*Psychological Medicine* (2017), **47**, 1848–1864. © Cambridge University Press 2017
doi:10.1017/S0033291717000307

ORIGINAL ARTICLE

# Characterizing cognitive heterogeneity on the schizophrenia–bipolar disorder spectrum

**T. E. Van Rheenen[1,2,3]*, K. E. Lewandowski[4,5], E. J. Tan[2,3], L. H. Ospina[6], D. Ongur[4,5], E. Neill[2,7], C. Gurvich[3], C. Pantelis[1,8,9], A. K. Malhotra[10], S. L. Rossell[2,3,7] and K. E. Burdick[6,11]**

[1] *Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Carlton, VIC, Australia;* [2] *Brain and Psychological Sciences Research Centre, Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, Hawthorn, VIC, Australia;* [3] *Cognitive Neuropsychiatry Laboratory, Monash Alfred Psychiatry Research Centre, The Alfred Hospital and Central Clinical School, Monash University, Melbourne, VIC, Australia;* [4] *Schizophrenia and Bipolar Disorder Program, McLean Hospital, Belmont, MA, USA;* [5] *Department of Psychiatry, Harvard Medical School, Boston, MA, USA;* [6] *Icahn School of Medicine, Mount Sinai, NY, USA;* [7] *Department of Psychiatry, St Vincent's Hospital, Melbourne, VIC, Australia;* [8] *Florey Institute for Neuroscience and Mental Health, Parkville, VIC, Australia;* [9] *Centre for Neural Engineering (CfNE), Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC, Australia;* [10] *Hofstra Northwell School of Medicine, Hempstead, NY, USA;* [11] *James J Peters VA Hospital, NY, USA*

**Background.** Current group-average analysis suggests quantitative but not qualitative cognitive differences between schizophrenia (SZ) and bipolar disorder (BD). There is increasing recognition that cognitive within-group heterogeneity exists in both disorders, but it remains unclear as to whether between-group comparisons of performance in cognitive subgroups emerging from within each of these nosological categories uphold group-average findings. We addressed this by identifying cognitive subgroups in large samples of SZ and BD patients independently, and comparing their cognitive profiles. The utility of a cross-diagnostic clustering approach to understanding cognitive heterogeneity in these patients was also explored.

**Method.** Hierarchical clustering analyses were conducted using cognitive data from 1541 participants (SZ $n = 564$, BD $n = 402$, healthy control $n = 575$).

**Results.** Three qualitatively and quantitatively similar clusters emerged within each clinical group: a severely impaired cluster, a mild-moderately impaired cluster and a relatively intact cognitive cluster. A cross-diagnostic clustering solution also resulted in three subgroups and was superior in reducing cognitive heterogeneity compared with disorder clustering independently.

**Conclusions.** Quantitative SZ–BD cognitive differences commonly seen using group averages did not hold when cognitive heterogeneity was factored into our sample. Members of each corresponding subgroup, irrespective of diagnosis, might be manifesting the outcome of differences in shared cognitive risk factors.

*Received 15 November 2016; Revised 23 January 2017; Accepted 24 January 2017; First published online 28 February 2017*

**Key words**: Clustering, cognition, heterogeneity, neuropsychology, psychosis spectrum.

## Introduction

Schizophrenia (SZ) and bipolar disorder (BD) are complex mental illnesses with elusive etiologies. Although nosologically distinct, these disorders are recognized to exist on a spectrum with substantial genetic, biological and phenomenological similarities (Craddock *et al.* 2006; International Schizophrenia Consortium *et al.* 2009; Rossell & Van Rheenen, 2013; Tamminga *et al.* 2013; Ruocco *et al.* 2014; Pearlson *et al.* 2016). Qualitative overlap is clearly evident in cognitive performance; patients of both disorders tend to show similar patterns of neuropsychological impairments compared with controls, and the same underlying cognitive factor structure (Schretlen *et al.* 2013; Harvey *et al.* 2014). While the magnitude of impairments tends to be larger in SZ compared with BD, recent work suggests that these effects lack diagnostic specificity (Bora *et al.* 2009*b*; Hill *et al.* 2013; Tamminga *et al.* 2014; Bora & Pantelis, 2015; Van Rheenen *et al.* 2016). Nonetheless, quantitative cognitive differences but qualitative similarities blur understandings of the extent to which the factors that contribute to cognitive impairments in SZ and BD are similar.

Traditional studies of cognition in BD and SZ have used group averages to study cognitive differences between these disorders. Although variance is considered to some extent, this approach does not fully

* Address for correspondence: T. Van Rheenen, Ph.D., Melbourne Neuropsychiatry Centre, Level 3, Alan Gilbert Building, 161 Barry Street, Carlton, VIC 3053, Australia.
(Email: tamsyn.van@unimelb.edu.au)

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

account for systematic inter-individual differences within diagnostic categories of SZ or BD, ignoring the considerable cognitive heterogeneity that exists within them (Bora *et al.* 2010). Studies clustering patients into subgroups on the basis of cognition increasingly show that some individuals with SZ have severely compromised cognition, whilst others appear to have mild to moderate impairments, and others, still, appear to be spared almost completely (Weickert *et al.* 2000; Lewandowski *et al.* 2014; Reser *et al.* 2015; Wells *et al.* 2015; TE Van Rheenen *et al.* unpublished observations). The few studies addressing cognitive heterogeneity in BD parallel these observations, suggesting that cognitive variability may be anchored by subgroups of relatively intact and significantly impaired BD patients (Green *et al.* 2013; Burdick *et al.* 2014; Bora *et al.* 2016; Jensen *et al.* 2016; Solé *et al.* 2016). Nonetheless, differences in the number of emergent subgroups and the extent of impairment across studies and diagnoses suggest that further replication is necessary.

Recently, the Bipolar–Schizophrenia Network for Intermediate Phenotypes Study has shown that differing cognitive outcomes in a cross-diagnostic sample of SZ and psychotic BD patients map onto 'cross-diagnostic biotypes' that were derived from several brain-based biomarkers (Clementz *et al.* 2016). Psychosis studies focused on cognition as the clustering variable specifically have provided preliminary evidence that membership of disparate cognitive clusters predicts different psychosocial outcomes (Burdick *et al.* 2014; Lewandowski *et al.* 2014; Wells *et al.* 2015). Together, these findings suggest diversity in the cognitive processes involved in disease phenotypes in SZ and BD, some of which are cross-diagnostic.

Empirical characterization of the structure of cognitive heterogeneity in SZ and BD is critical to our understanding of this key symptom dimension. Applying statistical clustering techniques in a large sample of patients with these disorders assessed with the same cognitive battery can facilitate our understanding of how clustering outcomes correspond between these nosologically differentiated groups. This can overcome limitations inherent in the group-average approach by addressing heterogeneity across patients explicitly. This is important given that whole-group comparisons of SZ and BD generally indicate quantitative differences in cognition that may represent manifestations of different underlying mechanisms. It is unknown whether these quantitative differences hold when cognitive heterogeneity within each disorder is explicitly considered.

Such an empirical refinement of disease phenotypes on the basis of cognition may make overlaps or disparities between SZ and BD more apparent, which could expedite the search for factors that influence different cognitive outcomes within these disorders. Such knowledge may also be informative to the formulation and refinement of approaches to treatment for BD, where cognitive remediation research is lacking but may benefit by drawing on approaches being trialed in SZ (Douglas & Van Rheenen, 2016). Cognitively speaking, the extent to which cognitive remediation approaches used in SZ can be applied to BD depends on the extent to which cognition in the disorders overlaps both qualitatively and quantitatively. For such approaches to be successful, both within- and between-group heterogeneity should be taken into account.

Here, we adopt a clustering approach in the interests of reducing variability in the broader psychiatric diagnoses of SZ and BD, by analysing the structure of cognitive heterogeneity using a transdiagnostic framework to further understand cognitive overlap between the disorders. Our study objectives were threefold and aimed to: (1) examine cognitive characteristics of emergent subgroups through cognitive cluster analysis in SZ and BD independently; (2) determine the extent of any overlaps in the profiles of the newly emergent clusters across diagnoses; and (3) examine how a cross-diagnostic approach to understanding cognitive variability was more meaningful than considering variability in BD and SZ patients separately. The degree to which emergent subgroups could be differentiated on demographic and symptom measures, and estimations of decline in cognitive functioning or accelerated aging was also explored.

