *81*, 470–477.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., … Price, A. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*, *97*, 576–592.

Wells, J. E., & Horwood, L. J. (2004). How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. *Psychological Medicine*, *34*, 1001–1011.

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*, 1520–1528.

Zhang, J.-P., Robinson, D., Yu, J., Gallego, J., Fleischhacker, W. W., Kahn, R. S., … Lencz, T. (2019). Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *American Journal of Psychiatry*, *176*, 21–28.

Schizophrenia Bulletin vol. 42 no. 4 pp. 984–991, 2016
doi:10.1093/schbul/sbw001
Advance Access publication February 12, 2016

Polygenic Risk of Schizophrenia and Cognition in a Population-Based Survey of Older Adults

David T. Liebers[1], Mehdi Pirooznia[1], Fayaz Seiffudin[1], Katherine L. Musliner[2], Peter P. Zandi[1,2], and Fernando S. Goes*[,1]

[1]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD; [2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

*To whom correspondence should be addressed; Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Meyer 4-119A, 600 N. Wolfe Street, Baltimore, MD 21287, US; tel: 443-287-6382, fax: 410-502-0065, e-mail: fgoes1@jhmi.edu

Cognitive impairment is a common feature of the major psychotic disorders, with deficits often present in at risk individuals and unaffected first-degree relatives. Previous studies have suggested that polygenic risk scores (PRS) for schizophrenia (SCZ) are associated with cognitive deficits, but there has been little examination of this association in longitudinal datasets, or comparison with other disorders. We used mixed models to study the association between PRS for 4 adult onset psychiatric disorders with cross-sectional cognitive performance and longitudinal cognitive decline in 8616 older adults from the Health and Retirement Study (HRS), followed for an average of 10 years. PRS were computed for SCZ, bipolar disorder (BD), Major Depressive Disorder (MDD), and Alzheimer's disease (ALZ). SCZ PRS associated with decreased cognitive function ($z = -3.00$, $P = .001$, $\Delta R^2 = 0.04\%$), which was largely driven by an association with impaired attention and orientation ($z = -3.33$, $P = 4.3\times10^{-4}$, $\Delta R^2 = 0.08\%$). We found no effect of BD or MDD PRS on cognition, in contrast to a robust effect of the APOE4/TOMM40 locus ($z = -5.05$, $P = 2.2\times10^{-7}$, $\Delta R^2 = 0.36\%$), which was primarily associated with impaired verbal memory ($z = -5.15$, $P = 1.3\times10^{-7}$, $\Delta R^2 = 0.21\%$). APOE4/TOMM40 locus and the ALZ PRS, but not the PRS for SCZ, were associated with greater cognitive decline. In summary, using a large, representative sample of older adults, we found evidence for different degrees of association between polygenic risk for SCZ and genetic risk factors for ALZ on cognitive function and decline, highlighting potential differences in the pathophysiology of cognitive deficits seen in SCZ and ALZ.

Key words: schizophrenia/polygenic risk/cognition/attention/cross-disorder/endophenotype /Alzheimer's/APOE4/bipolar disorder/major depressive disorder

Introduction

Impaired cognitive function is a common feature of the major psychotic disorders, with deficits seen across many cognitive domains.[1] Cognitive impairments are often present prior to the onset of psychotic symptoms, and may manifest, in more attenuated forms, in unaffected relatives of subjects with psychotic disorders.[2–5] As such, cognitive deficits represent classical "endophenotypes,"[6] although the precise nature of the causal relationship between psychosis and impaired cognitive function is still unclear. While cognitive decline is often seen at the time of first diagnosis, it may be confounded by the psychotic symptoms, or by the adverse effects of treatment targeted toward those symptoms. Hence, although cognitive difficulties and psychotic syndromes often co-occur, their etiological relationship remains uncertain.

The major psychotic syndromes, schizophrenia (SCZ) and bipolar disorder (BD), are highly heritable disorders[7] whose genetic architecture is likely to be polygenic, with few, if any, loci of major effect.[8] Similarly, most neurocognitive phenotypes have also been found to be heritable and polygenic in nature.[9] For example, a recent, large scale (N = 53 949) study of general cognitive ability of adults in middle and older age estimated that up to 29% of the phenotypic variation was attributable to common variation, but found only 3 genome-wide significant loci, despite the large sample size.[10] Among the genome-wide significant findings was the *APOE4/TOMM40* locus, which is a strong risk factor for Alzheimer's Disease (ALZ) and has previously been associated with modest cognitive decline in normal aging.[11,12]

Given the frequent overlap between psychotic disorders and cognitive deficits, there is increasing interest in determining whether the relationship is potentially mediated through a genetic mechanism. Family and twin studies have provided evidence for a modest degree of phenotypic covariance between cognitive function and SCZ that appears primarily attributable to additive genetic effects.[13,14] The familial relationship between cognitive function and other psychiatric disorders, such as BD

and major depression, has been less studied, although a recent Danish Registry based study has suggested that a mild degree of impaired cognitive function may be associated with a family history of a broad range of psychiatric disorders.[15]

With the increasing availability of large genome-wide association study (GWAS) samples it is now possible to use polygenic modeling techniques to examine the extent to which genetic factors associated with a phenotype may influence another independent, or closely allied phenotype.[16] Polygenic scores indexing susceptibility for a specific phenotype can be constructed from a GWAS "training" dataset and tested for an association with a second phenotype in an independent GWAS datasets. The question of whether liability for SCZ, as measured by SCZ polygenic risk scores (PRS), is associated with cognitive function or cognitive decline was initially addressed by McIntosh et al,[17] who found a modest association between PRS constructed from the first Psychiatric GWAS Consortium (PGC-1) meta-analysis of SCZ[18] and cognitive decline between ages 11 and 70. The relationship between the PGC-1 SCZ risk scores and cognitive function was similarly tested by the Cognitive Genomics Consortium (COGENT),[19] which found similar evidence for a negative association (meta-analytic $P = 1.4 \times 10^{-4}$) between the PGC-1 SCZ PRS and overall cognitive function in a combined meta-analysis of 4302 nonpsychiatric control subjects.

Although these studies have provided initial evidence for an association of SCZ risk scores with deficits in cognitive function, they have been limited by relatively small sample sizes for GWAS based analyses, and by the limited consideration of important confounders such as education and depressive symptoms. Moreover, since heritability estimates of cognitive function may vary with age,[20,21] it may be of particular importance to test the genetic influences on cognitive function in older samples. Hence, we sought to revisit the question of whether SCZ PRS are associated with cognitive function or cognitive decline by making use of the more powerful training dataset from the recent PGC-2 meta-analysis of SCZ[22] and the considerable resources of the Health and Retirement Study (HRS), a longitudinal cohort with GWAS and repeated cognitive evaluations available for 8616 middle-aged and older subjects of non-Hispanic Caucasian ancestry.

Materials and Methods

HRS and Inclusion Criteria

The HRS is a longitudinal study investigating the effects of aging and retirement on a representative sample of older Americans.[23] From 1992 onwards, several cohorts have been followed as part of the HRS, with a "core" interview performed every 2 years using a mixed mode design (in-person and telephone interviews) to monitor work, health, social, psychological, family and economic status.[24] More than 26 000 individuals have been studied, with DNA obtained from saliva in 12 507 participants. For the present analysis we selected participants from Waves 3 (1996) through Wave 10 (2010) when the same full-scale "total cognition" measure was consistently administered to the HRS participants. From these waves, we included individuals with: (1) at least 1 cognition measure taken at age 50 or later, (2) self-reported non-Hispanic Caucasian ancestry and (2) available genetic data. A total of 8616 individuals, with 40 257 cognition measures met these criteria.

Cognition Measures

The cognition measure in the HRS ("total cognition") is a customized version of the Telephone Interview for Cognitive Status (TICS).[25] The interview contained 3 phases: (1) a test for immediate recall of 10 nouns, (2) a mental status exam, including a task specific to attention (serial 7s) and a set of tasks for orientation and language (object and president naming, date recitation), and (3) a delayed recall test of the same 10 nouns from the immediate recall test, but given after 5 minutes had elapsed. Total cognition is a summary variable with a range of 0–35 that includes verbal memory (immediate and delayed recall) and mental status (which includes tests for orientation and attention). Our primary analysis focused on the aggregate (total cognition) score; however, we also performed secondary analyses on the attention/orientation and verbal memory subcomponents of the cognition measure to further delineate the effect of PRS on cognition.

Prior studies of the these measures have found that the telephone-administered cognitive battery is highly consistent with face-to-face interviews.[26] The test-retest reliability of the TICS interview has been previously found to be high ($r \geq .9$) in a number of studies of aging and dementia.[25,27]

Genotyping and Data Quality Control

DNA was collected from saliva samples of consenting HRS respondents in 2006 or 2008. DNA from these participants was genotyped by the Center for Inherited Disease Research (CIDR) on the Illumina Human Omni-2.5 Quad array. Quality control has been performed on this data by the original HRS investigators based on the guidelines described by Laurie et al[28] and described in the Quality Control Report for Genotypic Data downloaded from dbGaP. In brief, SNPs were excluded based on the following criteria: (1) minor allele frequency (MAF) = 0 to remove mono-allelic markers; (2) missing call rate ≥ 2%; (3) discordant calls or Mendelian errors in duplicate subjects or family based samples; (4) Hardy-Weinberg equilibrium (HWE) P-value $<10^{-4}$ in European or African samples; (5) sex difference in allelic frequency ≥0.2; (6) sex differences in heterozygosity >0.3; and (7)

MAF < 0.01. As part of the centralized quality control procedures, duplicate and subjects were removed (this information is made available "Sample analysis.csv" file in dbGaP). Based on this information, we excluded 113 subjects found to have been cryptically related. Finally, since training datasets for the polygenic scores were derived from samples primarily of European Ancestry, we restricted our analyses to HRS participants who were of non-Hispanic Caucasian descent ($N = 8616$). We subsequently used the results of the ancestry based principal component analyses to exclude any potential outliers (defined as having a principal component >6 standard deviations from the mean) that may have been misclassified in regard to their ancestry ($N = 59$).

