Characterizing cognitive heterogeneity on the schizophrenia – bipolar disorder spectrum

Running head: cognitive heterogeneity in SZ and BD

Tamsyn E. Van Rheenen*[a,b,c], PhD; Kathryn E. Lewandowski[e,f], PhD; Eric J. Tan[b,c], PhD; Luz H Ospina[g], PhD; Dost Ongur[e,f], MD, PhD; Erica Neill[b,d], PhD; Caroline Gurvich[c], DPsych; Christos Pantelis[a,j,k], MD, MRCPsych; Anil K Malhotra[h], MD; Susan L. Rossell[b,c,d], PhD and Katherine E Burdick[g,i], PhD.

[a] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, VIC, Australia
[b] Brain and Psychological Sciences Research Centre, Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, VIC, Australia
[c] Cognitive Neuropsychiatry Laboratory, Monash Alfred Psychiatry Research Centre, The Alfred Hospital and Central Clinical School, Monash University, Melbourne, Australia.
[d] Department of Psychiatry, St Vincent's Hospital, VIC, Australia
[e] Schizophrenia and Bipolar Disorder Program, McLean Hospital, Belmont, MA, USA
[f] Harvard Medical School, Department of Psychiatry, Boston, MA USA
[g] Icahn School of Medicine at Mount Sinai, NY, USA.
[h] Hofstra Northwell School of Medicine, Hempstead, NY, USA
[i] James J Peters VA Hospital, NY, USA
[j] Florey Institute for Neuroscience and Mental Health, VIC, Australia
[k] Centre for Neural Engineering (CfNE), Department of Electrical and Electronic Engineering, University of Melbourne, VIC, Australia

Address correspondence to: Dr Tamsyn Van Rheenen, Melbourne Neuropsychiatry Centre, Level 3, Alan Gilbert Building, 161 Barry St, Carlton, Vic 3053, Australia
tamsyn.van@unimelb.edu.au

Word count: 5113 (excluding abstract)

Abstract

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.

Methods: 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 to 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.

Keywords: cognition; neuropsychology; psychosis; psychosis spectrum; heterogeneity; variability; clustering; subtypes

Schizophrenia (SZ) and Bipolar Disorder (BD) are complex mental illnesses with elusive aetiologies. Although nosologically distinct, these disorders are recognized to exist on a spectrum with substantial genetic, biological and phenomenological similarities (Purcell, 2009, Craddock et al., 2006, Rossell and Van Rheenen, 2013, Ruocco et al., 2014, Tamminga et al., 2013, 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 to controls, and the same underlying cognitive factor structure (Harvey et al., 2014, Schretlen et al., 2013). While the magnitude of impairments tends to be larger in SZ compared to BD, recent work suggests that these effects lack diagnostic specificity (Van Rheenen et al., 2016, Bora et al., 2009b, Bora and Pantelis, 2015, Hill et al., 2013, Tamminga et al., 2014). 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 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, Wells et al., 2015, Reser et al., 2015, Van Rheenen et al., Under review). The few studies addressing cognitive heterogeneity in BD parallel these observations, suggesting that cognitive variability may be anchored by sub groups of relatively intact *and* significantly impaired BD patients (Bora et al., 2016, Burdick et al., 2014, Green et al., 2013, Solé et al., 2016, Jensen 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 (BSNIP) 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, Wells et al., 2015, Lewandowski et al., 2014). 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 and 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 analyzing the structure of cognitive heterogeneity using a trans-diagnostic framework to further understand cognitive overlap between the disorders. Our study objectives were threefold and aimed to; 1) exam 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 ageing was also explored.