## Method

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

### *Participants*

The data of 564 SZ patients, 402 BD patients and 575 healthy controls (HCs) were drawn from several independent research studies examining cognition in psychiatric illness at four sites; (1) Monash Alfred Psychiatry Research Centre, Melbourne Australia ($n$ = 288); (2) Zucker Hillside Hospital, NY, USA ($n$ = 874); (3) Icahn School of Medicine at Mount Sinai, NY, USA ($n$ = 221); (4) McLean Hospital, Belmont, MA, USA ($n$ = 158; see also online Supplementary material). Relevant hospital and university ethical review boards approved each study. Written informed consent was obtained from each participant. Psychiatric diagnoses

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

were confirmed by either the Mini-International Neuropsychiatric Interview (Sheehan *et al.* 1998) or the Structured Clinical Interview for DSM-IV (First *et al.* 1996), depending on the study through which participants were originally enrolled. Positive and negative symptomatology was assessed in all patients with either the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) or the Positive and Negative Syndrome Scale (PANSS; Kay *et al.* 1987). In the BD sample, depressive symptomatology was assessed with either the Montgomery–Åsberg Depression Rating Scale (MADRS; Montgomery & Åsberg, 1979) or the Hamilton Depression Rating Scale (HAMD; Hamilton, 1960). Manic symptomatology was assessed with either the Young Mania Rating Scale (YMRS; Young *et al.* 1978) or the Clinician Administered Rating Scale for Mania (CARS-M; Altman *et al.* 1994) in the BD patients[1]†. All participants were fluent in English, were between the ages of 18 and 65 years, and had an estimated premorbid intelligence quotient (IQ) > 70 as scored by either the Wechsler Test of Adult Reading (Holdnack, 2001), the North American Adult Reading Test (Blair & Spreen, 1989) or the word reading subtest from the Wide Range Achievement Test (WRAT; Jastak & Wilkinson, 1984). Participants with significant visual or verbal impairments, a known neurological disorder, and current substance/alcohol abuse or dependence were excluded.

### *Measures*

The US normed age- and gender-corrected T scores of the MATRICS Consensus Cognitive Battery (MCCB; Kern *et al.* 2008) were available for all participants for the following domains of cognition: speed of processing (SOP); attention/vigilance (AV); working memory (WM); visual learning (VisL); reasoning and problem solving (RPS); and social cognition (SC). The Hopkins Verbal Learning Test Revised (HVLT-R) T scores were available for a majority of participants; however, a subset of BD participants had only California Verbal Learning Test II (CVLT-II) scores available. To avoid having to exclude these participants and lose data, we merged the age- and gender-corrected T scores of HVLTR trials 1–3 and CVLT-II trials 1–5 into a single verbal learning (VerL) domain.

### *Statistical analysis*

Subsets of the current data have been published previously (Burdick *et al.* 2014; Lewandowski *et al.* 2014; Tan & Rossell, 2014; Van Rheenen & Rossell, 2014; Sperry *et al.* 2015; Van Rheenen *et al.* 2016). The online Supplementary material provides a detailed overview of the statistical analysis (including online Supplementary Table S1 and Fig. S1). Briefly, several hierarchical clustering analyses (HCA; Ward's method, squared Euclidean distance) were conducted using the cognitive domains scores, regressing out the effect of site. Analyses were conducted independently in the HC, SZ and BD groups first, then cross-diagnostically (i.e. combined SZ and BD)[2]. Once final clustering solutions were ascertained, differences in the descriptive and cognitive profiles of the newly emergent clusters were assessed. To determine if cross-diagnostic clustering was more meaningful in reducing cognitive variability than clustering within each diagnosis separately, comparisons were made between the quotient of the adjusted mean square errors for the separate BD and SZ clusters and for the cross-diagnostic sample from each cluster comparison. Next, estimated premorbid IQ[3] was assessed against a global estimate of current cognitive functioning (measured by a cognitive composite score) in a discrepancy analysis, to provide a proxy measure of stability or decline in cognition in HCs *v.* the clusters generated by the most meaningful clustering method (i.e. clustering BD or SZ independently or cross-diagnostically). The intra-individual standard deviation (is.d.) of variance on each cognitive domain (i.e. subtest dispersion) was also calculated, and the subgroup is.d. averages were compared between clusters as an indirect estimate of cognitive aging. Finally, to determine whether the cross-diagnostic cluster solution was uni- or multidimensional in its factorial structure, an exploratory factor analysis was conducted. This analysis used the maximum likelihood extraction method with a Varimax (orthogonal) rotation. The scree plot was examined and eigenvalues above 1 were considered to reference valid factors. All analyses were conducted in SPSS version 22, with *post-hoc* tests corrected at $p < 0.05$ using the Games–Howell method for unequal variance and unequal sample size.

## Results

Descriptive statistics and cognitive domain comparisons between SZ and BD participants prior to clustering are given in online Supplementary Table S2 and Fig. S2. Separate group analyses resulted in a two-cluster solution for HCs and three-cluster solutions for the patient groups.

### *HC analysis*

HC data were normally distributed. Two clusters were observed: cluster 1 (between 0.2 and 0.4 s.d.s above the

† The notes appear after the main text.

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

group mean) and cluster 2 (between 0.4 and 0.7 S.D.s below the group mean). While these clusters indicated groups with differential performance, both groups were within 1 S.D. of each other and their distinction is irrelevant in the current context. Thus, these clusters were collapsed and used as a single group comparator for the patient group analyses (Table 1).

### *SZ analysis*

There was substantial variability in cognitive profile severity of the three emergent SZ clusters, all differing from each other on all cognitive domains (Table 2). One cluster, labeled relatively intact, performed at an equivalent level to HCs on AV, VerL and RPS, with slight impairments in SOP and SC (range ±0.31 to −0.42 *v.* HCs). This cluster outperformed HCs in VisL and WM. A second cluster, labeled mild-moderate, was significantly impaired across all domains of cognition in comparison with HCs, but with effects ranging from 0.8 to 1.5 S.D.s below the HC means. The third cluster, labeled relatively severe, was also impaired across all domains relative to HCs, but with more severe performance deficits ranging from 1.2 to 2.7 S.D.s below HCs.

SZ cluster distributions are shown in Fig. 2*a*. There were no significant age differences between SZ clusters. However, the relatively intact cluster had a more even gender distribution[4], as well as lower BPRS scores compared with the relatively severe cluster[5]. The relatively intact cluster also had significantly higher pre-morbid IQ scores compared with the relatively severe and the mild-moderate clusters, which also differed from each other in terms of pre-morbid IQ (Table 1). Only scores for the latter two clusters were significantly lower than that of HCs.

### *BD analysis*

There were significant differences in cognitive profile severity of the three emergent BD clusters[6]. One cluster, labeled relatively intact, performed equivalently to HCs on the domains of WM, VerL, VisL, RPS and SC, with slightly worse performance in SOP and AV (range ±0.12 to −0.38 *v.* HCs). The second cluster, labeled mild-moderate, was significantly impaired across all cognitive domains in comparison with HCs, but with effects ranging from 0.4 to 1.3 S.D.s below the HC means. The third cluster, labeled relatively severe, showed widespread impairment across all domains relative to HCs, but with more severe performance deficits ranging from 0.9 to 2.4 S.D.s below HCs. Although there was differentiation between the BD clusters on most domains, the mild-moderate and relatively severe clusters were equivalent in performance on VisL and RPS domains and the relatively intact and mild-moderate clusters were equivalent in terms of SC.

BD cluster distributions are shown in Fig. 2*a*. No significant differences manifested between the clusters for age, gender distribution, BPRS scores, mood state, diagnostic subtype or psychosis history (Table 1). Across all cognitive domains, there were no differences between symptomatic or euthymic patients in any of the clusters (e.g. relatively intact euthymic *v.* symptomatic; mild-moderate euthymic *v.* symptomatic and relatively severe euthymic *v.* symptomatic; all $p$'s > 0.5). The relatively intact cluster had significantly higher pre-morbid IQ scores than the relatively severe and the mild-moderate clusters, who showed trending differences between themselves (Table 1). Only scores for the latter clusters were significantly reduced compared with HCs.

### *Comparison of clustered profiles between SZ and BD*

Although frequency of assignment to each cluster was significantly different by diagnosis (see online Supplementary Table S3), visual inspection of the emergent cluster profiles in SZ and BD indicated similarity in the clustering outcomes between them; similar patterns across domains and similar magnitude effects compared with HCs were evident (Fig. 1*a*). The relatively intact SZ and BD clusters did not significantly differ on any cognitive domains (all $p$'s > 0.05). The mild-moderate SZ and BD clusters did not differ on three domains (AV, VerL and RPS), with effect size differences on the other domains in the small to medium range (0.2–0.5). The relatively severe clusters had similarly sized performance deficits on four (SOP, AV, VerL and RPS) domains, with large effect size differences evident only for WM and SC. Pre-morbid IQ scores between corresponding SZ and BD clusters did not differ (e.g. SZ relatively intact *v.* BD relatively intact etc.; all $p$'s > 0.5) (Fig. 1*b*). Discriminant functions analyses (DFAs) predicting corresponding cluster membership based on diagnosis (e.g. relatively intact BD *v.* relatively intact SZ) indicated poor classification accuracy in all three analyses; BD participants were misclassified into the corresponding SZ cluster at a rate of more than 64% in analyses of both relatively severe and mild-moderate SZ/BD clusters. Only 2.7% of SZ participants were correctly classified in the relatively intact DFA (see online Supplementary Table S4).

