Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders

Maria Stella Calafato, Johan H. Thygesen, Siri Ranlund, Eirini Zartaloudi, Wiepke Cahn, Benedicto Crespo-Facorro, Álvaro Díez-Revuelta, Marta Di Forti, Genetic Risk and Outcome of Psychosis (GROUP) consortium*, Mei-Hua Hall, Conrad Ivyge, Assen Jablensky, Rene Kahn, Luba Kalaydjieva, Eugenia Kravariti, Kuang Lin, Colm McDonald, Andrew M. McIntosh, Andrew McQuillin, Psychosis Endophenotypes International Consortium (PEIC)*, Marco Picchioni, Dan Rujescu, Madiha Shaikh, Timothea Toulopoulou, Jim Van Os, Evangelos Vassos, Muriel Walshe, John Powell, Cathryn M. Lewis, Robin M. Murray, Elvira Bramon and Wellcome Trust Case Control Consortium 2 (WTCCC2)*

Background
There is increasing evidence for shared genetic susceptibility between schizophrenia and bipolar disorder. Although genetic variants only convey subtle increases in risk individually, their combination into a polygenic risk score constitutes a strong disease predictor.

Aims
To investigate whether schizophrenia and bipolar disorder polygenic risk scores can distinguish people with broadly defined psychosis and their unaffected relatives from controls.

Method
Using the latest Psychiatric Genomics Consortium data, we calculated schizophrenia and bipolar disorder polygenic risk scores for 1168 people with psychosis, 552 unaffected relatives and 1472 controls.

Results
Patients with broadly defined psychosis had dramatic increases in schizophrenia and bipolar polygenic risk scores, as did their relatives, albeit to a lesser degree. However, the accuracy of predictive models was modest.

Conclusions
Although polygenic risk scores are not ready for clinical use, it is hoped that as they are refined they could help towards risk reduction advice and early interventions for psychosis.

Declaration of interest
R.M.M. has received honoraria for lectures from Janssen, Lundbeck, Lilly, Otsuka and Sunovian.

Keywords
Bipolar disorder; polygenic; prediction; psychotic disorders; polygenic risk scores; schizophrenia.

Copyright and usage
© The Royal College of Psychiatrists 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Psychotic disorders affect approximately 4% of the general population. Epidemiological and genetic studies show that they have high heritability. A Psychiatric Genomics Consortium mega-analysis of genome-wide association studies (GWAS) for schizophrenia identified more than a hundred common single nucleotide polymorphisms (SNPs) with small individual effects conferring susceptibility to the disorder. A similar mega-analysis for bipolar disorder, albeit with a more modest sample size, identified common risk variants specific to bipolar disorder and some shared with schizophrenia. Genetic epidemiology studies have shown that when compared with controls, first-degree relatives of people with schizophrenia have increased risk for bipolar disorder and first-degree relatives of people with bipolar disorder have increased risk for schizophrenia. GWAS have now provided molecular evidence for this common genetic architecture between schizophrenia and bipolar disorder. Psychotic disorders are highly polygenic with thousands of contributing common genetic variants. Although each individual variant has a very low predictive power, their combination into a polygenic risk score (PRS) represents a stronger predictor of disease. Our primary aim was to evaluate whether PRSs specific for schizophrenia or bipolar disorder, could discriminate case–control status in our sample of patients with broadly defined psychosis. Our secondary aim was to investigate whether PRSs were different in the unaffected relatives of patients with broadly defined psychotic disorder compared with controls.

Method

Sample description
Samples were collected at research centres across Europe and Australia. Our study included patients with a range of psychotic disorders (1168), unaffected relatives of patients (552) and healthy controls with no personal or family history of psychosis (1472) (Table 1). The sample presented here was included in previous GWAS seeking to identify loci for schizophrenia or psychosis. Details of sample overlap are provided in the supplement (Supplementary Data 1; available at https://doi.org/10.1192/bjp.2018.89). In order to avoid any inflation of the PRS effect size, in each analysis we included only participants that were unrelated. This was achieved by random exclusion of related participants.
Calafato et al