Imputation was originally performed for 12 507 samples using the phase I of the 1000 genomes as a reference dataset. The genotype and imputed data was obtained from dbGAP (accession ID: phs000428.v1.p1) following appropriate IRB approval. The imputed files were initially available in IMPUTE2 format. We extracted markers used in the polygenic scores in each disorder's original training set marker and converted the IMPUTE2 dosage format into a single dosage format for polygenic risk scoring in PLINK1.9.[29]

Polygenic Risk Score

PRS were calculated using the method originally described by the International Schizophrenia Consortium.[30] We obtained summary results of the most recent PGC analyses of SCZ,[22] BD,[31] and Major Depressive Disorder (MDD).[32] These summary statistics were pruned by the original study investigators to exclude markers in linkage disequilibrium. We also obtained the results of a recent meta-analysis of ALZ performed by the IGAP consortium.[33] We subsequently used genotypes from the Caucasian participants of the 1000G project to prune the ALZ training set, removing markers in linkage disequilibrium using PLINK (clumping markers within 500kb and an $r^2 > .25$ of the index SNPs).

PRS for the HRS sample were calculated in PLINK using the imputed HRS data and the relevant training dataset. The logistic regression parameter (β) from the training dataset was used to weight the scoring. We calculated PRS for each individual for each of the 4 disorders at 8 *P*-value thresholds (P_T) in PLINK ($P_T < .5$; $P_T < .4$; $P_T < .3$; $P_T < .2$; $P_T < .1$; $P_T < .05$; $P_T < .01$; $P_T < .001$).

Analysis

To utilize the longitudinal nature of cognition data in the HRS, we used a mixed modeling approach accounting for the repeated measures on each participant. Analyses were performed in STATA 12.1 using the *xtmixed* command. The random-intercept and random slope model included fixed effects terms for age, quadratic age, sex, education, diabetes, stroke, depression and the first 4 ancestry-based principal components. Four principal components were chosen based on evaluation of a scree plot of 10 principal components, which showed minimal changes in eigenvalues starting between the third and fourth principal component. The phenotypic covariates were chosen based on evidence from the literature that they are associated with decreased cognitive function or cognitive decline.[33–36] Depression was measured using an 8-item version of the 20-item CES-D scale, which is widely used in population based studies and has been demonstrated to have high reliability and validity.[37] The CESD-D was also validated in 2 of the HRS waves and was found to have high internal reliability (Cronbach's $\alpha = .81$ and .83).[38] Diabetes and stroke status were operationalized as binary variables based on self-report of these conditions at any time during the study. Educational attainment was operationalized as a class variable (high school incomplete, general educational development [GED], high school graduate, college graduate, graduate degree holder).

Our primary analysis included all subjects of non-Hispanic Caucasian ancestry with at least 1 total cognition variable. As prior studies have shown that the major adult mental disorders are associated with lower cognitive function,[15] we performed polygenic association analyses using 1-sided tests. We fit mixed effect models using the previously described covariates and polygenic scores as fixed effects, while including each individual as a random effect (random intercept). Age and age^2 were included in the model both as fixed and random effects (random slopes). To further characterize the effect of the polygenic scores on cognition, we examined their association with the verbal memory and mental status (attention and language) components of the cognitive exam. The distribution of the primary total cognition and the verbal memory variable were normally distributed, however, the attention and language components of the cognitive measure showed a strong left skew with a prominent ceiling effect (supplementary figure 1).

For polygenic scores that showed a significant main effect in the primary analysis (table 2), we performed an additional analysis to test whether the effect of the polygenic score changes over time, either in a linear or nonlinear (quadratic) pattern. We included interaction terms to test the effect of polygenic score with both age and age^2. A significant interaction term ($P < .05$) was interpreted as evidence to suggest that polygenic scores were associated with differing rates of cognitive decline.

A frequently used metric to measure and compare the effect of PRS on a phenotype of interest is the estimated proportion of phenotypic variance (R^2) explained by the fitted model. The effect of the polygenic score can be estimated by subtracting a model that includes the score to a baseline model that includes all other covariates except for the polygenic score. Although the calculation of R^2 in traditional linear or logistic models is commonplace,

mixed models require methods that account for the correlation of individual data and the presence of hierarchical levels. In our analyses, we have used the Bosker-Snjiders method to estimate R^2 in our multilevel analysis.[39] This method accounts for hierarchical clustering (random intercepts) but does not take into account any potential explanatory effects from random slopes. It should therefore be seen as an approximation of R^2 best used to compare the results within our study, rather than across other studies, which may employ similar but not fully comparable study designs.[40]

Results

Participant Characteristics

A total of 8616 individuals were included in the sample. The average number of total cognition measures per individual was 4.6, providing a total of 40 257 measures to be included in the analysis. The majority of subjects (63.3%) had ≥4 cognition measures (table 1). Subjects tended to be evenly distributed across educational attainment levels. The mean CES-D score in this sample was 0.96 (SD = 1.6), while 21.8% reported a history of diabetes and 11.3% reported at least 1 stroke. Participants had a mean follow-up period of 10.0 years.

SCZ PRS and Cognition

The PRS for SCZ was significantly associated with lower total cognition scores across all training set *P*-value thresholds (supplementary table 1), with the strongest results found using a training set threshold of $P_T = .05$ ($\beta = -.04$, $Z = -3.00$. $P = .001$). The proportion of variance accounted for by the polygenic score (ΔR^2) was 0.04% (table 2). Further analysis of the polygenic association showed that it primarily affected the attention and orientation cognitive measure ($P_T = .05$: $\beta = -.02$, $z = -3.33$, $P = 4.3 \times 10^{-4}$, $\Delta R^2 = 0.08\%$), with relatively smaller effects on the verbal memory components of the cognitive score ($P_T = .05$: $\beta = -.02$, $z = -2.09$, $P = .02$, $\Delta R^2 = 0.02\%$; table 3).

Table 1. Sociodemographic Characteristics of Health and Retirement Study Participants of Individuals With Genotype Data and At Least 1 Total Cognition Measures Taken at Age ≥50[a] (*N* = 8616)

Characteristics	Mean (SD) or *n* (%)	Range
Age at entry (*M*, SD)	60.5 (8.5)	
Women (*N*, %)	4841 (56.2)	
Smoker (*N*, %)	1550 (18.0)	
Stroke (*N*, %)	975 (11.3)	
Diabetes (*N*, %)	1869 (21.7)	
CES-D (*M*, SD)	0.96 (1.6)	0–8
Psychiatric history (*N*, %)	1546 (17.9)	
Education (*N*, %)		
1. Limited HS	1158 (13.4)	1–5
2. GED	365 (4.3)	
3. HS graduate	2896 (33.6)	
4. Some college	2027 (23.5)	
5. College and above	2170 (25.2)	
Years of follow-up (*M*, SD)	10.0 (5.5)	0–14
Measures per person (*M*, SD)	4.6 (2.6)	1–8
% with 1 measure	1807 (21.0)	
% with 2 measures	608 (7.1)	
% with 3 measures	703 (8.1)	
% with ≥4 measures	5498 (63.8)	

Note: *M*, mean; GED, general educational development; HS, high school.

BD and MDD PRS and Cognition

Given the emerging evidence that cognition may be familially related to a broad range of psychiatric disorders, we tested whether polygenic scores derived from the PGC meta-analyses of other adult phenotypes (BD and MDD) were also associated with decline in cognitive function in the HRS. However, we found no significant association between polygenic risk for BD (best $P_T = .001$: $\beta = -.05$, $z = -1.50$, $P = .07$, $\Delta R^2 = 0.01\%$) or MDD (best $P_T = .2$: $\beta = -.000$, $z = -0.07$, $P = .37$, $\Delta R^2 = 0$). These cross-disorder results are shown in table 2.

Alzheimer's PRS, APOE4 Locus, and Cognition

To compare the effect of polygenic scores from the common adult psychiatric disorders to an established neurocognitive disorder with a known genetic risk factor of major effect (*APOE4* allele), we obtained summary data from the IGAP GWAS of ALZ and used it as a training dataset (after appropriate LD pruning) to calculate polygenic scores in the HRS dataset. The *APOE4/TOMM40* risk locus (rs769449) showed a strongly significant association with lower total cognition scores ($\beta = -.36$, $z = -5.05$, $P = 2.3 \times 10^{-7}$, $\Delta R^2 = 0.36\%$). The polygenic score for ALZ was more modestly associated with total cognition ($P_T = .01$: $\beta = -.02$, $z = -1.88$, $P = .03$, $\Delta R^2 = 0.05\%$), suggesting that the association with cognitive impairment is driven primarily by a single

Table 2. Association of Disorder Specific Polygenic Risk Scores With the Total Cognition Measure

	Total Cognition Measure			
Disorder (Training Set *P*-value Threshold)	β	*Z*-score	*P*-value	ΔR^2 (%)
SCZ ($P_T = .05$)	−.04	−3.00	.001	0.04
BD ($P_T = .001$)	−.05	−1.50	.07	0.01
MDD ($P_T = .2$)	−.0005	−0.07	.47	0.00
ALZ ($P_T = .01$)	−.02	−1.88	.03	0.05
APOE/TOMM40 locus	−.36	−5.05	2.2×10^{-7}	0.36

Note: P_T, *P*-value threshold, β, beta-value for fixed effects; ΔR^2, difference in Bosker-Snijders R^2; SCZ, schizophrenia; BD, bipolar disorder; ALZ = Alzheimer's disease.

locus rather than a polygenic component. We further characterized the effect of the *APOE4/TOMM40* locus and the ALZ polygenic score on the attention/orientation and verbal memory subcomponents of total cognition, finding that, in contrast to the results with the SCZ polygenic scores, both the ALZ polygenic score and the *APOE4/TOMM40* locus were more strongly associated with verbal memory components of the cognitive measure (table 3 and figure 1).

Effect of PRS on Decline in Cognition

Using the primary total cognition score, we tested whether the PRS that were significant in table 2 (SCZ, ALZ, and the *APOE4/TOMM40* locus) also showed an association with cognitive decline as subjects aged. We tested for an interaction between polygenic score and age, and found no effect of the SCZ polygenic score (P_T = .05) on cognitive decline (P > .1 for both linear and quadratic interaction terms). However, we found a strong effect of the *APOE4/TOMM40* locus with cognitive decline, with significant interactions seen between the *APOE4/TOMM40* risk locus and both linear ($P = 1.5\times10^{-21}$) and quadratic age ($P = 1.8\times10^{-4}$). The effect of ALZ polygenic score (P_T = .01) cognitive decline was more modest, with a significant interaction seen with linear age (P = .02) but not quadratic age (P = .30).