### *Cross-diagnostic analyses*

Cross-diagnostic reanalysis of all clinical participants replicated the three-cluster solution found in the diagnostic groups independently, both in quantitative magnitude of effects and qualitative patterns across domains. A relatively intact cluster performed

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press



**Table 1.** *Descriptive summary for the HC group and clustering outcomes in BD and SZ groups*[a]

| Diagnosis/cluster | Mood state, % euthymic[b] | Gender, % male | Mean age, years (S.D.) | Mean BPRS (S.D.) | Mean pre-morbid IQ[c] (S.D.) | Diagnostic subtype, % BD I | Psychosis history, % yes |
|---|---|---|---|---|---|---|---|
| (a) HC ($n = 575$) | N.A. | 46.60 | 38.15 (15.98) | N.A. | 0.00 (1.00) | N.A. | N.A. |
| (b) SZ ($n = 564$) | | | | | | | |
| Relatively intact ($n = 75$) | N.A. | 58.67 | 42.09 (10.82) | 29.30 (6.62) | −0.01 (1.00) | N.A. | N.A. |
| Mild-moderate ($n = 262$) | N.A. | 74.04 | 42.17 (10.77) | 31.37 (7.79) | −0.65 (1.14) | N.A. | N.A. |
| Relatively severe ($n = 227$) | N.A. | 76.65 | 43.77 (9.86) | 32.40 (8.36) | −1.56 (1.35) | N.A. | N.A. |
| Group comparisons | N.A. | $\chi^2_2 = 9.51$, $p = 0.01$ | $F_{2,561} = 1.65$, $p = 0.19$ | $F_{2,522} = 4.14$, $p = 0.02$; severe global > intact/selective ($p = 0.01$, $d = -0.41$) Severe global > moderate global ($p = 0.07$, $d = 0.12$) | $F_{2,559} = 58.78$, $p < 0.001$; mild-moderate < intact/selective ($p < 0.001$, $d = 0.60$) Relatively severe < relatively intact $p < 0.001$, $d = 1.30$) Relatively severe < mild-moderate ($p < 0.001$, $d = 0.73$) | N.A. | N.A. |
| (c) BD ($n = 402$) | | | | | | | |
| Relatively intact ($n = 200$) | 46.02 | 49.00 | 40.70 (12.07) | 24.68 (5.69) | −0.04 (1.17) | 79.50 | 59.38 |
| Mild-moderate ($n = 128$) | 36.45 | 52.34 | 39.14 (11.49) | 25.68 (6.01) | −0.71 (1.23) | 85.16 | 66.67 |
| Relatively severe ($n = 74$) | 52.45 | 54.05 | 40.80 (13.99) | 25.39 (5.71) | −1.16 (1.50) | 86.49 | 68.49 |
| Group comparisons | $\chi^2_2 = 4.53$, $p = 0.10$ | $\chi^2_2 = 0.69$, $p = 0.71$ | $F_{2,398} = 0.73$, $p = 0.48$ | $F_{2,375} = 2.92$, $p = 0.06$ | $F_{2,399} = 25.81$, $p < 0.001$ Mild-moderate < relatively intact ($p < 0.001$, $d = 0.56$) Relatively severe < relatively intact ($p < 0.001$, $d = 0.83$) Relatively severe < mild-moderate (trend $p = 0.07$, $d = 0.33$) | $\chi^2_2 = 2.70$, $p = 0.26$ | $\chi^2_2 = 2.74$, $p = 0.26$ |

HC, Healthy control; BD, bipolar disorder; SZ, schizophrenia; S.D., standard deviation; BPRS, Brief Psychiatric Rating Scale; IQ, intelligence quotient; N.A., not applicable.

[a] *Post-hoc p* values are Games–Howell corrected for unequal samples sizes/unequal variances.

[b] Euthymic criteria = ⩽ 8 mania and depression rating scale (see online Supplementary material for further details).

[c] Pre-morbid IQ values represent residual scores (from which site differences have been regressed out) that are standardized against HC values (mean = 0, S.D. = 1) (see online Supplementary material for further details).

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press



**Table 2.** *Cognitive domain comparisons between HC and BD/SZ clusters*[a]

| Domain | Relatively intact: mean (S.D.) | Mild-moderate: mean (S.D.) | Severe: mean (S.D.) | Omnibus/main effect | Cluster comparisons: *post-hoc* comparisons |
|---|---|---|---|---|---|
| SZ | | | | $F_{21,3393} = 61.31$, $p < 0.001$ | |
| SOP | −0.42 (0.82) | −1.43 (0.89) | −2.69 (0.99) | $F_{3,1135} = 461.67$, $p < 0.001$ | SZ relatively intact *v.* HC $p \leqslant 0.001$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| AV | 0.03 (1.07) | −1.17 (1.00) | −2.16 (1.09) | $F_{3,1135} = 275.33$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.996$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| WM | 0.31 (0.88) | −0.81 (0.82) | −2.55 (0.94) | $F_{3,1135} = 432.90$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.030$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| VerL | −0.15 (0.79) | −0.87 (0.76) | −1.62 (0.64) | $F_{3,1135} = 208.55$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.436$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| VisL | 0.30 (0.71) | −0.82 (0.90) | −1.92 (0.79) | $F_{3,1135} = 264.53$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.009$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| RPS | 0.13 (0.88) | −0.77 (0.78) | −1.15 (0.79) | $F_{3,1135} = 112.07$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.641$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press



**Table 2** (*cont.*)

| Domain | Relatively intact: mean (S.D.) | Mild-moderate: mean (S.D.) | Severe: mean (S.D.) | Omnibus/main effect | Cluster comparisons: *post-hoc* comparisons |
|---|---|---|---|---|---|
| SC | −0.35 (0.99) | −0.95 (1.02) | −1.56 (0.92) | $F_{3,1135} = 154.42$, $p < 0.001$ | SZ relatively intact *v.* HC $p = 0.022$<br>SZ mild-moderate *v.* HC $p < 0.001$<br>SZ relatively severe *v.* HC $p < 0.001$<br>SZ relatively intact *v.* SZ mild-moderate $p < 0.001$<br>SZ relatively intact *v.* SZ relatively severe $p < 0.001$<br>SZ mild-moderate *v.* SZ relatively severe $p < 0.001$ |
| BD | | | | $F_{21,2907} = 34.62$, $p < 0.001$ | |
| SOP | −0.38 (0.88) | −1.13 (0.70) | −2.41 (0.84) | $F_{3,973} = 177.73$, $p < 0.001$ | BD relatively intact *v.* HC $p < 0.001$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p < 0.001$ |
| AV | −0.29 (1.08) | −0.90 (0.91) | −2.112 (0.91) | $F_{3,973} = 113.51$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.005$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p < 0.001$ |
| WM | 0.12 (0.87) | −1.24 (0.83) | −2.0847 (1.08) | $F_{3,973} = 159.46$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.401$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p < 0.001$ |
| VerL | −0.11 (1.07) | −0.73 (0.89) | −1.4348 (0.77) | $F_{3,973} = 59.44$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.564$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p < 0.001$ |
| VisL | 0.06 (0.90) | −1.25 (0.93) | −1.2669 (0.90) | $F_{3,973} = 93.31$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.882$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD severe $p = 0.1.00$ |