PRs calculation
We calculated the PRs separately for schizophrenia and for bipolar disorder in all our study participants following established methodology.4,5,28,29 Odds ratios (ORs) of allelic association tests were obtained from the most recent Psychiatric Genomics Consortium mega-analysis of GWAS for schizophrenia4 and for bipolar disorder,7 excluding all samples overlapping with the current study. For schizophrenia, the used discovery sample included 31 658 case participants and 42 022 controls, and for bipolar disorder, it included 7481 case participants and 9250 controls.4,5 In each discovery sample, SNPs were selected at nine significance thresholds ($P_r<5 \times 10^{-8}$, $1 \times 10^{-8}$, $1 \times 10^{-9}$, $1 \times 10^{-10}$, 0.01, 0.05, 0.1, 0.2, 0.5, 1). Linkage disequilibrium pruning was used to identify SNPs in linkage equilibrium with each other. The number of SNPs included at each $P$-value threshold is shown in Supplementary Table 4. In order to obtain PRs in each individual, for each SNP the number of risk alleles carried by the individual (0, 1, 2) was multiplied by the log of the OR of the allelic association test. The PRs were then calculated adding up the values obtained for each SNP.

Statistical analysis
We used logistic regression, with the first three population structure principal components and the centre of ascertainment of the samples as covariates to test whether the PRs were predictive of case–control or relative–control status in our study. The proportion of the variance explained by the PRs was calculated as Nagelkerke’s pseudo-$R^2$, by comparing a full model (PRS plus covariates) to a reference model (covariates only). The R package pROC30 was used to calculate the area under the receiver operator characteristic curve (AUC) in both the full and reference models.

In the primary analysis, the schizophrenia PRSs and the bipolar disorder PRSs were compared between 1168 case participants and 1472 controls. In the secondary analysis, we split the 1168 case participants with broadly defined psychosis into three subcategories, depending on the DSM diagnosis: schizophrenia/schizoaffective disorder, bipolar disorder and all other psychotic disorders. We then compared schizophrenia and bipolar disorder PRSs between 552 unaffected relatives and healthy controls. See Supplementary Table 5 for a breakdown of these secondary analysis subgroups. In order to divide case participants and controls into decile categories, we calculated Z-standardised PRSs, using the mean and s.d. of controls in each centre.

Results

Analysis of PRs in psychotic disorders
We calculated PRs for schizophrenia and bipolar disorder in 1472 controls and 1168 people diagnosed with a range of psychotic disorders. Density plots of schizophrenia and bipolar disorder PRs are shown in the Supplementary Fig. 2).

Using logistic regression, we found highly significant differences for both schizophrenia and bipolar disorder PRs between case participants with psychosis and controls (Table 2 and Supplementary Table 6). The difference was greater for increasingly liberal $P$-value thresholds (Table 2 and Supplementary Table 6). Compared with the bipolar disorder PRs, the schizophrenia PRs had a better ability to discriminate between case participants and controls.

The proportion of the variance in psychosis risk explained by the schizophrenia PR increased with progressively more inclusive $P$-value thresholds, reaching a plateau of 9% variance explained at the 0.05 $P$-value threshold (Nagelkerke’s pseudo-$R^2 = 9$;
Use of polygenic risk scores to identify psychotic disorders

Table 2  Comparison of schizophrenia and bipolar disorder polygenic risk scores between patients with psychotic disorders and controls

| Polygenic risk score | Polygenic risk score P-value thresholds |
|----------------------|----------------------------------------|
|                      | 5 × 10^{-08} | 1 × 10^{-06} | 0.05 | 1 |
| Schizophrenia        |              |              |      |      |
| P-value              | 1.3 × 10^{-06} | 6.8 × 10^{-21} | 7.6 × 10^{-30} | 5.7 × 10^{-40} |
| Variance explained, %| 1.1          | 4.4          | 9    | 9    |
| Bipolar disorder     |              |              |      |      |
| P-value              | 0.6          | 0.25         | 2.8 × 10^{-09} | 5.7 × 10^{-11} |
| Variance explained, %| <0.1        | <0.1         | 1.7  | 2.1  |