Discussion

In this study we tested whether polygenic risk for a number of psychiatric disorders was associated with decreased general cognitive function, and whether this effect increased with age. We found that PRS for SCZ was associated with decreased total cognition scores (P = .001, ΔR^2 = 0.04%), and that most of this association was driven by decreased performance on a subcomponent of the cognitive score measuring attention and language ($P = 4.3\times10^{-4}$, ΔR^2 = 0.08%). Not unexpectedly, the extent of the variance in cognition explained by SCZ PRS was modest, especially when compared to the *APOE4/TOMM40* locus, a well know risk factor for ALZ and cognitive decline.[41,42]

Interestingly, the pattern of cognitive deficits affected by the SCZ polygenic risk alleles differed from those of the *APOE4/TOMM40* loci, the former being primarily driven by deficits in mental status (language and attention), with the latter manifesting mostly in verbal

Table 3. Association of Disorder-Specific Polygenic Risk Scores With Mental Status (Attention/Language) and Verbal Memory Components of Total Cognition

Disorder (Training Set *P*-value Threshold)	Attention/Language				Verbal Memory			
	β	*z*	*P*-value	ΔR^2 (%)	β	*z*	*P*-value	ΔR^2 (%)
SCZ (P_T = .05)	−.02	−3.33	4.3×10^{-4}	0.08	−.02	−2.09	.02	0.02
BD (P_T = .001)	−.01	−0.37	.36	0	−.03	−1.32	.09	0.01
MDD (P_T = .2)	.00	−0.21	.42	0	.00	0.25	.60	0
ALZ (P_T = .01)	−.01	−0.97	.17	0.03	−.02	−2.61	.005	0.05
APOE/TOMM40 locus	−.08	−2.52	.0058	0.21	−.25	−5.15	1.3×10^{-7}	0.21

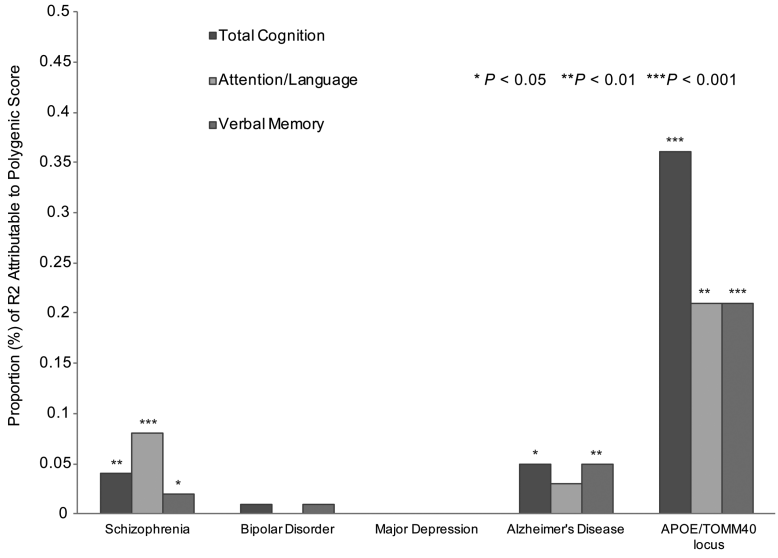


Fig. 1. Cross-disorder polygenic effects on specific cognitive domains.

memory. Verbal memory, language and attention, as measured by the TICS, have been previously found to be heritable,[43] suggesting that this may reflect domain-specific impairments that are related to the differing genetic risk of SCZ risk alleles and *APOE4*. An additional difference between the effects of the SCZ polygenic score and the *APOE4* locus was the lack of effect on cognitive decline seen in the former and the prominent effect seen in the latter. These results are consistent with a modest but "static" effect of SCZ risk alleles on overall cognition, in contrast to the deteriorating, age related effect of the major risk locus of Alzheimer's disorder. As such, these results are broadly consistent with these disorders being traditionally considered as neurodevelopmental vs neurodegenerative disorders.

In contrast to SCZ, we did not find an association between polygenic risk for BD or MDD and total cognition (figure 1). This is perhaps not surprising, given that the cognitive deficits seen in these disorders is milder than in SCZ.[4,44,45] However, one important consideration is that the primary meta-analyses of these disorders used in the polygenic training sets were smaller in sample size than that of SCZ, and were likely less predictive.

While PRS from the most recent PGC2 meta-analysis were found to explain 7%–18% of the phenotypic variance of SCZ case status in independent samples,[22,46] the proportion of variance explained by polygenic scores when applied to alternative phenotypes, particularly in unselected participants, has been much smaller and more consistent with the findings of our study (supplementary table 2).[17,19,47,48] Indeed, our findings were consistent with the modest effects found by the recent COGENT study,[19] where pseudo-R^2 values in the individual studies ranged from 0% to 2%. McIntosh et al[17] also found a small but significant effect of SCZ polygenic risk on both cognitive decline and ability in the Lothian Birth Cohort ($R^2 <$ 1%). Moreover, while most studies of the effect of polygenic risk on cognition have so far focused on generalized measures of cognitive function,[17,19] a recent study of the effects of polygenic risk for SCZ on cognitive performance also found evidence for varying associations with differing cognitive phenotypes.[48] Consistent with our results, the 2 cognitive phenotypes that showed significant associations with the SCZ polygenic score were measures of attention and spatial working memory.

The significant but modest association of polygenic scores with cognition suggests that other factors beyond common variation are likely to be associated with the cognitive impairment seen in SCZ. However, one important limitation of currently available GWAS data is that they do not fully account for all the heritability attributed to common variation.[49] As GWAS meta-analyses increase in sample size, this gap, which has been termed "hidden" heritability, will lessen and the ensuing polygenic scores will become more predictive.[50] Polygenic scores also do not reflect the contribution of rare variation and copy-number variants, which play an important role in neurodevelopmental disorders with intellectual disability,[51,52] although their role in cognitive function in the general population is unclear. In addition, polygenic scores, by definition, do not index any significant environmental factors involved in general cognition and cognitive decline. In this study, we have controlled for the role of education, but did not specifically evaluate the role of additional environmental factors, such as life stressors,[53] physical activity,[54] occupational activity,[55] and social engagement,[56] all of which have been previously shown to have associations with cognition in later life.

Several additional limitations of our study should be taken into consideration. First, although the HRS sample was designed to specifically test cognition in later life, its large ascertainment and longitudinal nature necessitated the use of an abbreviated cognitive measure designed to be administrable by lay interviewers over the telephone. This could introduce potential bias if subjects with higher rates of cognitive or mental disorders are less likely to respond to the study surveys. Although such a bias could lead to a loss of power, it is likely to be limited given the high rates of study participation (>80%) and reinterview participation (>90%) seen across all of the HRS data collection waves.[57] Second, while previous studies have demonstrated performance on the TICS to be both heritable and a well validated dementia-screening tool,[43,58,59] it is less comprehensive than more traditional neurocognitive batteries and may have limited power to detect associations seen primarily with specific cognitive domains. However, the consistency of our findings with the prior literature provides reassurance that the cognitive measurement was sufficiently robust. Third, the HRS sample was comprised entirely of older individuals, which may limit the generalization of our results to a younger population, although it may also hold the advantage of testing cognition during an age range when it is highly heritable.[20] Finally, the greater medical comorbidity in the elderly could have confounded our results, particularly since common disorders such as stroke and diabetes may also be associated with lower cognitive function. However, in our primary analyses, we included both stroke and diabetes as time-dependent covariates. Moreover, in an additional sensitivity analyses, we excluded subjects with a history of stroke of diabetes and found reassuringly similar results (supplementary table 3).

In sum, we found an association between increased polygenetic risk of SCZ and a modest decline in overall cognitive performance in older adults. Moreover, we provide initial evidence that this decline may be domain-specific and potentially distinguishable from the cognitive deficits associated with the *APOE4* risk loci. As such, our findings are consistent with a modest degree of shared genetic risk between SCZ and overall cognition, but they also point to the importance of measuring domain specific cognitive phenotypes to help delineate the specific type of deficits associated with of SCZ.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

Supported, in part, by National Institute of Mental Health (NIMH) grant R00MH86049 to FSG. The Health and Retirement Study genetic data is sponsored by the National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and was conducted by the University of Michigan.

Acknowledgment

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. van Os J, Kapur S. Schizophrenia. *Lancet*. 2009;374:635–645.
2. Kremen WS, Seidman LJ, Pepple JR, Lyons MJ, Tsuang MT, Faraone SV. Neuropsychological risk indicators for schizophrenia: a review of family studies. *Schizophr Bull*. 1994;20:103–119.
3. Watters AJ, Gotlib IH, Harris AWF, Boyce PM, Williams LM. Using multiple methods to characterize the phenotype of individuals with a family history of major depressive disorder. *J Affect Disord*. 2013;150:474–480.
4. Hill SK, Reilly JL, Keefe RSE, et al. Neuropsychological impairments in schizophrenia and psychotic bipolar disorder: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *Am J Psychiatry*. 2013;170:1275–1284.
5. Sitskoorn MM, Aleman A, Ebisch SJH, Appels MCM, Kahn RS. Cognitive deficits in relatives of patients with schizophrenia: a meta-analysis. *Schizophr Res*. 2004;71:285–295.
6. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003;160:636–645.
7. Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 2009;373:234–239.
8. Neale BM, Sklar P. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. *Curr Opin Neurobiol*. 2015;30:131–138.
9. Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatry*. 2015;20:98–108.
10. Davies G, Armstrong N, Bis J, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium. *Mol Psychiatry*. 2015;183:183–192.
11. Zhang C, Pierce BL. Genetic susceptibility to accelerated cognitive decline in the US Health and Retirement Study. *Neurobiol Aging*. 2014;35:1512.e11–1512.e18.
12. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39:17–23.
13. Fowler T, Zammit S, Owen MJ, Rasmussen F. A population-based study of shared genetic variation between premorbid IQ and psychosis among male twin pairs and sibling pairs from Sweden. *Arch Gen Psychiatry*. 2012;69:460–466.
14. Toulopoulou T, Goldberg TE, Mesa IR, et al. Impaired intellect and memory: a missing link between genetic risk and schizophrenia? *Arch Gen Psychiatry*. 2010;67:905–913.
15. McGrath JJ, Wray NR, Pedersen CB, Mortensen PB, Greve AN, Petersen L. The association between family history of mental disorders and general cognitive ability. *Transl Psychiatry*. 2014;4:e412.
16. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry Allied Discip*. 2014;10:1068–1087.
17. McIntosh AM, Gow A, Luciano M, et al. Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biol Psychiatry*. 2013;73:938–943.
18. Psychiatric GWAS Consortium Schizophrenia Disorder Working Group. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969–978.
19. Lencz T, Knowles E, Davies G, et al. Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). *Mol Psychiatry*. 2014;19:168–174.
20. Deary IJ, Johnson W, Houlihan LM. Genetic foundations of human intelligence. *Hum Genet*. 2009;126:215–232.
21. Reynolds CA, Finkel D. A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychol Rev*. 2015;25:97–112.
22. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.
23. Juster FT, Suzman R. An overview of the health and retirement study. *J Hum Resour*. 1995;30:S7–S56.
24. Growing Older in America: The Health and Retirement Study. 2015. https://www.nia.nih.gov/health/publication/growing-older-america-health-and-retirement-study/preface. Accessed October 1, 2015.
25. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol*. 1988;1:111–117.
26. Herzog A, Rodgers W. Cognitive Performance in Survey Research in Older Adults in Cognition, Aging and Self-Repeorts. In: Schwarz N, Park D, Knauper B, Sudman S, eds. *Cognition, Aging, and Self-Reports*. Philadelphia, PA: Psychology Press; 1999:337.
27. Desmond DW, Tatemichi TK, Hanzawa L. The Telephone Interview for Cognitive Status (TICS): reliability and validity in a stroke sample. *Int J Geriatr Psychiatry*. 1994;9:803–807.
28. Laurie CC, Doheny KF, Mirel DB, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol*. 2010;34:591–602.
29. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
30. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.