Material and Methods

The authors assert that 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, New York, USA (n=874); 3) Icahn School of Medicine at Mount Sinai, New York, USA (n=221); 4) McLean Hospital, Belmont, MA, USA (n=158; see also supplementary material). Relevant Hospital and University ethical review boards approved each study. Written informed consent was obtained from each participant. Psychiatric diagnoses 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 and 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 Asberg Depression Rating Scale ((MADRS; Montgomery and Asberg, 1979) or the Hamilton Depression Rating Scale (HRSD; Hamilton, 1960). Manic symptomatology was assessed with either the Young Mania Rating Scale (YMRS; Young et al., 1978) or the Clinician Administered Rating Scale (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 pre-morbid IQ >70 as scored by either the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001), the North American Adult Reading Test (NAART; Blair and Spreen, 1989) or the word reading subtest from the Wide Range Achievement Test

[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., 2013, Leucht et al., 2006), 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 HRSD >8) or euthymic (CARS-M or YMRS score $\leq$ 8 and MADRS or HRSD $\leq$ 8) at the time of assessment.

(WRAT; Jastak and Wilkinson, 1984). Participants with significant visual or verbal impairments, a known neurological disorder, and current substance/alcohol abuse or dependence were excluded.

Measures

The U.S normed age- and gender-corrected T scores of the MATRICS Consensus Cognitive Battery(Kern et al., 2008) were available for all participants for the following domains of cognition; Processing Speed (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 trial 1-3 and CVLT-II trial1-5 into a single Verbal Learning (VerL) domain.

Statistical analysis

Subsets of the current data have been published previously (Lewandowski et al., 2014, Burdick et al., 2014, Van Rheenen et al., 2016, Van Rheenen and Rossell, 2014, Sperry et al., 2015, Tan and Rossell, 2014). The supplementary material provides a detailed overview of the statistical analysis (including Table S1 and Figure 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 versus the clusters generated by the most meaningful clustering method (i.e., clustering BD or SZ independently or cross-diagnostically). The intra-individual standard deviation (iSD) of variance on each cognitive domain (i.e., subtest dispersion) was also calculated, and the subgroup iSD averages were compared between clusters as an indirect estimate of cognitive ageing. Finally, to determine whether the cross-diagnostic cluster solution was uni or multi-dimensional 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 $p<.05$ 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 Supplementary Table S2 and Figure S2. Separate

[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 Wide Range Achievement Test.

group analyses resulted in a two-cluster solution for HCs and three-cluster solutions for the patient groups.

HC analysis

HC data was normally distributed. Two clusters were observed: Cluster 1 (between 0.2 and 0.4 SD's above the group mean) and Cluster 2 (between 0.4 and 0.7 below the group mean). While these clusters indicated groups with differential performance, both groups were within 1 SD 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.

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. One cluster labelled *Relatively Intact,* performed at an equivalent level to HCs on AV, VerL and RPS, with slight impairments in SOP and SC (range $\pm$ 0.31 to -0.42 vs. HCs). This cluster outperformed HCs in VisL and WM. A second cluster labelled *Mild-Moderate,* was significantly impaired across all domains of cognition in comparison to HCs, but with effects ranging from 0.8-1.5 SD's below the HC means. The third cluster labelled *Relatively Severe,* was also impaired across all domains relative to HCs, but with more severe performance deficits ranging from 1.2-2.7 SD's below HCs.

SZ cluster distributions are shown in Figure 2a. 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 Brief Psychiatric Rating Scale (BPRS) scores compared

[4] Given the difference in gender distribution across the SZ clusters, we re-ran the ANOVAs comparing cognitive performance in these clusters stratified by gender. This indicated that the better WM (p=.010) and VisL (p=.002) performance in the Relatively Intact SZ cluster compared to HCs was largely driven by performance

to the Relatively Severe cluster[5]. The Relatively Intact cluster also had significantly higher premorbid IQ scores compared to the Relatively Severe and the Mild-Moderate clusters, which also differed from each other in terms of premorbid 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, labelled *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 $\pm$ 0.12 to -0.38 vs. HC's). The second cluster, labelled *Mild-Moderate,* was significantly impaired across all cognitive domains in comparison to HCs, but with effects ranging from 0.4-1.3 SD's below the HC means. The third cluster labelled *Relatively Severe,* showed widespread impairment across all domains relative to HC's, but with more severe performance deficits ranging from 0.9-2.4 SD'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 Figure 2a. 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.,

---

by SZ males, whereas the SOP deficit was driven by SZ females (p=.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 $p<.01$) 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 ($r = -.17$, $p=.01$) and SC and BPRS in the Mild-Moderate SZ cluster ($r=-.18$, $p=.004$), no other correlations reached significance.