a. Schizophrenia polygenic risk scores and bipolar disorder polygenic risk scores were calculated using as reference, respectively, the outcome of the schizophrenia and bipolar disorder mega-analyses conducted by the Psychiatric Genomics Consortium. We then compared the scores between 1168 case participants and 1472 controls using standard logistic regression at ten different P-value thresholds (P_T 5 × 10^{-08}, 1 × 10^{-06}, 1 × 10^{-04}, 1 × 10^{-03}, 0.01, 0.05, 0.1, 0.2, 0.5, 1). Regression models included the first three ancestry-based principal components and a cohort indicator as covariates. For clarity, here we report P-values and the variance explained in disease risk as measured by Nagelkerke’s pseudo-R^2 at four P-value thresholds (P_T 5 × 10^{-08}, 1 × 10^{-06}, 1 × 10^{-04}, 0.05, 1). Results at each one of the ten different thresholds are available in Supplementary Table 6.

P = 7.6 × 10^{-40} (Table 2, Supplementary Table 6 and Fig. 1). At the same P-value threshold the variance explained by the bipolar disorder PRS was only 1.7% (P_T = 0.05, Nagelkerke’s pseudo-R^2 = 1.7%) (Table 2, Supplementary Table 6 and Fig. 1). Results for all the P-value thresholds used are reported in Fig. 1 and in the Supplementary Table 6.

Given that 68% of our case participants had a diagnosis of schizophrenia/schizoaffective disorder, to rule out the possibility that the results obtained were driven by this subgroup, we tested whether the schizophrenia and bipolar disorder PRSs were able to discriminate between case participants and controls in each of the three diagnostic subcategories included in our study (schizophrenia/schizoaffective disorder combined, bipolar disorder or other psychotic disorders). We demonstrated that even if the discriminative ability of the schizophrenia PRS was highest in the schizophrenia/schizoaffective disorder subcategory, it was also able to discriminate case participants with either bipolar disorder or other psychotic disorders from controls with highly significant group differences. At P_T = 0.05 the variance in case-control status explained by the schizophrenia PRS (Nagelkerke’s pseudo-R^2) in the bipolar disorder and other psychotic disorders subcategory was 3.4%, providing evidence that our results were not only driven by schizophrenia/schizoaffective disorder subcategory (Table 3 and Supplementary Table 7).

To evaluate the accuracy of the schizophrenia and bipolar disorder PRSs in the detection of broadly defined psychotic disorders, we calculated the AUC. For the model containing only covariates (cohort and three population structure principal components) the AUC was 0.63. Adding the schizophrenia PRS to the model increased the AUC to 0.7, whereas adding the bipolar PRS increased it to 0.65 (Supplementary Fig. 3).

We then divided our sample into deciles based on schizophrenia and bipolar disorder PRSs and calculated the ORs for affected status for each decile using as reference the central risk deciles (fifth and sixth). As expected, we observed an increase in the case-to-control ratio in progressively higher decile categories.
Similarly, the odds of having broadly defined psychosis increased progressively across PRS deciles. Compared with individuals in the central deciles (fifth and sixth), those at the tenth and highest decile had an OR for psychosis of 3.53 (95% CI 2.53–4.97) for schizophrenia PRS (Fig. 3 and Supplementary Table 9). For the bipolar PRS no difference was found between central and highest deciles (OR = 1, 95% CI 0.73–1.35) (Fig. 3 and Supplementary Table 9).

Analysis of PRSs in the unaffected relatives of people with psychosis

Given the established heritability of psychotic disorders, we evaluated whether schizophrenia and bipolar disorder PRSs could discriminate between unaffected relatives, who had never experienced any psychotic symptoms and healthy controls (Supplementary Fig. 4). Compared with controls, unaffected relatives had significantly higher PRSs both for schizophrenia ($P = 6.1 \times 10^{-39}$) and bipolar disorder ($P = 2.1 \times 10^{-7}$). Analyses at the $P$-value threshold of 0.05 are shown in Table 3 and full details are in Supplementary Table 7.

Discussion

In this study, we have shown that PRSs specific for schizophrenia or for bipolar disorder obtained from a large international cohort are also associated with broadly defined psychosis in an independent sample. Compared with controls, patients with a range of psychotic disorders have significantly higher PRSs for both schizophrenia and for bipolar disorder. The schizophrenia and bipolar disorder PRSs explained, respectively, 9 and 2% of the variance in psychosis risk, which is substantial for a single variable.