31. Group PGCBDW. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011;43:977–983.
32. Major Depressive Disorder Work Group of the Psychiatric GWAS Consortium; Ripke S, Wray NR, Lewis CM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2013;18:497–511.
33. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452–1458.
34. Plassman B, Williams J. Systematic review: factors associated with risk for and possible prevention of cognitive decline in later life. *Ann Intern Med*. 2010;153:182–193.
35. Biessels GJ, Deary IJ, Ryan CM. Cognition and diabetes: a lifespan perspective. *Lancet Neurol*. 2008;7:184–190.
36. Mok VCT. Cognitive impairment and functional outcome after stroke associated with small vessel disease. *J Neurol Neurosurg Psychiatry*. 2004;75:560–566.
37. Turvey CL, Wallace RB, Herzog R. A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. *Int Psychogeriatr*. 1999;11:139–148.
38. Health and Retirement Study: Documentation of Affective Functioning Measures in the Health and Retirement Study. 2000. http://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf. Accessed October 15, 2015.
39. Snijders TAB, Bosker RJ. Modeled variance in two-level models. *Sociol Methods Res*. 1994;22:342–363.
40. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol*. 2012;36:214–224.
41. Deary IJ, Whiteman MC, Pattie A, et al. Cognitive change and the APOE epsilon 4 allele. *Nature*. 2002;418:932.
42. Schiepers OJG, Harris SE, Gow AJ, et al. APOE E4 status predicts age-related cognitive decline in the ninth decade: longitudinal follow-up of the Lothian Birth Cohort 1921. *Mol Psychiatry*. 2012;17:315–324.
43. Brandt J, Welsh KA, Breitner JC, Folstein MF, Helms M, Christian JC. Hereditary influences on cognitive functioning in older men. A study of 4000 twin pairs. *Arch Neurol*. 1993;50:599–603.
44. Hill SK, Keshavan MS, Thase ME, Sweeney JA. Neuropsychological dysfunction in antipsychotic-naive first-episode unipolar psychotic depression. *Am J Psychiatry*. 2004;161:996–1003.
45. Depp CA, Moore DJ, Sitzer D, et al. Neurocognitive impairment in middle-aged and older adults with bipolar disorder: comparison to schizophrenia and normal comparison subjects. *J Affect Disord*. 2007;101:201–209.
46. Agerbo E, Sullivan PF, Vilhjálmsson BJ, et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia. *JAMA Psychiatry*. 2015;72:635–641.
47. Hill WD, Davies G, Liewald DC, McIntosh AM, Deary IJ. Age-dependent pleiotropy between general cognitive function and major psychiatric disorders. *Biol Psychiatry*. In press. doi:10.1016/j.biopsych.2015.08.033.
48. Hatzimanolis A, Bhatnagar P, Moes A, et al. Common genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168B:392–401.
49. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet*. 2014;15:765–776.
50. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348.
51. Deciphering T, Disorders D. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2014;10:223–228.
52. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511:344–347.
53. Gould F, Clarke J, Heim C, Harvey PD, Majer M, Nemeroff CB. The effects of child abuse and neglect on cognitive functioning in adulthood. *J Psychiatr Res*. 2012;46:500–506.
54. Zhu N, Jacobs DR, Schreiner PJ, et al. Cardiorespiratory fitness and cognitive function in middle age: the CARDIA study. *Neurology*. 2014;82:1339–1346.
55. Then FS, Luck T, Luppa M, König H-H, Angermeyer MC, Riedel-Heller SG. Differential effects of enriched environment at work on cognitive decline in old age. *Neurology*. 2015;84:2169–2176.
56. Bassuk SS, Glass TA, Berkman LF. Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Ann Intern Med*. 1999;131:165–173.
57. Health and Retirement Study: Sample Sizes and Response Rates. 2001. http://hrsonline.isr.umich.edu/sitedocs/sampleresponse.pdf. Accessed October 15, 2015.
58. De Jager CA, Budge MM, Clarke R. Utility of TICS-M for the assessment of cognitive function in older adults. *Int J Geriatr Psychiatry*. 2003;18:318–324.
59. Knopman DS, Roberts RO, Geda YE, et al. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology*. 2010;34:34–42.

HHS Public Access

Author manuscript

Psychol Med. Author manuscript; available in PMC 2017 November 17.

Published in final edited form as:

Psychol Med. 2015 August ; 45(11): 2437–2446. doi:10.1017/S0033291715000446.

Distinguishing bipolar from unipolar depression: the importance of clinical symptoms and illness features

A. K. Leonpacher[1,†], **D. Liebers**[1,†], **M. Pirooznia**[1], **D. Jancic**[2], **D. F. MacKinnon**[1], **F. M. Mondimore**[1], **B. Schweizer**[1], **J. B. Potash**[2], **P. P. Zandi**[1,3], **NIMH Genetics Initiative Bipolar Disorder Consortium, GenRED Consortium**, and **F. S. Goes**[1,*]

[1]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

[2]Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA, USA

[3]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Abstract

Background—Distinguishing bipolar disorder (BP) from major depressive disorder (MDD) has important relevance for prognosis and treatment. Prior studies have identified clinical features that differ between these two diseases but have been limited by heterogeneity and lack of replication. We sought to identify depression-related features that distinguish BP from MDD in large samples with replication.

Method—Using a large, opportunistically ascertained collection of subjects with BP and MDD we selected 34 depression-related clinical features to test across the diagnostic categories in an initial discovery dataset consisting of 1228 subjects (386 BPI, 158 BPII and 684 MDD). Features significantly associated with BP were tested in an independent sample of 1000 BPI cases and 1000 MDD cases for classifying ability in receiver operating characteristic (ROC) analysis.

Results—Seven clinical features showed significant association with BPI compared with MDD: delusions, psychomotor retardation, incapacitation, greater number of mixed symptoms, greater number of episodes, shorter episode length, and a history of experiencing a high after depression treatment. ROC analyses of a model including these seven factors showed significant evidence for discrimination between BPI and MDD in an independent dataset (area under the curve = 0.83). Only two features (number of mixed symptoms, and feeling high after an antidepressant) showed an association with BPII *versus* MDD.

[*]Address for correspondence: F. S. Goes, M.D., Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Meyer 4-119A, 600 N. Wolfe Street, Baltimore, MD 21287, USA. (fgoes1@jhmi.edu).

[†]These authors contributed equally.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291715000446

Declaration of Interest

None.

Conclusions—Our study suggests that clinical features distinguishing depression in BPI *versus* MDD have important classification potential for clinical practice, and should also be incorporated as 'baseline' features in the evaluation of novel diagnostic biomarkers.

Keywords

Bipolar disorder; diagnosis; major depressive disorder; mixed symptoms; psychosis

Introduction

The syndrome of depression is a functionally debilitating condition common in both community and treatment settings (Murray *et al.* 2012). Although symptoms of depression are present in many psychiatric and somatic illnesses, depressive episodes are most prominent in bipolar disorder (BP) and in major depressive disorder (MDD), the two most common mood disorders that, together, afflict up to one-fifth of the world's population (Bromet *et al.* 2011; Merikangas *et al.* 2011). The distinction between bipolar and unipolar illness, first made decades ago, was originally based on differing patterns of external validators such as family history, sex and premorbid personality (Leonhard *et al.* 1962; Angst, 1966; Perris, 1966). The discovery of psychopharmacological agents, with their relative specificity for depressive or manic syndromes, largely supported this distinction (Pacchiarotti *et al.* 2013), and highlighted the need to minimize the misdiagnosis of MDD in patients with BP, since antidepressant treatment may tend to worsen mood stability in BP (Wehr & Goodwin, 1987; Altshuler *et al.* 1995; Henry *et al.* 2001). Nevertheless, misdiagnosis between BP and MDD depression remains common in both primary care and psychiatric clinics (Ghaemi *et al.* 1999, 2000; Angst *et al.* 2011).

A major challenge in the diagnosis of BP is the relative infrequency of episodes of mania and hypomania in comparison with the longer and more frequent periods of depression (Judd *et al.* 2003; Altshuler *et al.* 2010). In addition, most individuals diagnosed with BP experience the onset of their illness with a depressive rather than manic episode (Lish *et al.* 1994). In recent decades, studies have attempted to identify features of illness that help distinguish patients with bipolar depression from those with MDD. Several of these features have been consistently found to be more prominent in bipolar depression: earlier age at onset, increased number of depressive episodes, and greater propensity for hypersomnia, psychomotor abnormalities, and psychotic symptoms (Mitchell *et al.* 2001, 2011; Serretti *et al.* 2002; Perlis *et al.* 2006; Goes *et al.* 2007; Souery *et al.* 2012; Tondo *et al.* 2014). However, differing phenotype assessments and widely varying ascertainment schemes have limited the comparison of anything but broad trends across studies.