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press



| | | | | | |
|---|---|---|---|---|---|
| RPS | 0.09 (0.91) | −0.94 (0.69) | −1.1581 (0.65) | $F_{3,973} = 69.41$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.637$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p < 0.001$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p = 0.107$ |
| SC | −0.19 (0.99) | −0.41 (1.09) | −0.8993 (0.93) | $F_{3,973} = 21.45$, $p < 0.001$ | BD relatively intact *v.* HC $p = 0.079$<br>BD mild-moderate *v.* HC $p < 0.001$<br>BD relatively severe *v.* HC $p < 0.001$<br>BD relatively intact *v.* BD mild-moderate $p = 0.222$<br>BD relatively intact *v.* BD relatively severe $p < 0.001$<br>BD mild-moderate *v.* BD relatively severe $p = 0.006$ |
| BD *v.* SZ | | | | | |
| SOP | | | | $F_{5,960} = 189.47$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p > 0.999$[b]<br>SZ relatively intact *v.* BD mild-moderate $p < 0.001$<br>SZ relatively intact *v.* BD relatively severe $p < 0.001$<br>SZ mild-moderate *v.* BD relatively intact $p < 0.001$<br>SZ mild-moderate *v.* BD mild-moderate $p = 0.006$[b]<br>SZ mild-moderate *v.* BD relatively severe $p < 0.001$<br>SZ relatively severe *v.* BD relatively intact $p < 0.001$<br>SZ relatively severe *v.* BD mild-moderate $p < 0.001$<br>SZ relatively severe *v.* BD severe $p = 0.178$[b] |
| AV | | | | $F_{5,960} = 106.43$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p = 0.244$[b]<br>SZ relatively intact *v.* BD mild-moderate $p < 0.001$<br>SZ relatively intact *v.* BD relatively severe $p < 0.001$<br>SZ mild-moderate *v.* BD relatively intact $p < 0.001$<br>SZ mild-moderate *v.* BD mild-moderate $p = 0.091$[b]<br>SZ mild-moderate *v.* BD relatively severe $p < 0.001$<br>SZ relatively severe *v.* BD relatively intact $p < 0.001$<br>SZ relatively severe *v.* BD mild-moderate $p < 0.001$<br>SZ relatively severe *v.* BD relatively severe $p > 0.999$[b] |
| WM | | | | $F_{5,960} = 260.58$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p = 0.593$[b]<br>SZ relatively intact *v.* BD mild-moderate $p < 0.001$<br>SZ relatively intact *v.* BD relatively severe $p < 0.001$<br>SZ mild-moderate *v.* BD relatively intact $p < 0.001$<br>SZ mild-moderate *v.* BD mild-moderate $p < 0.001$[b]<br>SZ mild-moderate *v.* BD relatively severe $p < 0.001$<br>SZ relatively severe *v.* BD relatively intact $p < 0.001$<br>SZ relatively severe *v.* BD mild-moderate $p < 0.001$<br>SZ severe *v.* BD relatively severe $p = 0.014$[b] |

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press



**Table 2** (*cont.*)

| Domain | Relatively intact: mean (S.D.) | Mild-moderate: mean (S.D.) | Severe: mean (S.D.) | Omnibus/main effect | Cluster comparisons: *post-hoc* comparisons |
|---|---|---|---|---|---|
| VerL | | | | $F_{5,960} = 88.73$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p \geqslant 0.999$[b] |
| | | | | | SZ relatively intact *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively intact *v.* BD relatively severe $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD mild-moderate $p = 0.643$[b] |
| | | | | | SZ mild-moderate *v.* BD relatively severe $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively severe $p = 0.438$[b] |
| VisL | | | | $F_{5,960} = 147.86$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p = 0.191$[b] |
| | | | | | SZ relatively intact *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively intact *v.* BD relatively severe $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD mild-moderate $p = 0.004$[b] |
| | | | | | SZ mild-moderate *v.* BD relatively severe $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively severe $p < 0.001$[b] |
| RPS | | | | $F_{5,960} = 76.81$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p > 0.999$[b] |
| | | | | | SZ relatively intact *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively intact *v.* BD severe $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD mild-moderate $p = 0.244$[b] |
| | | | | | SZ mild-moderate *v.* BD relatively severe $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD mild-moderate $p = 0.088$ |
| | | | | | SZ relatively severe *v.* BD relatively severe $p > 0.999$[b] |
| SC | | | | $F_{5,960} = 50.266$, $p < 0.001$ | SZ relatively intact *v.* BD relatively intact $p = 0.822$[b] |
| | | | | | SZ relatively intact *v.* BD mild-moderate $p = 0.998$ |
| | | | | | SZ relatively intact *v.* BD relatively severe $p = 0.008$ |
| | | | | | SZ mild-moderate *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ mild-moderate *v.* BD mild-moderate $p < 0.001$[b] |
| | | | | | SZ mild-moderate *v.* BD relatively severe $p > 0.999$ |
| | | | | | SZ relatively severe *v.* BD relatively intact $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD mild-moderate $p < 0.001$ |
| | | | | | SZ relatively severe *v.* BD relatively severe $p < 0.001$[b] |

HC, Healthy control; BD, bipolar disorder; SZ, schizophrenia; S.D., standard deviation; SOP, speed of processing; AV, attention/vigilance; WM, working memory; VerL, verbal learning; VisL, visual learning; RPS, reasoning and problem solving; SC, social cognition.

[a] All SZ and BD patient scores were standardized (*Z* scores) against HC group means and standard deviations for all domains. All HC means = 0 and standard deviations = 1. *Post-hoc* comparisons were corrected using the Games–Howell test for unequal variances/unequal sample sizes.

[b] *Post-hoc* comparisons between corresponding SZ and BD clusters.

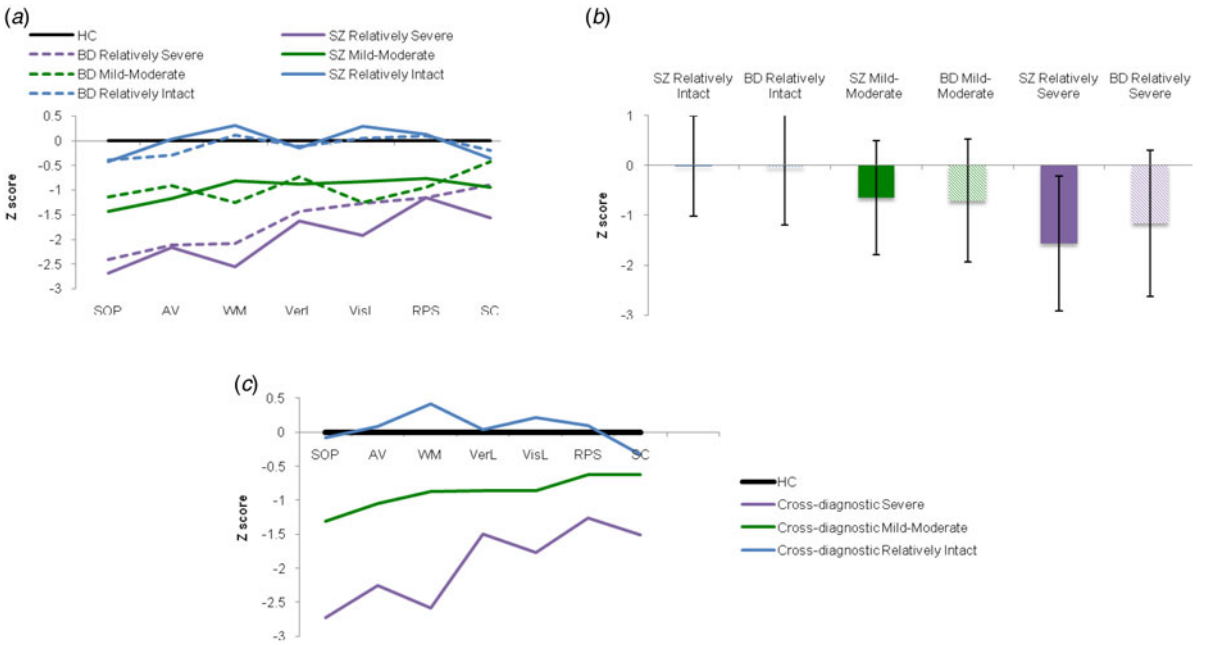



**Fig. 1.** Cognitive profile and pre-morbid intelligence quotient (IQ) comparisons of newly derived schizophrenia (SZ), bipolar disorder (BD) and cross-diagnostic clusters standardized against healthy control (HC) values (mean = 0, S.D. = 1). (*a*) Cognitive domain scores comparing SZ clusters with BD clusters. Mild-moderate SZ/BD significant differences: speed of processing (SOP) $d = 0.37$, working memory (WM) $d = 0.52$, visual learning (VisL) $d = 0.20$, social cognition (SC) $d = 0.50$. Relatively severely SZ/BD significant differences: WM $d = 0.46$, VisL $d = 0.77$, SC $d = 0.71$. (*b*) Pre-morbid IQ scores. (*c*) Cross-diagnostic clustering. AV, Attention/vigilance; VerL, verbal learning; RPS, reasoning and problem solving. Error bars represent standard deviations. For a color figure, see the online version of the paper.