The PRS for schizophrenia had a much better performance than the PRS for bipolar disorder and this could be because of several factors. First, the schizophrenia PRS contains a more
accurate measure of genetic susceptibility, as it is derived from a much larger discovery sample than the bipolar PRS.3,5 The last Psychiatric Genomics Consortium schizophrenia meta-analysis provided evidence that increasing the size of discovery samples leads to a significant increase in the variance explained by PRS.4,8 Second, our case participants with a range of psychotic disorders included a majority of patients with schizophrenia and schizoaffective disorder (68%), which drives these performance results. However, our secondary analyses subdividing in three diagnostic categories, also showed a better performance for the schizophrenia PRS in discriminating case participants with bipolar disorder and other psychotic disorders from controls. Therefore, the use of larger discovery sample sizes appears to be the best way forward to further enhance the accuracy of PRS.8

GWAS have provided evidence for genetic overlap between schizophrenia and bipolar disorder.5,7–14 Our findings add evidence to the hypothesis of shared genetic architecture across the psychosis spectrum, supporting a continuum model for the aetiology of these disorders.31,32 The patients with bipolar disorder included in this study had type I bipolar disorder with a history of psychotic symptoms at some point in their illness. Therefore, in our sample it was not possible to make any comparison of schizophrenia and bipolar PRSs in patients with bipolar disorder with and without psychotic features. A study just published showed the existence of a gradient of schizophrenia PRSs across bipolar disorder subtypes (bipolar disorder type I with psychosis > bipolar disorder type I without psychosis > bipolar disorder type II).33

Given the heritability and familial aggregation patterns in schizophrenia and bipolar disorder, we expected unaffected relatives to have higher PRSs than the general population.34–36 In a recent study, Bigdeli et al showed that 217 healthy first-degree relatives of patients with schizophrenia and healthy controls could be distinguished by schizophrenia PRSs.36 We replicated their findings using an independent sample with 552 unaffected relatives of patients diagnosed with a wide range of psychotic disorders. Furthermore, we showed that the bipolar disorder PRS is significantly higher among healthy relatives compared with controls.

**Strengths and limitations of PRSs**

Even if the schizophrenia and bipolar PRSs can discriminate case participants from controls, their accuracy is currently modest, as indicated by the AUC of 0.7 and 0.65 for schizophrenia and bipolar disorder, respectively. The AUC is an estimate of diagnostic accuracy which equals to 0.5 when a diagnostic test is no better than chance and reaches 1 if the test could discriminate patients from controls to perfection.37,38 Typically an AUC of 0.7 is considered to have moderate discriminatory power and only when reaching 0.9 it is deemed to have high discriminatory power.39,40 For example, the models used in general practice to estimate cardiovascular disease risk and to offer preventative interventions have reached AUCs in the range of 0.74–0.85.41,42 In the case of psychotic disorders, currently the moderate accuracy precludes the use of schizophrenia and bipolar PRSs as a diagnostic or prognostic tool in clinical practice.

Current genetic findings explain only about a third of the genetic variance of these disorders. The so-called ‘missing heritability’ may reside in further common variants yet to be identified, rare mutations, copy number variants and gene–gene interactions.12 As larger samples are being collected through international efforts, additional common and rare genetic variants will be identified and the performance of PRSs is expected to improve.17,43

In the future PRSs may also incorporate socioenvironmental factors as well as gene–gene and gene–environment interactions, thus eventually enabling their use in clinical practice for risk reduction advice as it is happening in cardiovascular disease.44–52 There is growing interest in the potential of PRSs in public health campaigns to reduce environmental risks and to facilitate access to early treatment for psychosis.51 Finally, PRSs constitute a
Neale BM, Sklar P. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. Curr Opin Neurol 2015; 30: 131–8.

Cardno AG, Owen MJ. Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. Schizophrenia Bull 2014; 40: 504–15.

Harrison PJ. Recent genetic findings in schizophrenia and their therapeutic relevance. J Psychopharmacol 2015; 29: 85–96.

Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. Science 2015; 349: 1489–94.

Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013; 381: 1371–9.

Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet 2013; 9: e1003348.

Wray NR, Visscher PM. Narrowing the boundaries of the genetic architecture of schizophrenia. Schizophrenia Bull 2010; 36: 14–23.

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. Biol Psychiatry 2017; 81: 470–7.