More recently, attention has also focused on the identification of biomarkers that could assist in differentiating BP from MDD depression. Structural and functional MRI studies have been performed, though their interpretation is limited by small sample size, differing protocols, and often divergent results (Cardoso de Almeida & Phillips, 2013). Similarly, biomarker studies, mainly focused on peripheral proteins such as brain-derived neurotrophic factor, have identified potential differences across the diagnoses, but these findings remain preliminary (Fernandes *et al.* 2009; Li *et al.* 2014). Finally, while robust genetic association

findings are emerging for BP, the modest effect sizes of both individual markers and their combinations limit their diagnostic utility at present (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011). Hence, while biomarker studies will probably provide insights into illness pathophysiology, they are currently neither sufficiently robust nor sufficiently predictive to aid in the important differentiation between depressive episodes between BP compared with MDD.

Given the limited current availability of diagnostic biomarkers and the difficulties in interpreting across previous studies of differential diagnostic features, we sought to revisit the question of which symptoms, clinical characteristics, and co-morbidities may be of use in identifying patients with depression due to BP. We took advantage of a large dataset of subjects with BP and recurrent major depression diagnosed with substantively identical semi-structured interviews and best-estimate procedures, thus allowing for appropriate cross-diagnosis and cross-study comparability. We identify depression-related features specifically associated with BP and implement a simple predictive model that performs robustly in a fully independent dataset.

Method

Subjects

We analysed diagnostic and interview data from large BP [National Institute of Mental Health (NIMH) Genetics Initiative] and MDD (Genetics of Recurrent Early-Onset Depression; GenRED) collections, initially focused on ascertainment of samples for genetic studies. Both studies began by ascertaining families with at least two affected family members and transitioned to the collection of singleton subjects. Diagnoses for both studies were based on the Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger *et al.* 1994), followed by best-estimate procedures. In this analysis, we utilized samples with high confidence BPI, BPII or MDD diagnoses from both familial and singleton samples in both the BP and MDD samples. For the family samples, only one affected member was used per family. To minimize potential effects from ascertainment bias, we excluded all probands from each family and used a randomly selected affected relative.

We split the BP and MDD samples into a discovery and replication dataset. For both samples, the discovery dataset consisted of an affected relative (one per family) from the family-based collection. To increase the number of BPII cases, we preferentially selected BPII relatives if they were available. The final number of subjects in the discovery dataset was: 386 BPI, 158 BPII and 684 MDD. The replication datasets were drawn from the later singleton collections of the BP and MDD studies. We randomly selected 1000 cases with BPI and 1000 cases with MDD.

Phenotype selection

We examined three categories of clinical features: (*a*) symptoms during the most severe depression; (*b*) lifetime clinical characteristics of depression; and (*c*) co-morbidities. Diagnoses were obtained from the best-estimate consensus, while symptoms were obtained directly from the DIGS interview (most severe depression section). Clinical characteristics

were obtained from the DIGS and the best-estimate interviews. In total, we tested 34 features, including both categorical and dimensional variables (Tables 1–3)

Analysis

We compared (*a*) symptoms, (*b*) clinical characteristics and (*c*) co-morbidities, between BP and MDD, using univariate and multivariate logistic regression. Significance testing was initially performed with nominal two-sided *p* values. We subsequently selected features that remained significant after correction for the 34 clinical variables examined (Bonferroni $p < 0.0014$). These features were included in a multivariate logistic regression model, controlling for age and sex. Features that remained significant in the full model were retained and this final model was then tested for its ability to distinguish cases of BP and MDD using a receiver operating characteristic (ROC) analysis in both the discovery and independent replication samples. In secondary analyses, we compared BPII subjects with MDD subjects. All analyses were conducted in Stata 12.1 (USA).

Results

Sociodemographic characteristics

Table 1 shows the major demographic features of the selected subjects by diagnosis, combining the discovery and independent target datasets. Of the 1228 participants initially analysed in the first dataset, 386 had a BPI diagnosis, 158 had a BPII diagnosis and 684 had an MDD diagnosis. Major demographic variables were consistent across the various diagnostic groups, with certain expected differences (Table 1): for example, BPI participants were more likely to be on disability (17.1%) compared with BPII and MDD depression participants (3.2% and 4.2%, respectively).

Distinguishing BPI from MDD

We divided the associated phenotypes into three categories: (1) symptoms during the most severe depressive episode; (2) lifetime clinical characteristics; and (3) co-morbidities. In the primary analysis of subjects with BPI *versus* those with MDD (Table 2), six symptoms showed evidence of association with BPI (using Bonferroni-corrected *p* value cut-off of 0.0014): psychomotor agitation [odds ratio (OR) 1.56, $p < 0.001$], psychomotor retardation (OR 2.51, $p < 0.001$), thoughts of self-harm (OR 2.11, $p < 0.001$), delusions (OR 7.72, $p < 0.001$), hallucinations (OR 4.75, $p < 0.001$) and functional incapacitation (OR 5.00, $p < 0.001$). Similarly, six lifetime clinical characteristics were also found to be strongly associated with BPI (Table 3): high after anti-depressive treatment (OR 6.24, $p < 0.001$), hospitalization for depression (OR 3.40, $p < 0.001$), greater number of lifetime depressive episodes (OR 1.02, $p < 0.001$), shorter most severe episode (OR 0.97, $p < 0.001$), greater number of mixed symptoms (defined as the count of seven potential mixed symptoms during the most severe depression) (OR 1.43, $p < 0.001$), and greater number of lifetime suicide attempts (OR 1.20, $p < 0.001$). The differences for the co-morbid diagnoses were less pronounced: a nominal association was seen for social phobia, but only the category of alcohol abuse and dependence was found to be significantly associated with BPI (OR 1.66, $p < 0.001$) after correcting for multiple testing.

We subsequently included all 13 associated features into a multiple logistic regression, and found that seven features remained significant in that model after Bonferroni correction (Table 4). These seven features were included in a final model and its ROCs were examined for their ability to distinguish BPI from MDD. ROC analysis of the discovery dataset showed, as expected, evidence for very good differentiation [AUC 0.84, 95% confidence interval (CI) 0.81– 0.87] (see online Supplementary Fig. S1). Importantly, very similar results were found when we tested these seven factors in an independent dataset of 1000 cases with BPI and 1000 cases with MDD. As shown in Fig. 1, the ROC analysis in the independent dataset had a significant AUC of 0.83 (95% CI 0.82–0.85), with a sensitivity of 75.6% and specificity of 77.8% at the point on the curve furthest from the null. To explore which features were driving the classification performance, we split the features into symptoms during a most severe depressive episode and those lifetime clinical characteristics. As shown in Fig. 1, the four clinical characteristics (AUC = 0.78, 95% CI 0.76–0.80) and the three symptoms (AUC = 0.74, 95% CI 0.72– 0.76) contributed about equally to the classification model.

Distinguishing BPII from MDD

In contrast to the BPI comparisons, there were very few differences in symptoms of depression between BPII and MDD, and none were significant after correction for multiple testing (see online Supplementary Table S1). Further, only two clinical characteristics were significant after correcting for multiple testing: the number of mixed symptoms during the most severe depression (OR 1.48, $p < 0.001$) and feeling ‘high after depression treatment’ (OR 3.41, $p < 0.001$) (see online Supplementary Table S2). Comparison of clinical co-morbidities between BPII and MDD showed a similar pattern as was observed between BPI and MDD, with only alcohol abuse being significantly associated (OR 2.21, $p < 0.001$). When these three features were examined in a multiple logistic regression model, only the two clinical characteristics remained significant (see online Supplementary Table S3). As expected, ROC analyses of these features showed a much more limited ability to discriminate between BPII and MDD (AUC = 0.63, 95% CI 0.57–0.70) (see online Supplementary Fig. S2). Since the AUC value was low in the discovery dataset, we did not seek to test the final model in the independent dataset.

Discussion

This study provided a broad evaluation of depressive symptoms and illness features that differ between subjects with bipolar and unipolar depressions. Seven such distinguishing features were found to be statistically significant after correction for multiple testing, showing a significant ability to discriminate between the BPI and MDD diagnoses (AUC = 0.83) when tested in an independent sample. In contrast, we found relatively few differences between features of depression in BPII compared with MDD, highlighting the potentially intermediary role that BPII may occupy across the MDD–BP spectrum.

The main differences between BPI and MDD were driven by symptoms during the most severe episodes and by lifetime clinical characteristics. Among the most important symptom differences (with ORs >2) were increased likelihood of psychomotor retardation, suicidal

behavior, psychotic symptoms and overall incapacitation. Importantly, they generally indicate a pattern of greater severity in bipolar depression, which is consistent with much of the published literature (Souery *et al.* 2012) albeit with some exceptions potentially due to differences in ascertainment and diagnosis. Comparison of other associated clinical characteristics showed a greater risk of hospitalization, a higher number of shorter depressive episodes, and a greater likelihood of experiencing a greater number of mixed symptoms during the most severe depression. In agreement with prior studies (Mitchell *et al.* 2001; Moreno *et al.* 2012), we found that BPI subjects experienced more hypersomnia (Table 2), although this difference did not survive correction for multiple testing in our study. A number of other important differences in symptoms and clinical characteristics were consistent with prior studies, which have shown BPI subjects to be more likely to experience mixed symptoms during a depression (Angst *et al.* 2011), psychomotor abnormalities, delusions (Mitchell *et al.* 2011) and overall impairment (Das *et al.* 2005). Both the significant effect sizes and the consistency with the prior literature suggest that these may be particularly important features and worthy of further study.

The strongest association with BPI in our sample was experiencing a 'high after depression treatment' (OR = 6.24, $p < 0.001$), which has been shown in a multinational, large-scale study to be one of the most strongly associated clinical features with a BP diagnosis (Angst *et al.* 2011). However, although pharmacologically induced switching of mood may be one of the most distinctive clinical features associated with BP, it is feature limited to subjects who have been treated and will not aid the clinician in distinguishing first-onset or untreated depressions.

Perhaps surprisingly, there were relatively few differences among co-morbid diagnoses, with only an increased rate of alcoholism being significantly associated with BPI (OR = 1.66, $p < 0.001$). This supports a similar finding by Souery *et al.* (2012), but differs from other reports that suggest that anxiety disorders and drug use may also be more prevalent in epidemiologically ascertained BPI patients (Moreno *et al.* 2012).