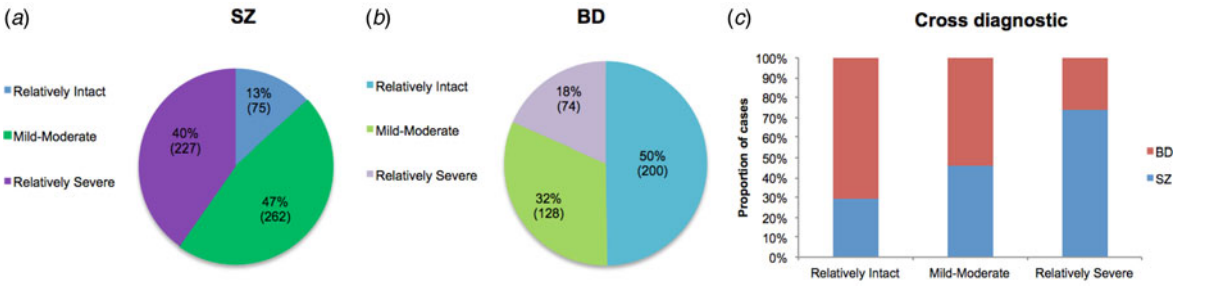



**Fig. 2.** Percentage distribution of cases falling into each cluster when classified within categorical diagnoses and a cross-diagnostic approach. (*a*) Schizophrenia (SZ) clusters. (*b*) Bipolar disorder (BD) clusters. Numbers in parentheses in (*a*) and (*b*) represent the sample number per cluster. (*c*) Cross-diagnostic clusters: the proportion of BD or SZ participants (relative to the total sample number for each cohort) falling into each cluster. For a color figure, see the online version of the paper.

equivalently to HCs on all domains except WM and VisL, where performance was slightly superior ($p <$ 0.05); and SC, where performance was slightly worse ($p < 0.05$). A mild-moderate cluster was significantly impaired across all domains of cognition in comparison with HCs, but with effects ranging from 0.6 to 1.3 S.D.s below the HC means. A relatively severe cluster also showed impairment across all domains relative to HCs, with more severe performance deficits ranging from 1.2 to 2.7 S.D.s below HCs (Fig. 1*c*).

Of the cases that fell into each of the clusters in the independent BD or SZ analyses, 81.3% fell into the corresponding cluster in the overall cross-diagnostic analysis (see online Supplementary Table S5). There was also significantly less error variation for the cross-diagnostic method than the separate BD/SZ clustering methods in the case of SOP, AV, WM and the composite score (all $p$'s =/< 0.05). There were no differences in error variation between clustering methods for the remaining cognitive domains.

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

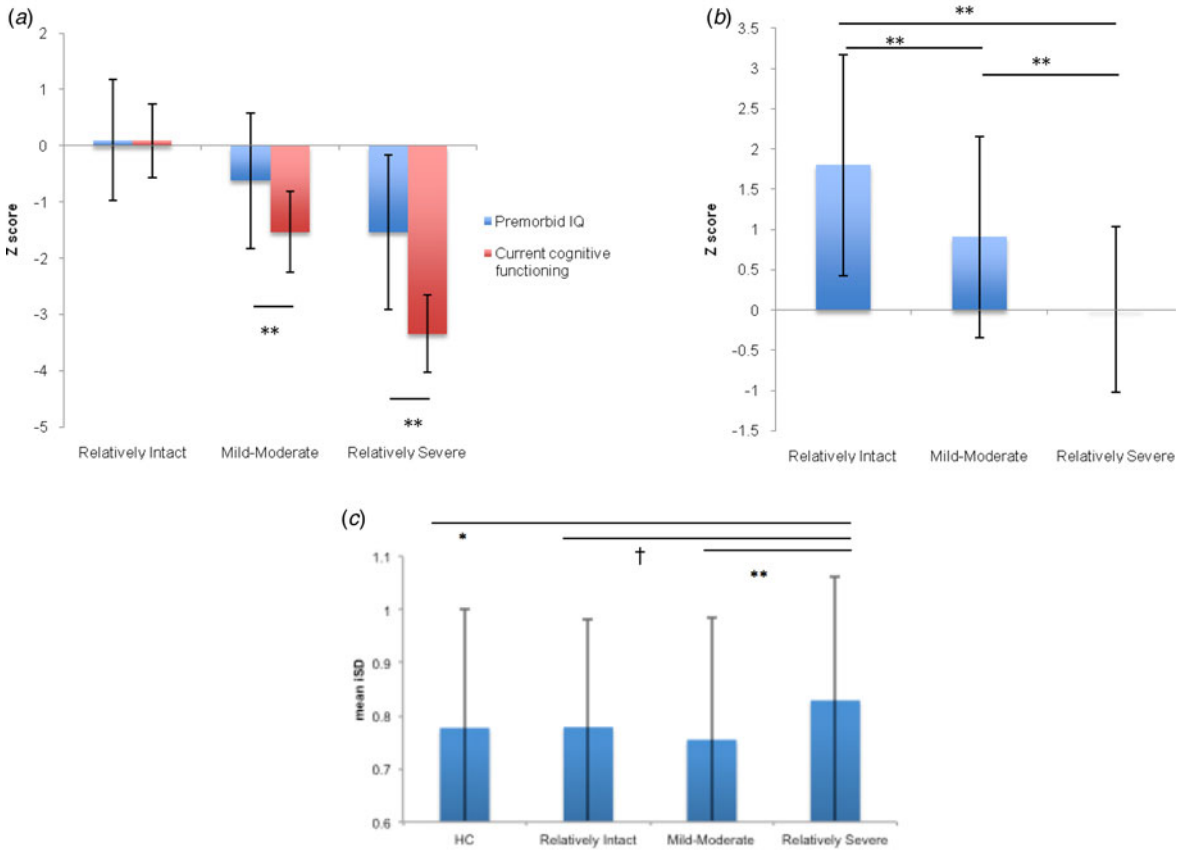



**Fig. 3.** Analysis of estimated cognitive course in the cross-diagnostic clusters. (*a*) Estimated pre-morbid intelligence quotient (IQ) *v.* current cognitive functioning [*Z* scores are standardized against healthy control (HC) mean/S.D.]. (*b*) Discrepancy score indicating estimated improvement or decline in cognition (*Z* scores are standardized against HC mean/S.D.). Larger scores indicate greater discrepancy between pre-morbid IQ and current cognitive functioning. (*c*) Dispersion (mean individual standard deviation; iS.D.) of cognitive domain scores. * $p = 0.01$, ** $p = 0.000$, † $p = 0.06$. Error bars represent standard deviations. For a color figure, see the online version of the paper.

The cross-diagnostic clusters differed significantly in terms of the distribution of the proportion of SZ/BD participants comprised within each cluster ($\chi^2_2 = 92.59$, $p < 0.001$); the relatively intact cluster contained a greater proportion of BD relative to SZ participants, whilst the mild-moderate and relatively severe clusters contained a greater proportion of SZ relative to BD participants (Fig. 2*c*). A step-down pattern was evident in terms of pre-morbid IQ: the relatively severe cluster had the lowest scores (1.5 S.D.s below HCs); the mild-moderate cluster had scores 0.6 S.D.s below HCs; and the relatively intact cluster did not differ from HCs. All pre-morbid IQ scores significantly differed between the emergent cross-diagnostic clusters.

### *Discrepancy and dispersion analysis in emergent cross-diagnostic clusters*

Given that performance variation for some domains was significantly reduced for the cross-diagnostic clustering method compared with independent diagnostic clustering, discrepancy and dispersion analysis was conducted on the emergent cross-diagnostic clusters only. The HCs and relatively intact cross-diagnostic cluster showed 'stable' performance in pre-morbid IQ *v.* current cognitive functioning ($t_{574} = 0.00$, $p < 0.999$ and $t_{192} = 0.063$, $p < 0.95$), whilst the mild-moderate and relatively severe clusters showed a 'decline', with current cognitive functioning lower than pre-morbid IQ scores ($t_{500} = 16.27$, $p < 0.001$ and $t_{269} = 21.64$, $p < 0.001$, respectively) (Fig. 3*a*). Discrepancy effects (the difference between pre-morbid IQ and current cognitive functioning) differed between cross-diagnostic clusters ($F_{3,1535} = 93.50$, $p < 0.001$); they were larger in the relatively severe compared with the mild-moderate clusters ($p < 0.001$), but scores in both groups significantly differed from the relatively intact cluster and HC group ($p < 0.001$). These latter two clusters, in turn, did not differ from each other ($p > 0.99$) (Fig. 3*b*).

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

Differences were also observed in the dispersion of variance on cognitive tests between groups ($F_{3,1537}$ = 6.42, $p < 0.001$). Dispersion, a measure of intra-individual variability, was increased in the relatively severe cross-diagnostic cluster compared with HCs ($p = 0.012$), the mild-moderate cross-diagnostic cluster ($p < 0.001$) and the relatively intact cross-diagnostic cluster (trend level $p = 0.063$). Dispersion scores did not differ between the relatively intact or mild-moderate cross-diagnostic clusters ($p = 0.549$) and neither of these clusters differed from HCs ($p = 0.375$ and $p > 0.999$, respectively) (Fig. 3*c*).