Ripke S, O’Dushlaine C, Chamberl K, Moran J, Köhler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013; 45: 1150–9.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM-IV). American Psychiatric Association, 2000.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1999; 47: 589–93.

Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978; 35: 837–44.

Spitzer RL, Williams JB, Gibbon M, First MB. The structured clinical interview for DSM-IV-R (SCID). I: history, rationale, and description. Arch Gen Psychiatry 1992; 49: 624–9.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006; 2: 2004–9.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38: 904–9.

Euesden J, Lewis CM, O’Reilly P. PRSice: polygenic Risk Score software. Bioinformatics 2014; 31: 1464–6.

Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet 2009; 18: 3525–31.

Robin X, Tuck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12: 77.

Kendler KS, Walsh D. Schizophreniform disorder, delusional disorder and psychotic disorder not otherwise specified: clinical features, outcome and familial psychopathology. Acta Psychiatr Scand 1995; 91: 370–8.

Ivleva EI, Morris DW, Moates AF, Suppes T, Thaker GK, Tamminga CA. Genetics and intermediate phenotypes of the schizophrenia-borderline disorder boundary. Neurosci Biobehav Rev 2010; 34: 897–921.

Allardyce J, Leonenko G, Harnshire M, Pardistas AF, Forty L, Knott S, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. JAMA Psychiatry 2018; 75: 28–35.

Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. Genome Med 2014; 6: 29.

Bigdell TB, Ripke S, Bacanu SA, Lee SH, Wray NR, Geiman PV, et al. Genome-wide association study reveals greater polygenic loading for schizophrenia in cases with a family history of illness. Am J Med Genet Part B Neuropsychiatr Genet 2016; 171: 276–89.

Bigdell TB, Bacanu S-A, Webb BT, Walsh D, O’Neill FA, Fanous AH, et al. Molecular validation of the schizophrenia spectrum. Schizophrenia Bull 2014; 40: 60–5.

Zou KH, O’Malley AJ, Murre L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 2007; 115: 654–7.

Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp J Intern Med 2013; 4: 627–35.

Grzybowsk M, Younger JG. Statistical methodology: III. Receiver operating characteristic (ROC) curves. Acad Emerg Med 1997; 4: 818–26.

Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285–93.

Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem 2008; 54: 17–23.

Harrison SL, Ding J, Tang EYH, Siervo M, Robinson L, Jagger C, et al. Cardiovascular disease risk models and longitudinal changes in cognition: a systematic review. PLoS One 2014; 9: e0114431.

Purcell SM, Moran J, Fromer M, Ruderfer D, Soloviev N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014; 506: 185–90.

Pepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. European guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2016; 37: 2315–81.

Mak TSW, Khan JSH, Campbell DD, Sham PC. Local true discovery rate weighted polygenic scores using GWAS summary data. Behav Genet 2016; 46: 573–82.

Kirkbride JB. The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for gxe research. Soc Psychiatry Psychiatr Epidemiol 2014; 49: 1531–4.

Marconni A, Di Forti M, Lewis CM, Murray RM, Vassos E. Meta-analysis of the association between the level of cannabis use and risk of psychosis. Schizophrenia Bull 2016; 42: 1262–9.

Morgan C, Charalambides M, Hutchinson G, Murray RM. Migration, ethnicity, and psychosis: toward a sociodevelopmental model. Schizophrenia Bull 2010; 36: 655–64.

Van Os J, Rutten BP, Myln-Germey I, Delespaul P, Vliechbauer W, van Zelst C, et al. Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations. Schizophrenia Bull 2014; 40: 729–36.

Agerto E, Sullivan PF, Vihjalmsson BJ, Pedersen CB, Mors O, Bargum AD, et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population based study and meta-analysis. Jama Psychiatry 2015; 72: 635–41.

Mcgrath JJ, Mortensen PB, Visscher PM, Wray NR. Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. Schizophrenia Bull 2013; 39: 955–9.

Tzeng JY, Zhang D, Ponpanich M, Smith C, McCarthy M, Sale MM, et al. Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. Am J Hum Genet 2011; 89: 277–88.

Vinkhuyzen AAE, Wray NR. Novel directions for G x E analysis in psychiatry. Epidemiol Psychiatr Sci 2015; 24: 12–9.