The comparison between depressions in BPII *versus* MDD yielded very few differences. The overall pattern was one of more subtle differences, with only two features (high after depression treatment and a greater number of mixed symptoms) remaining significant in the multivariate logistic regression models; consequently, there was limited discrimination seen in the ROC analysis (AUC = 0.63, 95% CI 0.57–0.70). Notably, the number of mixed symptoms during a most severe depression was also elevated in BPII (OR = 1.48, $p < 0.001$), consistent with a number of prior studies more focused on BPII (Benazzi, 2007). An important caveat is that our BPII sample size was less than half that for BPI, and thus our power to detect significant differences was substantially less. However, the ORs were much smaller in most of the BPII *versus* MDD comparisons, suggesting that our failure to detect significant results was largely driven by a lack of true differences.

Having identified associated features in our discovery sample, we subsequently performed ROC analysis in an independent sample, also diagnosed with the DIGS, but which focused on collection of singleton subjects rather than families. ROC analysis of the independent sample showed an essentially identical ability of the selected seven features to discriminate

between BPI and MDD (AUC = 0.83, (95% CI 0.82–0.85). Under optimal conditions (the point on the ROC curve most displaced from the null), the clinical features would have a sensitivity of 76% and a specificity of 78%. Although there is no specific AUC score cut-off that renders a test or diagnostic procedure clinically 'useful', values > 0.70 are usually recommended, with AUC scores > 0.9 being particularly desirable for tests that require particularly high accuracy (Swets, 1988). Our AUC value in the independent sample of 0.83 falls within a range that has been typically described as providing 'very good discrimination' in studies screening for psychopathology in the population (Kessler *et al.* 2003). As a comparison, this AUC is similar, if not slightly superior to widely used cardiac outcome predictor models, which range in AUC from 0.7 to 0.8 (Cook, 2012).

As research increasingly turns towards blood-based or imaging biomarkers, what role should such clinical findings play in decision-making? Because of their ready availability, clinically based markers should represent a baseline for prediction of diagnosis and/or illness course upon which biomarkers are tested. Hence, if the goal of a hypothetical biomarker were to help distinguish between two diagnoses, the important outcome of the biomarker study would be to determine the extent to which the use of a biomarker aids in classification over and above what is provided by the 'baseline' of clinical features (Kendler, 2014). For example, a recent volumetric imaging study of BP compared with MDD identified a number of gray matter differences, which yielded slightly inferior classification performance (depending on classifier models, sensitivity ranged from 66% to 76% with specificity of 59% to 73%) compared with the clinical results presented in this paper (Redlich *et al.* 2014). In such studies, it would be of interest to know what proportion of the classification performance may be indexed by the clinical features alone, and how much additional ability to classify is provided by the putative biomarker(s). Of course, biomarkers have the added potential of providing new mechanistic insights, but in their predictive role, they are probably most likely to be useful in combination with clinical features (Ioannidis & Tzoulaki, 2012).

An important limitation of our study, common to most previous studies, is the cross-sectional nature of the diagnoses. A more informative longitudinal design would have allowed us to test whether the identified illness features could predict the development of mania in a patient who initially presents with depression. Further limitations arising from the cross-sectional design include reliance on recalled symptoms and, for feasibility, a focus only on a single 'worst' depressive episode. A few prior studies have benefitted from a longitudinal design, but they have been also limited by the pragmatic need to follow fewer patients (Akiskal *et al.* 1995) or to perform a more limited phenotypic assessment (Tondo *et al.* 2014). At present, the feasibility of collecting sufficient information on a sufficiently large number of affected subjects remains a challenge for the field. A second important potential limitation might be potential ascertainment bias that may emerge from collecting cases for genetic studies. For example, one potential bias is that the GenRED sample specifically recruited early-onset cases of MDD (age of onset <31 years), which precludes an inquiry into whether age at onset could be used as a distinguishing feature between bipolar and unipolar patients as has been found in previous studies (Souery *et al.* 2012). To limit the potential for ascertainment for more severe cases, we excluded probands from the family dataset and, reassuringly, found few differences across the discovery and independent

datasets. Although population-based surveys may provide more external validity, they are by necessity limited to self-report questionnaires [such as in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) study] (Moreno *et al.* 2012) or lay interviewer-based diagnoses (as used in the Epidemiologic Catchment Area study or the National Comorbidity Survey) (Robins & Regier, 1991; Kessler *et al.* 1994). In contrast, our diagnoses were confirmed by a well-validated interview with best-estimate diagnoses made by two supervising clinicians. Additionally, we included only diagnoses that were made with a high degree of confidence; however, one potential limitation of excluding low confidence diagnoses is that our results may be less informative to cases with more nuanced presentations. Finally, in an era when diagnostic boundaries are being increasingly called into question, our diagnostic instrument was based on Statistical Manual of Mental Disorders (DSM) criteria, which limits its use for exploration of alternative and more dimensional methods of classification.

In summary, our study, encompassing one of the largest collections of subjects with BP and MDD mood disorders, identified seven important clinical features that successfully distinguished BP from MDD patients, both in an initial and an independent dataset, which suggests that our result may have sufficient accuracy to be relevant for clinical use. Despite increasing emphasis on the discovery of novel neuroimaging and peripheral biomarkers, our study suggests that clinical features continue to have important classification potential that should not be ignored. Rather, these features should be integrated with biological markers in future studies aiming to predict diagnosis and course of illness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

NIMH Genetics Initiative: Some of the data and biomaterials were collected in four projects that participated in the NIMH BP Genetics Initiative from 1991 to 1998. The principal investigators and co-investigators were: Indiana University, Indianapolis, IN (U01 MH46282), J. Nurnberger, M.D., Ph.D., M. Miller, M.D. and E. Bowman, M.D.; Washington University, St Louis, MO (U01 MH46280), T. Reich, M.D., A. Goate, Ph.D. and J. Rice, Ph.D.; Johns Hopkins University, Baltimore, MD (U01 MH46274), J. R. DePaulo Jr, M. D., S. Simpson, M.D., M.Ph. and C. Stine, Ph.D.; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD, E. Gershon, M. D., D. Kazuba, B.A. and E. Maxwell, M.S.W. Other data and biomaterials were collected in 10 NIMH BP Genetics Initiative projects from 1999 to 2003. The principal investigators and co-investigators were: Indiana University, Indianapolis, IN (R01 MH59545), J. Nurnberger, M.D., Ph.D., M. J. Miller, M.D., E. S. Bowman, M.D., N. L. Rau, M.D., P. R. Moe, M. D., N. Samavedy, M.D., R. El-Mallakh, M.D. (at University of Louisville), H. Manji, M.D., D.A. Glitz, M.D. (at Wayne State University), E. T. Meyer, M.S., C. Smiley, R.N., T. Foroud, Ph.D., L. Flury, M.S., D. M. Dick, Ph.D. and H. Edenberg, Ph.D.; Washington University, St Louis, MO (R01 MH059534), J. Rice, Ph.D., T. Reich, M.D., A. Goate, Ph.D. and L. Bierut, M.D.; Johns Hopkins University, Baltimore, MD (R01 MH59533), M. McInnis M.D., J. R. DePaulo Jr, M.D., D. F. MacKinnon, M.D., F. M. Mondimore, M.D., J. B. Potash, M.D., P. P. Zandi, Ph.D., D. Avramopoulos, Ph.D. and J. Payne, M.D.; University of Pennsylvania, PA (R01 MH59553), W. Berrettini, M.D., Ph.D.; University of California at Irvine, CA (R01 MH60068), W. Byerley M.D. and M. Vawter M.D.; University of Iowa, IA (R01 MH059548), W. Coryell M.D. and R. Crowe M. D.; University of Chicago, IL (R01 MH59535), E. Gershon, M.D., J. Badner Ph.D., F. McMahon M. D., C. Liu Ph.D., A. Sanders M.D., M. Caserta, S. Dinwiddie M.D., T. Nguyen and D. Harakal; University of California at San Diego, CA (R01 MH59567), J. Kelsoe, M.D. and R. McKinney, B.A.; Rush University, IL (R01 MH059556), W. Scheftner M.D., H. M. Kravitz, D.O., M.P.H., D. Marta, B.S., A. Vaughn-Brown, M.S.N., R.N. and L. Bederow, M. A.; NIMH Intramural Research Program, Bethesda, MD (1Z01MH002810-01), F. McMahon, M.D., L. Kassem, Psy.D., S. Detera-Wadleigh, Ph.D., L. Austin, Ph.D. and D. L. Murphy, M.D.; GenRED Consortium: Data and biomaterials were collected in six projects that participated in the NIMH Genetics of Recurrent Early-Onset Depression (GenRED) project. From 1999 to 2003, the principal investigators and co-investigators were: New York State

Psychiatric Institute, New York, NY, Myrna M. Weissman, Ph.D. and James K. Knowles, M.D., Ph.D.; University of Pittsburgh, Pittsburgh, PA, George S. Zubenko, M.D., Ph.D. and Wendy N. Zubenko, Ed.D., R.N., C.S.; Johns Hopkins University, Baltimore, J. Raymond DePaulo, M.D., Melvin G. McInnis, M.D. and Dean MacKinnon, M.D.; University of Pennsylvania, Philadelphia, PA, Douglas F. Levinson, M.D. (GenRED coordinator), Madeleine M. Gladis, Ph.D., Kathleen Murphy-Eberenz, Ph.D. and Peter Holmans, Ph.D. (University of Wales College of Medicine); University of Iowa, Iowa City, IW, Raymond R. Crowe, M.D. and William H. Coryell, M.D.; Rush University Medical Center, Chicago, IL, William A. Scheftner, M.D. (Rush-Presbyterian).