### *Exploratory cross-diagnostic factor analysis and reclassification*

Given the possibility that severity differences in the cross-diagnostic clusters could be accounted for by differences in an underlying general cognitive factor '*g*' (Fig. 1*c*), we explored whether the three-cluster cross-diagnostic solution would hold if the factorial structure of the combined patient sample was actually unidimensional. An exploratory factor analysis indicated a single-factor solution in the combined patient data explaining 40.5% of the variance and comprising all of the cognitive domains employed as the clustering variables in the previous HCAs. When the unidimensional factor score was used as a clustering variable in a subsequent HCA, the three-cluster solution was replicated with the three new clusters showing good correspondence (cross-over = 89%) to the original cross-diagnostic subgroups (see online Supplementary Figs S3 and S4, and Tables S6 and S7)[7].

## Discussion

This study examined the structure of cognitive heterogeneity on the SZ–BD spectrum in a large sample, demonstrating evidence of correspondence between the cognitive clustering outcomes across disorders. A data-driven hierarchical clustering approach indicated the presence of three cognitive subgroups in the SZ and BD groups, and in the combined patient sample. In SZ and BD patients, the three emergent clusters were differentiated in the extent to which they were impaired relative to HCs, such that each disorder encompassed three impairment subgroups: relatively severe, mild to moderate, and relatively intact. These findings align with previous work indicating that cognitive variability is anchored by spared and severely impaired subgroups of SZ or BD patients (Burdick *et al.* 2014; Lewandowski *et al.* 2014; Bora *et al.* 2016). Aside from lower symptom scores and more females in the relatively intact SZ subgroup, no other demographic factors predicted subgroup membership in either disorder. Pre-morbid IQ scores, however, were always lowest in the most impaired cognitive subgroups.

Although the numbers of subgroups for SZ and BD were identical, the frequency of assignment into each subgroup differed by diagnosis. BD patients were over-represented in the relatively intact subgroup and SZ patients over-represented in the relatively severe subgroup. This important point of differentiation between the disorders may help to explain why diagnostic comparisons using measures of central tendency are weighted toward greater magnitude deficits in SZ than BD. Nonetheless, the mere presence of some BD patients in a more severely impaired subgroup and some SZ patients in a relatively intact subgroup suggests that the traditional method of mean-level disorder comparisons is inadequate; critically, some BD patients do appear to have similar magnitudes of severe impairment as those commonly seen in SZ.

Indeed, the corresponding SZ and BD subgroups were very similar in the magnitudes of their cognitive performance for a number of domains assessed, despite differences in their sample numbers; and each paired group was poorly statistically differentiated from each other. Thus, with the exception of social cognition, the common observation of quantitative performance differences in group-wise comparisons of SZ and BD did not generally hold when cognitive heterogeneity was factored into our sample[8]. The overlap between the more homogeneous BD and SZ subgroups seen here, both qualitative and quantitative, may index common disease processes that contribute to different cognitive outcomes irrespective of differences in overt disease presentation.

Critically, our cross-diagnostic clustering analysis produced less error variance than the independent diagnostic clustering, suggesting that this method may be a useful means of studying factors contributing to differing cognitive phenotypes in these disorders. That the cross-diagnostic cluster solution was replicated using only a general cognitive factor as a clustering variable suggests that a '*g*' factor can be examined with some confidence in empirical clustering studies in even larger, 'messier' datasets that comprise different cognitive batteries. This has obvious implications for the genomic analyses of large-scale datasets focused on intra-individual differences in cognition in BD and SZ (Harvey *et al.* 2014, 2016), and studies measuring the overlap between cognition and risk for psychiatric disease more broadly (Gale *et al.* 2008; Lencz *et al.* 2014; Hatzimanolis *et al.* 2015; Hill *et al.* 2016).

Both relatively severe and mild-moderate cross-diagnostic subgroups had lower estimated pre-morbid IQ scores than HCs here. The step-down pattern between the two subgroups, however, suggests a

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

potentially stronger influence of neurodevelopmental factors in the disease course of patients with a more severe cognitive burden. Both subgroups also showed significant discrepancies in estimated pre-morbid IQ *v.* current cognitive performance, possibly reflecting a 'decline' in the course of intellectual functioning and, subsequently, neuroprogression. Although speculative, this more pronounced pattern in the relatively severe subgroup is consistent with the concept of reduced cognitive reserve (Barnett *et al.* 2006; Stern, 2009), whereby greater neurobiological liability or poorer pre-morbid intellectual neurodevelopment is thought to be associated with greater cognitive degeneration.

The intra-individual dispersion of variance across subtests was more pronounced in the relatively severe cross-diagnostic subgroup – the only one to differ significantly from HCs. Increased intra-individual variation in cognitive subtest performance is increasingly recognized as a measure of accelerated aging (MacDonald *et al.* 2006; Morgan *et al.* 2011; Kälin *et al.* 2014). Converging evidence suggests a link between such increased variation and various markers of neural integrity (MacDonald *et al.* 2009; Garrett *et al.* 2011). These neural factors may be genetically influenced (Lacritz & Cullum, 1998; Barnett *et al.* 2006; MacDonald *et al.* 2009), suggesting that the cognitive outcomes of the emergent subgroups may represent the manifestation of partially divergent etiologies that affect the neural processing systems mediating cognition. This remains speculative and future confirmatory work is necessary.

To our knowledge this is the largest clustering analysis using a validated cognitive battery in both BD and SZ individuals, and the first to explicitly examine heterogeneity when comparing these disorders on the basis of cognition. However, there are some caveats. First, we were unable to account for the effects of medication in the sample, given limited data. Second, different scales were used to assess mania/depression symptoms in the BD patients, which limited our ability to compare continuous symptom scores across subgroups. However, attempts were made to determine the influence of these variables by categorizing patients as symptomatic *v.* euthymic. Although we found no differences in the distribution of these symptom subgroups per cluster and no differences in cognitive performance between symptomatic and euthymic patients within each cluster, we cannot discount that this may have had an effect on the results. Third, it can be argued that some of the MCCB subtests may not be sufficiently sensitive to detect cognitive deficits in BD. For example, the MCCB contains only one measure of executive functioning that assesses reasoning and problem solving [Neuropsychological Assessment Battery (NAB)-Mazes]. This measure has not been widely assessed in BD and elicits smaller effects than those seen for measures of set shifting and response inhibition (Bora *et al.* 2009*a*; Burdick *et al.* 2011; Van Rheenen & Rossell, 2014). Although it is possible that this may have influenced our findings, the results of the factor analysis indicating that our subgroups are reflective of differences in an underlying '*g*' factor suggest that any potential influence is unlikely to have been substantial. Fourth, there were some VerL test scores derived from the CVLT-II. Although this had a possible influence on the results, past work has shown the total recall of words on the CVLT-II and HVLT-R to be highly correlated (Lacritz *et al.* 2001). Although this does not suggest that these tests have similar efficacy in these patient groups, our BD clustering solution was replicated when we excluded the patients without HVLT-R scores. Further, our results indicated that three clusters still emerge when using a general cognitive factor as the clustering variable. Thus, combining the CVLT-II and the HVLT-R into a single domain is likely to have had a minimal effect on the clustering outcomes in these data. Fifth, the cross-sectional design of this study limited our ability to directly assess differences between subgroups in terms of estimated decline or stability in intellectual functioning; therefore proxy measures were derived. Finally, due to limitations in the availability of clinical/demographic data across sites, we could not compare confounders such as treatment history, socio-economic status, race/ethnicity, years of education, age of onset or hospitalization and episode history, the latter of which may contribute to identifying neurodevelopmental or neuroprogressive factors that influence outcome in each of the subgroups.

One further point warrants discussion. This relates to the argument that these cognitive subgroups are not necessarily reflective of distinct 'subtypes', but rather a simple reflection of a spectrum of cognitive ability that crosses diagnostic boundaries. This argument cannot be fully addressed with the present data. However, our findings of differences in pre-morbid IQ, estimated cognitive decline and intra-individual variability do indicate that the underlying pathways leading to the different cognitive outcomes seen here are not likely to be wholly overlapping between subgroups. In more comprehensive SZ and BD datasets with biological or neuroimaging data available, the classification of individuals into more homogeneous subgroups – even within a spectrum – may therefore prove a useful means by which to determine if there are biological or clinical differences amongst individuals with these disorders. This will guide future targeted enquiries into the influential

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

factors that lead some individuals with SZ or BD to experience severe cognitive impairment, whilst others remain cognitively spared (Clementz *et al.* 2016). It will also probably have implications for the formulation of preventative strategies or treatment, where a one-size-fits-all approach is not effective for all patients.