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

Relatively Intact euthymic vs symptomatic; Mild-Moderate euthymic vs symptomatic and Relatively Severe euthymic vs symptomatic; all p>'s 0.5). The Relatively Intact cluster had significantly higher premorbid 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 to HCs.

Comparison of clustered profiles between SZ and BD.

Although frequency of assignment to each cluster was significantly different by diagnosis (Supplemental 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 to HCs were evident (Figure 1a). The Relatively Intact SZ and BD clusters did not significantly differ on any cognitive domains (all p's >.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-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. Premorbid IQ scores between corresponding SZ and BD clusters did not differ (e.g., SZ Relatively Intact vs. BD Relatively Intact etc.; all p>'s 0.5) (Figure 1b). DFAs predicting corresponding cluster membership based on diagnosis (e.g., Relatively Intact BD vs. 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 (Supplemental Table S4).

Cross-diagnostic analyses

Cross-diagnostic re-analysis 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 equivalently to HCs on all domains except WM and VisL; where performance was slightly superior ($p<.05$), and SC; where performance was slightly worse ($p<.05$). A *Mild-Moderate* cluster was significantly impaired across all domains of cognition in comparison to HCs, but with effects ranging from 0.6-1.3 SD's below the HC means. A *Relatively Severe* cluster also showed impairment across all domains relative to HC's, with more severe performance deficits ranging from 1.2-2.7 SD's below HCs (Figure 1c).

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 (Supplemental 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 =/< .05). There were no differences in error variation between clustering methods for the remaining cognitive domains.

The cross-diagnostic clusters differed significantly in terms of the distribution of the proportion of SZ/BD participants comprised within each cluster ($\chi^2(2)=92.59, p<.001$); 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 (Figure 2c). A step-down pattern was evident in terms of premorbid IQ: the Relatively Severe cluster had the lowest scores (1.5 SDs below HCs); the Mild-Moderate cluster had scores 0.6 SD's below HCs; and the Relatively Intact cluster did not differ from HCs. All premorbid 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 to 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 premorbid IQ versus current cognitive functioning ($t(574)=.00$, $p<.999$ and $t(192)=.063, p<.95$), whilst the Mild-Moderate and Relatively Severe clusters showed a 'decline,' with current cognitive functioning lower than premorbid IQ scores ($t(500)=16.27, p<.001$ and $t(269)=21.64, p<.001$ respectively) (Figure 3a). Discrepancy effects (the difference between premorbid IQ and current cognitive functioning) differed between cross-diagnostic clusters ($F(3, 1535)=93.50, p<.001$); they were larger in the Relatively Severe compared to the Mild-Moderate clusters ($p<.001$), but scores in both groups significantly differed from the Relatively Intact cluster and HC group ($p<.001$). These latter two clusters in turn, did not differ from each other ($p>.99$) (Figure 3b).

Differences were also observed in the dispersion of variance on cognitive tests between groups ($F(3, 1537)=6.42, p<.001$). Dispersion, a measure of intra-individual variability, was increased in the Relatively Severe cross-diagnostic cluster compared to HCs ($p=.012$), the Mild-Moderate cross-diagnostic cluster ($p<.001$) and the Relatively Intact cross-diagnostic cluster (trend level $p=.063$). Dispersion scores did not differ between the Relatively Intact or Mild-Moderate cross-diagnostic clusters ($p=.549$) and neither of these clusters differed from HCs ($p=.375$ and $p>.999$ respectively) (Figure 3c).

Exploratory cross-diagnostic factor analysis and re-classification.