References

Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Keller M, Warshaw M, Clayton P, Goodwin F. Switching from 'unipolar' to bipolar II. An 11-year prospective study of clinical and temperamental predictors in 559 patients. Archives of General Psychiatry. 1995; 52:114–123. [PubMed: 7848047]

Altshuler LL, Kupka RW, Hellemann G, Frye MA, Sugar CA, McElroy SL, Nolen WA, Grunze H, Leverich GS, Keck PE, Zermeno M, Post RM, Suppes T. Gender and depressive symptoms in 711 patients with bipolar disorder evaluated prospectively in the Stanley Foundation bipolar treatment outcome network. American Journal of Psychiatry. 2010; 167:708–715. [PubMed: 20231325]

Altshuler LL, Post RM, Leverich GS, Mikalauskas K, Rosoff A, Ackerman L. Antidepressant-induced mania and cycle acceleration: a controversy revisited. American Journal of Psychiatry. 1995; 152:1130–1138. [PubMed: 7625459]

Angst J. On the etiology and nosology of endogenous depressive psychoses. A genetic, sociologic and clinical study [article in German]. Monographien aus dem Gesamtgebiete der Neurologie und Psychiatrie. 1966; 112:1–118. [PubMed: 5943846]

Angst J, Azorin JM, Bowden CL, Perugi G, Vieta E, Gamma A, Young AH. BRIDGE Study Group. Prevalence and characteristics of undiagnosed bipolar disorders in patients with a major depressive episode: the BRIDGE study. Archives of General Psychiatry. 2011; 68:791–798. [PubMed: 21810644]

Benazzi F. Bipolar disorder – focus on bipolar II disorder and mixed depression. Lancet. 2007; 369:935–945. [PubMed: 17368155]

Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lépine J-P, Levinson D, Matschinger H, Mora MEM, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC. Cross-national epidemiology of DSM-IV major depressive episode. BMC Medicine. 2011; 9:90. [PubMed: 21791035]

Cardoso de Almeida JR, Phillips ML. Distinguishing between unipolar depression and bipolar depression: current and future clinical and neuroimaging perspectives. Biological Psychiatry. 2013; 73:111–118. [PubMed: 22784485]

Cook NR. Clinically relevant measures of fit? A note of caution. American Journal of Epidemiology. 2012; 176:488–491. [PubMed: 22875759]

Das AK, Olfson M, Gameroff MJ, Pilowsky DJ, Lantigua R, Shea S, Weissman MM. Screening for bipolar disorder in a primary care practice. Journal of the American Medical Association. 2005; 293:956–963. [PubMed: 15728166]

Fernandes BS, Gama CS, Kauer-Sant'Anna M, Lobato MI, Belmonte-de-Abreu P, Kapczinski F. Serum brain-derived neurotrophic factor in bipolar and unipolar depression: a potential adjunctive tool for differential diagnosis. Journal of Psychiatric Research. 2009; 43:1200–1204. [PubMed: 19501841]

Ghaemi SN, Boiman EE, Goodwin FK. Diagnosing bipolar disorder and the effect of antidepressants: a naturalistic study. Journal of Clinical Psychiatry. 2000; 61:804–808. quiz 809. [PubMed: 11078046]

Ghaemi SN, Sachs GS, Chiou AM, Pandurangi AK, Goodwin K. Is bipolar disorder still underdiagnosed? Are antidepressants overutilized? Journal of Affective Disorders. 1999; 52:135–144. [PubMed: 10357026]

Goes FS, Sadler B, Toolan J, Zamoiski RD, Mondimore FM, Mackinnon DF, Schweizer B, Raymond Depaulo J, Potash JB. Psychotic features in bipolar and unipolar depression. Bipolar Disorders. 2007; 9:901–906. [PubMed: 18076541]

Henry C, Sorbara F, Lacoste J. Antidepressant-induced mania in bipolar patients: identification of risk factors. Journal of Clinical Psychiatry. 2001; 62:249–255. [PubMed: 11379838]

Ioannidis JP, Tzoulaki I. Minimal and null predictive effects for the most popular blood biomarkers of cardiovascular disease. Circulation Research. 2012; 110:658–662. [PubMed: 22383708]

Judd LL, Akiskal HS, Schettler PJ, Coryell W, Maser J, Rice JA, Solomon DA, Keller MB. The comparative clinical phenotype and long term longitudinal episode course of bipolar I and II: a clinical spectrum or distinct disorders? Journal of Affective Disorders. 2003; 73:19–32. [PubMed: 12507734]

Kendler KS. DSM issues: incorporation of biological tests, avoidance of reification, and an approach to the “Box Canyon Problem”. American Journal of Psychiatry. 2014; 171:1248–1250. [PubMed: 25756765]

Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand SL, Manderscheid RW, Walters EE, Zaslavsky AM. Screening for serious mental illness in the general population. Archives of General Psychiatry. 2003; 60:184–189. [PubMed: 12578436]

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Archives of General Psychiatry. 1994; 51:8–19. [PubMed: 8279933]

Leonhard K, Korff I, Schulz H. Temperament in families with monopolar and bipolar phasic psychoses. Psychiatria et Neurologia. 1962; 143:416–434. [PubMed: 14464117]

Li Z, Zhang C, Fan J, Yuan C, Huang J, Chen J, Yi Z, Wang Z, Hong W, Wang Y, Lu W, Guan Y, Wu Z, Su Y, Cao L, Hu Y, Hao Y, Liu M, Yu S, Cui D, Xu L, Song Y, Fang Y. Brain-derived neurotrophic factor levels and bipolar disorder in patients in their first depressive episode: 3-year prospective longitudinal study. British Journal of Psychiatry: The Journal of Mental Science. 2014; 205:29–35.

Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RM. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. Journal of Affective Disorders. 1994; 31:281–294. [PubMed: 7989643]

Merikangas KR, Jin R, He J-P, Kessler RC, Lee S, Sampson NA, Viana MC, Andrade LH, Hu C, Karam EG, Ladea M, Medina-Mora ME, Ono Y, Posada-Villa J, Sagar R, Wells JE, Zarkov Z. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. Archives of General Psychiatry. 2011; 68:241–251. [PubMed: 21383262]

Mitchell PB, Frankland A, Hadzi-Pavlovic D, Roberts G, Corry J, Wright A, Loo CK, Breakspear M. Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. British Journal of Psychiatry: The Journal of Mental Science. 2011; 199:303–309.

Mitchell PB, Wilhelm K, Parker G, Austin MP, Rutgers P, Malhi GS. The clinical features of bipolar depression: a comparison with matched major depressive disorder patients. Journal of Clinical Psychiatry. 2001; 62:212–216. quiz 217. [PubMed: 11305713]

Moreno C, Hasin DS, Arango C, Oquendo MA, Vieta E, Liu S, Grant BF, Blanco C. Depression in bipolar disorder versus major depressive disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Bipolar Disorders. 2012; 14:271–282. [PubMed: 22548900]

Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne

C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fèvre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gonzalez-Medina D, Gosselin R, Grainger R, Grant B, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laden F, Lalloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Levinson D, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mock C, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Nevitt MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA 3rd, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leòn FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivakoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ, Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiebe N, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD, AlMazroa MA, Memish ZA. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012; 380:2197–2223. [PubMed: 23245608]

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic Interview for Genetic Studies. Rationale, unique features, and training. NIMH Genetics Initiative. Archives of General Psychiatry. 1994; 51:849–859. discussion 863–864. [PubMed: 7944874]

Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, Post RM, Berk M, Goodwin GM, Sachs GS, Tondo L, Findling RL, Youngstrom EA, Tohen M, Undurraga J, González-Pinto A, Goldberg JF, Yildiz A, Altshuler LL, Calabrese JR, Mitchell PB, Thase ME, Koukopoulos A, Colom F, Frye MA, Malhi GS, Fountoulakis KN, Vázquez G, Perlis RH, Ketter TA, Cassidy F, Akiskal H, Azorin J-M, Valentí M, Mazzei DH, Lafer B, Kato T, Mazzarini L, Martínez-Aran A, Parker G, Souery D, Ozerdem A, McElroy SL, Girardi P, Bauer M, Yatham LN, Zarate CA, Nierenberg AA, Birmaher B, Kanba S, El-Mallakh RS, Serretti A, Rihmer Z, Young AH, Kotzalidis GD, MacQueen GM, Bowden CL, Ghaemi SN, Lopez-Jaramillo C, Rybakowski J, Ha K, Perugi G, Kasper S, Amsterdam JD, Hirschfeld RM, Kapczinski F, Vieta E. The International Society for Bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. American Journal of Psychiatry. 2013; 170:1249–1262. [PubMed: 24030475]

Perlis RH, Brown E, Baker RW, Nierenberg AA. Clinical features of bipolar depression *versus* major depressive disorder in large multicenter trials. American Journal of Psychiatry. 2006; 163:225–231. [PubMed: 16449475]

Perris C. A study of bipolar (manic–depressive) and unipolar recurrent depressive psychoses. Acta Psychiatrica Scandinavica. 1966; 194:9–14. [PubMed: 5229234]

Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature Genetics. 2011; 43:977–983. [PubMed: 21926972]

Redlich R, Almeida JJR, Grotegerd D, Opel N, Kugel H, Heindel W, Arolt V, Phillips ML, Dannlowski U. Brain morphometric biomarkers distinguishing unipolar and bipolar depression: a voxel-based morphometry-pattern classification approach. JAMA Psychiatry. 2014; 71:1222–1230. [PubMed: 25188810]

Robins, LN., Regier, DA. Introduction. In: Robins, LN., Regier, DA., editors. Psychiatric Disorders in America: the Epidemiologic Catchment Area Study. The Free Press; New York: 1991. p. 1-10.