In sum, our work supports previous findings of qualitative overlaps in the cognitive profiles of SZ and BD. However, it goes further by demonstrating the utility of considering cognitive heterogeneity when characterizing these overlaps, such that quantitative similarities also become apparent. Indeed, the current findings indicate correspondence in the profiles of more homogeneous subgroups of BD and SZ patients; while there are clear differences in the number of patients with either disorder who fall into each subgroup, the pattern and extent of cognitive ability within those subgroups is similar for most domains of general cognition. In fact in these data, the structure of heterogeneity is best portrayed cross-diagnostically, by classifying three subgroups with similar cognitive profiles that differ from each other in terms of the severity of their general cognitive performance. Such findings may assist in the search for influences on different cognitive outcomes in SZ and BD, and may help to target cognitive remediation treatments at differing levels of impairment severity irrespective of diagnosis.

## Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S0033291717000307

## Acknowledgements

This research was supported by the Australian National Health and Medical Research Council [fellowships to T.E.V.R. (APP1088785) and C.G. (APP546262) and project grant to S.L.R. (APP1060664)], the Helen McPherson Smith Trust (to T.E.V.R., S.L.R.), the Australian Rotary Health/Bipolar Expedition (to T.E.V.R., S.L.R.), the National Institute of Mental Health (MH101267; R01 MH102257, R03MH079995; K23MH077807; R01MH100125 to K.E.B.; K23MH091210 to K.E.L.; K24MH104449 to D.O.) and the Veterans Health Administration (I01CX000995 to K.E.B.).

The authors would like to thank Daniela Calafiore and Lesley Norris for their assistance with data entry.

## Declaration of Interest

K.E.B. has served on advisory boards for Sumitomo Dainippon Pharma, Takeda-Lundbeck and NeuralStem, which had no impact upon the work presented in this paper.

## Notes

[1] There were differences in the use of the PANSS and the BPRS across studies. However, given that these scales have very similar scoring formats and show good correspondence in terms of treatment response (Leucht *et al.* 2006, 2013), we created a consistent metric across participants by taking the PANSS items that are encompassed within the BPRS for those participants without BPRS scores, and merged them with the participant scores of those that had enrolled in studies that had used the BPRS. Thus, we present ratings based on the items of the BPRS in Table 1. We were unable to do the same for the differing measures of mood symptomatology in the BD patients, due to disparities in their scale formats and scoring method. Thus, we used cut-off scores on each respective scale to define BD individuals considered to meet criteria for being symptomatic (CARS-M or YMRS score >8 and/or MADRS or HAMD >8) or euthymic (CARS-M or YMRS score $\leqslant$8 and MADRS or HAMD $\leqslant$ 8) at the time of assessment.

[2] We also conducted a secondary HCA in the BD group excluding the patients who had only CVLT administered (thus were missing HVLT-R scores), to ascertain reliability of the clustering solutions. This did not change the results and for brevity is not presented.

[3] As assessed by collated, site-corrected residual scores on the Wechsler Test of Adult Reading, the North American Adult Reading Test or the word reading subtest from the WRAT.

[4] Given the difference in gender distribution across the SZ clusters, we re-ran the analyses of variance comparing cognitive performance in these clusters stratified by gender. This indicated that the better WM ($p = 0.010$) and VisL ($p = 0.002$) performance in the relatively intact SZ cluster compared with HCs was largely driven by performance by SZ males, whereas the SOP deficit was driven by SZ females ($p = 0.012$). All other group differences in performance remained.

[5] Given differences in BPRS scores between clusters, we conducted bivariate correlations (conservatively corrected for multiple testing using an $\alpha$ of $p < 0.01$) between BPRS and the standardized cognitive domain scores in each group separately. With the exception of very weak significant correlations between WM and BPRS in the relatively severe SZ cluster ($r = -0.17$, $p = 0.01$) and SC and BPRS in the mild-moderate SZ cluster ($r = -0.18$, $p = 0.004$), no other correlations reached significance.

[6] This clustering solution was replicated when we excluded the BD patients who were administered the CVLT and, thus, did not have the HVLT-R data available.

[7] This factor analysis was initially run in the two diagnostic groups independently, and produced evidence for a unifactorial model in both. Therefore, it was repeated in the cross-diagnostic sample.

[8] Differences in WM between SZ and BD clusters may be explained in part by the influence of BPRS

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

symptomatology. A significant correlation between the BPRS and this domain was observed in the SZ but not the BD relatively severe sample; the former had higher BPRS scores than BD patients.

## References


**Altman EG, Hedeker DR, Janicak PG, Peterson JL, Davis JM** (1994). The Clinician-Administered Rating Scale for Mania (CARS-M): development, reliability, and validity. *Biological Psychiatry* **36**, 124–134.

**Barnett J, Salmond C, Jones P, Sahakian B** (2006). Cognitive reserve in neuropsychiatry. *Psychological Medicine* **36**, 1053–1064.

**Blair JR, Spreen O** (1989). Predicting premorbid IQ: a revision of the National Adult Reading Test. *Clinical Neuropsychologist* **3**, 129–136.

**Bora E, Hıdıroğlu C, Özerdem A, Kaçar ÖF, Sarısoy G, Civil Arslan F, Aydemir Ö, Cubukcuoglu Tas Z, Vahip S, Atalay A, Atasoy N, Ateşci F, Tümkaya S** (2016). Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. *European Neuropsychopharmacology* **26**, 1338–1347.

**Bora E, Pantelis C** (2015). Meta-analysis of cognitive impairment in first-episode bipolar disorder: comparison with first-episode schizophrenia and healthy controls. *Schizophrenia Bulletin* **41**, 1095–1104.

**Bora E, Yucel M, Pantelis C** (2009*a*). Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders* **113**, 1–20.

**Bora E, Yucel M, Pantelis C** (2009*b*). Cognitive functioning in schizophrenia, schizoaffective disorder and affective psychoses: meta-analytic study. *British Journal of Psychiatry* **195**, 475–482.

**Bora E, Yücel M, Pantelis C** (2010). Neurocognitive markers of psychosis in bipolar disorder: a meta-analytic study. *Journal of Affective Disorders* **127**, 1–9.

**Burdick KE, Goldberg TE, Cornblatt BA, Keefe RS, Gopin CB, Derosse P, Braga RJ, Malhotra AK** (2011). The MATRICS Consensus Cognitive Battery in patients with bipolar I disorder. *Neuropsychopharmacology* **36**, 1587–1592.

**Burdick KE, Russo M, Frangou S, Mahon K, Braga RJ, Shanahan M, Malhotra AK** (2014). Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. *Psychological Medicine* **44**, 3083–3096.

**Clementz BA, Sweeney JA, Hamm JP, Ivleva EI, Ethridge LE, Pearlson GD, Keshavan MS, Tamminga CA** (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry* **173**, 373–384.

**Craddock N, O'Donovan MC, Owen MJ** (2006). Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophrenia Bulletin* **32**, 9–16.

**Douglas KM, Van Rheenen TE** (2016). Current treatment options for cognitive impairment in bipolar disorder: a review. *Current Treatment Options in Psychiatry* **3**, 330–355.

**First MB, Spitzer RL, Gibbon M, Williams JB** (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders Clinician Version (SCID-CV)*. American Psychiatric Press: Washington, DC.

**Gale CR, Deary IJ, Boyle SH, Barefoot J, Mortensen LH, Batty G** (2008). Cognitive ability in early adulthood and risk of 5 specific psychiatric disorders in middle age: the Vietnam Experience Study. *Archives of General Psychiatry* **65**, 1410–1418.

**Garrett DD, Kovacevic N, McIntosh AR, Grady CL** (2011). The importance of being variable. *Journal of Neuroscience* **31**, 4496–4503.

**Green M, Cairns M, Wu J, Dragovic M, Jablensky A, Tooney P, Scott R, Carr V** (2013). Genome-wide supported variant MIR137 and severe negative symptoms predict membership of an impaired cognitive subtype of schizophrenia. *Molecular Psychiatry* **18**, 774–780.

**Hamilton M** (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–62.

**Harvey PD, Aslan M, Du M, Zhao H, Siever LJ, Pulver A, Gaziano JM, Concato J** (2016). Factor structure of cognition and functional capacity in two studies of schizophrenia and bipolar disorder: implications for genomic studies. *Neuropsychology* **30**, 28–39.

**Harvey PD, Siever LJ, Huang GD, Muralidhar S, Zhao H, Miller P, Aslan M, Mane S, McNamara M, Gleason T, Brophy M, Przygodszki R, O'Leary TJ, Gaziano M, Concato J** (2014). The genetics of functional disability in schizophrenia and bipolar illness: methods and initial results for VA cooperative study #572. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **165B**, 381–389.