Serretti A, Mandelli L, Lattuada E, Cusin C, Smeraldi E. Clinical and demographic features of mood disorder subtypes. Psychiatry Research. 2002; 112:195–210. [PubMed: 12450629]

Souery D, Zaninotto L, Calati R, Linotte S, Mendlewicz J, Sentissi O, Serretti A. Depression across mood disorders: review and analysis in a clinical sample. Comprehensive Psychiatry. 2012; 53:24–38. [PubMed: 21414619]

Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988; 240:1285–1293. [PubMed: 3287615]

Tondo L, Visioli C, Preti A, Baldessarini RJ. Bipolar disorders following initial depression: modeling predictive clinical factors. Journal of Affective Disorders. 2014; 167:44–49. [PubMed: 25082113]

Wehr TA, Goodwin FK. Can antidepressants cause mania and worsen the course of affective illness? American Journal of Psychiatry. 1987; 144:1403–1411. [PubMed: 3314536]

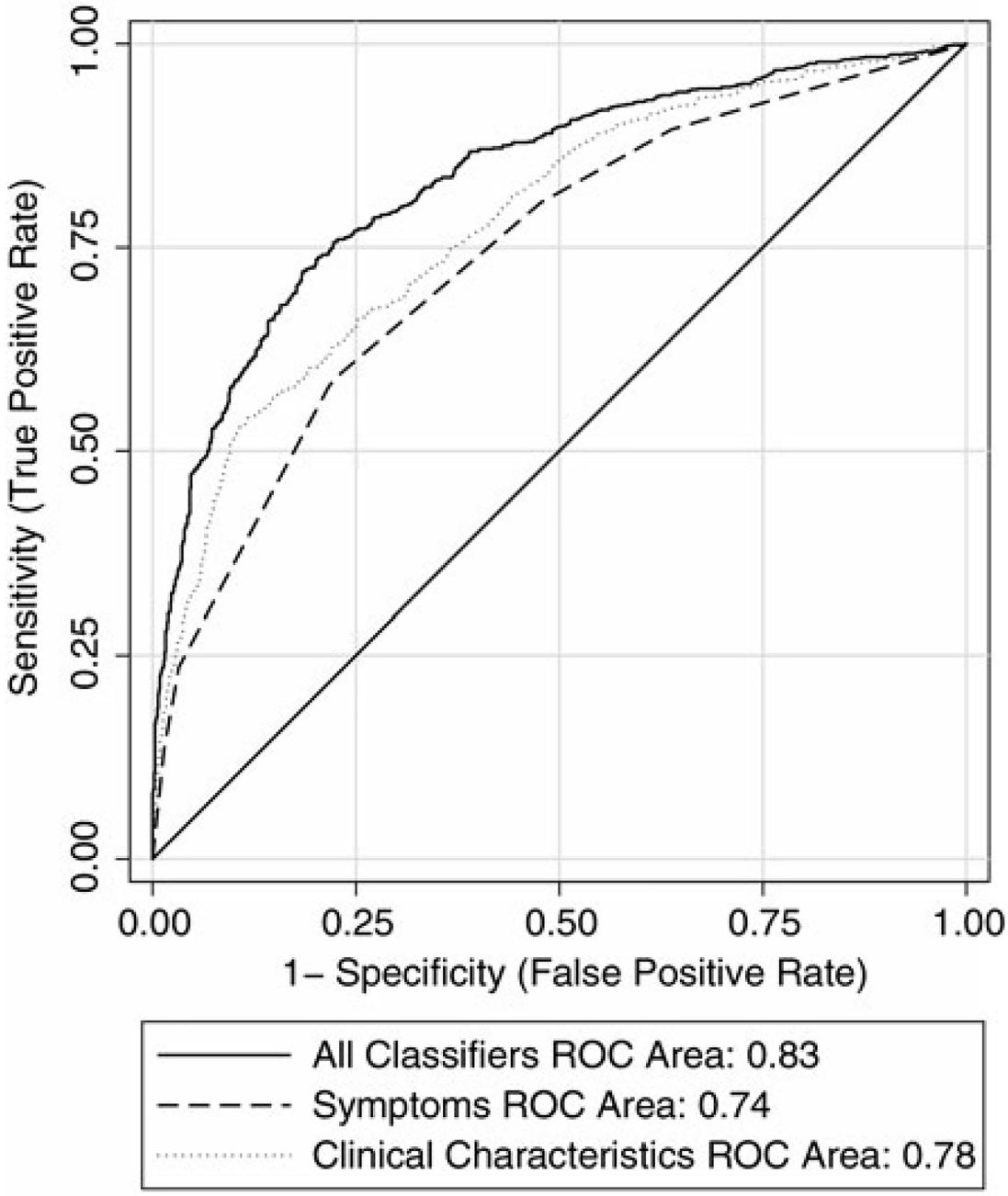


Fig. 1.
Receiver-operating characteristic (ROC) curves for seven clinical features criteria in an independent dataset.

Table 1

Sociodemographic characteristics of participants in BPI, BPII and MDD diagnosis groups encompassing both the discovery and target datasets

Characteristic	BPI (*n* = 1386)	BPII (*n* = 158)	MDD (*n* = 1684)
Mean age at interview, years (S.D.)	42.3 (12.1)	41.3 (14.4)	39.8 (12.8)
Mean age at most severe episode, years (S.D.)	30.9 (11.4)	30.8 (12.2)	29.1 (11.0)
Women, *n* (%)	886 (63.9)	104 (65.8)	1356 (80.6)
Married, *n* (%)	460 (33.2)	91 (58.0)	680 (40.4)
Disabled, *n* (%)	371 (26.8)	5 (3.2)	108 (6.4)
Mean duration of schooling, years (S.D.)	14.7 (2.6)	14.3 (3.1)	15.7 (2.8)

BP, Bipolar disorder; MDD, major depressive disorder; S.D., standard deviation.

Table 2

Prevalence of depressive symptoms during most severe episode in BPI and MDD samples in discovery dataset

Symptom	BPI, % (*n* = 386)	MDD, % (*n* = 684)	OR[a] (95% CI)	*p*
Decrease in appetite	58.0	53.2	1.35 (1.04–1.75)	0.025
Increase in appetite	18.4	22.2	0.85 (0.62–1.17)	0.321
Difficulty falling asleep	54.1	49.9	1.26 (0.97–1.65)	0.088
Early morning awakening	30.1	35.8	0.82 (0.62–1.08)	0.157
Oversleeping	56.5	49.4	1.43 (1.09–1.87)	0.009
Psychomotor agitation	46.6	36.8	1.56 (1.21–2.02)	<0.001 *
Psychomotor retardation	51.3	31.1	2.51 (1.93–3.26)	<0.001 *
Anhedonia	93.0	93.9	0.94 (0.54–1.62)	0.821
Fatigue	93.5	92.8	1.58 (0.88–2.84)	0.127
Feeling guilty	74.1	71.5	1.23 (0.92–1.65)	0.157
Feelings of worthlessness	86.0	84.2	1.31 (0.91–1.91)	0.148
Poor concentration	90.7	90.2	1.21 (0.76–1.93)	0.415
Passive death wish	71.5	64.5	1.46 (1.11–1.93)	0.007
Actually harmed self	23.3	13.3	2.11 (1.52–2.93)	<0.001 *
Worse in morning	27.8	27.6	1.05 (0.80–1.40)	0.706
Worse in evening	20.0	28.7	0.64 (0.47–0.87)	0.004
Delusions	22.0	3.8	7.72 (4.85–12.3)	<0.001 *
Hallucinations	10.4	2.5	4.75 (2.64–8.55)	<0.001 *
Incapacitated	72.5	36.0	5.00 (3.79–6.62)	<0.001 *

BPI, Bipolar disorder I; MDD, major depressive disorder; OR, odds ratio; CI, confidence interval.

[a] ORs controlled for age at interview and sex. ORs >1 indicate that a symptom is associated with greater likelihood of BPI diagnosis.

* Association *p* values that meet Bonferroni-corrected threshold of 0.0014.

Table 3

Clinical characteristics and co-morbidities of depressive episodes in discovery datasets

Feature	BPI (*n* = 386)	MDD (*n* = 684)	OR[a] (95% CI)	*p*
Clinical characteristic				
Sought professional treatment, %	81.0	80.9	1.01 (0.77–1.47)	0.718
Took medication, %	72.1	64.0	1.46 (1.15–1.92)	0.003
High after depression treatment, %	27.7	6.2	6.24 (4.20–9.28)	<0.001 *
Hospitalized for depression, %	41.7	17.4	3.40 (2.56–4.52)	<0.001 *
Mean number of lifetime depressive episodes (S.D.)	14.2 (36.5)	6.7 (11.9)	1.02 (1.01–1.03)	<0.001 *
Mean number of suicide attempts (S.D.)	1.1 (4.3)	0.4 (0.89)	1.20 (1.09–1.32)	<0.001 *
Mean duration of most severe episode, months (S.D.)	8.2 (13.5)	16.8 (31.2)	0.97 (0.95–0.98)	<0.001 *
Mean number of mixed symptoms (S.D.)	1.2 (1.9)	0.5 (1.0)	1.43 (1.30–1.58)	<0.001 *
Any mixed symptoms, %	37.1	29.6	1.45 (1.10–1.93)	0.009
Three or more mixed symptoms, %	21.0	4.4	6.05 (3.84–9.54)	<0.001 *
Co-morbidity				
Alcoholism, %	35.2	23.8	1.66 (1.25–2.19)	<0.001 *
Substance abuse, %	12.4	9.6	1.25 (0.84–1.87)	0.275
Panic disorder, %	26.9	24.4	1.20 (0.90–1.61)	0.207
Simple phobia, %	10.1	10.1	1.10 (0.73–1.68)	0.642
Social phobia, %	9.1	13.3	0.67 (0.44–1.01)	0.056
OCD, %	6.7	7.4	0.91 (0.56–1.49)	0.712
Anorexia/bulimia, %	6.2	5.8	1.22 (0.71–2.07)	0.468

BPI, Bipolar disorder I; MDD, major depressive disorder; OR, odds ratio; CI, confidence interval; S.D., standard deviation; OCD, obsessive–compulsive disorder.

[a]ORs controlled for age at interview and sex. ORs >1 indicate that a symptom is associated with greater likelihood of BPI diagnosis.

*Association *p* values that meet Bonferroni-corrected threshold of 0.0014.

Table 4

Multivariate logistic regression model for factors associated with BPI versus MDD in discovery dataset

Feature	Multivariate OR (95% CI)	*P*
High after depression treatment	5.92 (3.50–10.01)	<0.001 *
Number of suicide attempts	1.07 (0.92–1.24)	0.367
Actually harmed self	0.82 (0.44–1.50)	0.518
Hallucinations	2.24 (0.96–5.27)	0.063
Delusions	4.27 (2.16–8.45)	<0.001 *
Incapacitated	2.94 (1.87–4.60)	<0.001 *
Psychomotor agitation	1.21 (0.82–1.80)	0.332
Psychomotor retardation	1.62 (1.10–2.39)	0.015 *
Number of mixed symptoms	1.32 (1.15–1.52)	<0.001 *
Alcoholism	1.27 (0.83–1.94)	0.265
Hospitalized for depression	1.38 (0.83–2.28)	0.216
Number of lifetime depressive episodes	1.01 (1.00–1.02)	0.013 *
Most severe episode, months	0.95 (0.93–0.97)	<0.001 *

BPI, Bipolar disorder I; MDD, major depressive disorder; OR, odds ratio; CI, confidence interval.

* *p* Values meet Bonferroni threshold and were used as factors in the receiver-operating characteristic analysis.