**Hatzimanolis A, Bhatnagar P, Moes A, Wang R, Roussos P, Bitsios P, Stefanis CN, Pulver AE, Arking DE, Smyrnis N, Stefanis NC, Avramopoulos D** (2015). Common genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **168B**, 392–401.

**Hill SK, Reilly JL, Keefe RSE, Gold JM, Bishop JR, Gershon ES, Tamminga CA, Pearlson GD, Keshavan MS, Sweeney JA** (2013). Neuropsychological impairments in schizophrenia and psychotic bipolar disorder: findings from the Bipolar–Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *American Journal of Psychiatry* **170**, 1275–1284.

**Hill WD, Davies G; CHARGE Cognitive Working Group, Liewald DC, McIntosh AM, Deary IJ** (2016). Age-dependent pleiotropy between general cognitive function and major psychiatric disorders. *Biological Psychiatry* **80**, 266–273.

**Holdnack JA** (2001). *Wechsler Test of Adult Reading*. The Psychological Corporation: San Antonio, TX.

**International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P** (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752.

**Jastak S, Wilkinson GS** (1984). *The Wide Range Achievement Test – Revised Administration Manual*. Jastak Associates: Wilmington, DE.

**Jensen JH, Knorr U, Vinberg M, Kessing LV, Miskowiak KW** (2016). Discrete neurocognitive subgroups in fully or


https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

partially remitted bipolar disorder: associations with functional abilities. *Journal of Affective Disorders* **205**, 378–386.

**Kälin AM, Pflüger M, Gietl AF, Riese F, Jäncke L, Nitsch RM, Hock C** (2014). Intraindividual variability across cognitive tasks as a potential marker for prodromal Alzheimer's disease. *Frontiers in Aging Neuroscience* **6**, 147.

**Kay SR, Flszbein A, Opfer LA** (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin* **13**, 261–276.

**Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JG, Keefe RSE, Mesholam-Gately R, Mintz J, Seidman LJ, Stover E, Marder SR** (2008). The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. *American Journal of Psychiatry* **165**, 214–220.

**Lacritz LH, Cullum CM** (1998). The Hopkins Verbal Learning Test and CVLT: a preliminary comparison. *Archives of Clinical Neuropsychology* **13**, 623–628.

**Lacritz LH, Cullum CM, Weiner MF, Rosenberg RN** (2001). Comparison of the Hopkins Verbal Learning Test-revised to the California Verbal Learning Test in Alzheimer's disease. *Applied Neuropsychology* **8**, 180–184.

**Lencz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM, Djurovic S, Melle I, Sundet K, Christoforou A, Reinvang I, Mukherjee S, DeRosse P, Lundervold A, Steen VM, John M, Espeseth T, Räikkönen K, Widen E, Palotie A, Eriksson JG, Giegling I, Konte B, Ikeda M, Roussos P, Giakoumaki S, Burdick KE, Payton A, Ollier W, Horan M, Donohoe G, Morris D, Corvin A, Gill M, Pendleton N, Iwata N, Darvasi A, Bitsios P, Rujescu D, Lahti J, Hellard SL, Keller MC, Andreassen OA, Deary IJ, Glahn DC, Malhotra AK** (2014). Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT). *Molecular Psychiatry* **19**, 168–174.

**Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR** (2006). Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology* **31**, 2318–2325.

**Leucht S, Rothe P, Davis J, Engel R** (2013). Equipercentile linking of the BPRS and the PANSS. *European Neuropsychopharmacology* **23**, 956–959.

**Lewandowski KE, Sperry SH, Cohen BM, Öngür D** (2014). Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. *Psychological Medicine* **44**, 3239–3248.

**MacDonald SWS, Li S-C, Bäckman L** (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychology and Aging* **24**, 792–808.

**MacDonald SWS, Nyberg L, Bäckman L** (2006). Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences* **29**, 474–480.

**Montgomery SA, Åsberg M** (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382–389.

**Morgan EE, Woods SP, Delano-Wood L, Bondi MW, Grant I; HIV Neurobehavioral Research Program (HNRP) Group** (2011). Intraindividual variability in HIV infection: evidence for greater neurocognitive dispersion in older HIV seropositive adults. *Neuropsychology* **25**, 645–654.

**Overall JE, Gorham DR** (1962). The Brief Psychiatric Rating Scale. *Psychological Reports* **10**, 799–812.

**Pearlson GD, Clementz BA, Sweeney JA, Keshavan MS, Tamminga CA** (2016). Does biology transcend the symptom-based boundaries of psychosis? *Psychiatric Clinics of North America* **39**, 165–174.

**Reser MP, Allott KA, Killackey E, Farhall J, Cotton SM** (2015). Exploring cognitive heterogeneity in first-episode psychosis: what cluster analysis can reveal. *Psychiatry Research* **229**, 819–827.

**Rossell SL, Van Rheenen TE** (2013). Theory of mind performance using a story comprehension task in bipolar mania compared to schizophrenia and healthy controls. *Cognitive Neuropsychiatry* **18**, 409–421.

**Ruocco AC, Reilly JL, Rubin LH, Daros AR, Gershon ES, Tamminga CA, Pearlson GD, Hill SK, Keshavan MS, Gur RC, Sweeney JA** (2014). Emotion recognition deficits in schizophrenia-spectrum disorders and psychotic bipolar disorder: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *Schizophrenia Research* **158**, 105–112.

**Schretlen DJ, Peña J, Aretouli E, Orue I, Cascella NG, Pearlson GD, Ojeda N** (2013). Confirmatory factor analysis reveals a latent cognitive structure common to bipolar disorder, schizophrenia, and normal controls. *Bipolar Disorders* **15**, 422–433.

**Sheehan DV, Lecrubier Y, Harnett-Sheehan K, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC** (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry* **59** (Suppl. 20), 22–33.

**Solé B, Jiménez E, Torrent C, Mar Bonnin C, Torres I, Reinares M, Priego Á, Salamero M, Colom F, Varo C, Vieta E, Martínez-Arán A** (2016). Cognitive variability in bipolar II disorder: who is cognitively impaired and who is preserved. *Bipolar Disorders* **18**, 288–299.

**Sperry SH, O'Connor LK, Öngür D, Cohen BM, Keshavan MS, Lewandowski KE** (2015). Measuring cognition in bipolar disorder with psychosis using the MATRICS Consensus Cognitive Battery. *Journal of the International Neuropsychological Society* **21**, 468–472.

**Stern Y** (2009). Cognitive reserve. *Neuropsychologia* **47**, 2015–2028.

**Tamminga CA, Ivleva EI, Keshavan MS, Pearlson GD, Clementz BA, Witte B, Morris DW, Bishop J, Thaker GK, Sweeney JA** (2013). Clinical phenotypes of psychosis in the Bipolar–Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *American Journal of Psychiatry* **170**, 1263–1274.

**Tamminga CA, Pearlson G, Keshavan M, Sweeney J, Clementz B, Thaker G** (2014). Bipolar And Schizophrenia Network For Intermediate Phenotypes: outcomes across the psychosis continuum. *Schizophrenia Bulletin* **40**, S131–S137.

**Tan EJ, Rossell SL** (2014). Building a neurocognitive profile of thought disorder in schizophrenia using a standardized test battery. *Schizophrenia Research* **152**, 242–245.

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press

**Van Rheenen TE, Bryce S, Tan EJ, Neill E, Gurvich C, Louise S, Rossell SL** (2016). Does cognitive performance map to categorical diagnoses of schizophrenia, schizoaffective disorder and bipolar disorder? A discriminant functions analysis. *Journal of Affective Disorders* **192**, 109–115.

**Van Rheenen TE, Rossell SL** (2014). An empirical evaluation of the MATRICS Consensus Cognitive Battery in bipolar disorder. *Bipolar Disorders* **16**, 318–325.

**Weickert TW, Goldberg TE, Gold JM, Bigelow LB, Egan MF, Weinberger DR** (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Archives of General Psychiatry* **57**, 907–913.

**Wells R, Swaminathan V, Sundram S, Weinberg D, Bruggemann J, Jacomb I, Cropley V, Lenroot R, Pereira AM, Zalesky A, Bousman C, Pantelis C, Weickert CS, Weickert TW** (2015). The impact of premorbid and current intellect in schizophrenia: cognitive, symptom, and functional outcomes. *NPJ Schizophrenia* **1**, 15043.

**Young R, Biggs J, Ziegler V, Meyer D** (1978). A rating scale for mania: reliability, validity and sensitivity. *British Journal of Psychiatry* **133**, 429–435.

https://doi.org/10.1017/S0033291717000307 Published online by Cambridge University Press