Given the possibility that severity differences in the cross-diagnostic clusters could be accounted for by differences in an underlying general cognitive factor '*g*' (Figure 1c), we

explored whether the three-cluster cross-diagnostic solution would hold if the factorial structure of the combined patient sample was actually uni-dimensional. 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 Supplementary Figures S3 and S4 and Tables S6-7)[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 - 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 (Lewandowski et al., 2014, Bora et al., 2016, Burdick et al., 2014). Aside from lower symptom scores and more females in the Relatively Intact SZ subgroup, no other demographic factors predicted subgroup membership in either disorder. Premorbid IQ scores however, were always lowest in the most impaired cognitive subgroups.

[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.

Although the number of subgroups for SZ and BD were identical, the frequency of assignment into each subgroup differed by diagnosis. BD patients were overrepresented in the Relatively Intact subgroup and SZ patients overrepresented 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 didn't 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

[8] Differences in WM between SZ and BD clusters may be explained in part, by the influence of BPRS symptomatology. A significant correlation between 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.

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., 2016, Harvey et al., 2014), and studies measuring the overlap between cognition and risk for psychiatric disease more broadly (Lencz et al., 2014, Hatzimanolis et al., 2015, Hill et al., 2016, Gale et al., 2008).

Both Relatively Severe and Mild-Moderate *cross-diagnostic subgroups* had lower estimated premorbid IQ scores than HCs here. The step-down pattern between the two subgroups however, suggests a 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 premorbid IQ vs. 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 premorbid 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 ageing(Morgan et al., 2011, Kälin et al., 2014, MacDonald et al., 2006). Converging evidence suggests a link between such increased variation and various markers of neural integrity(Garrett et al., 2011, MacDonald et al., 2009). These neural factors may be genetically influenced(MacDonald et al., 2009, Lacritz and Cullum, 1998, Barnett et al., 2006), 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 vs. 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 (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., 2009a, Van Rheenen and Rossell, 2014, Burdick et al., 2011). 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 this 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, socioeconomic 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 premorbid 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 homogenous 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 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 likely 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 homogenous 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.

Acknowledgments

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

Financial Support

This research was supported by the Australian National Health and Medical Research Council (Fellowships to TVR (APP1088785) and CG (APP546262) and Project Grant to SR (APP1060664)), the Helen McPherson Smith Trust (to TVR, SR), the Australian Rotary Health/Bipolar Expedition (to TVR, SR), the National Institute of Mental Health (MH101267; R01 MH102257, R03MH079995; K23MH077807; R01MH100125 to KEB; K23MH091210 to KEL; K24MH104449 to DO) and the Veterans Health Administration (I01CX000995 to KEB).

Declarations of Interest

Dr. Burdick has served on advisory boards for Sumitomo Dainippon Pharma, Takeda-Lundbeck, and NeuralStem, which had no impact upon the work presented in this manuscript.

References

ALTMAN, E. G., HEDEKER, D. R., JANICAK, P. G., PETERSON, J. L. & DAVIS, J. M. 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, J. R. & SPREEN, O. 1989. Predicting premorbid IQ: a revision of the National Adult Reading Test. *The Clinical Neuropsychologist,* 3**,** 129-136.

BORA, E., HIDIROĞLU, C., ÖZERDEM, A., KAÇAR, Ö. F., SARISOY, G., ARSLAN, F. C., AYDEMIR, Ö., TAS, Z. C., VAHIP, S. & ATALAY, A. 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. *European Neuropsychopharmacology*.

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*.

BORA, E., YUCEL, M. & PANTELIS, C. 2009a. 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. 2009b. Cognitive functioning in schizophrenia, schizoaffective disorder and affective psychoses: meta-analytic study. *The 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, K. E., GOLDBERG, T. E., CORNBLATT, B. A., KEEFE, R. S., GOPIN, C. B., DEROSSE, P., BRAGA, R. J. & MALHOTRA, A. K. 2011. The MATRICS Consensus Cognitive Battery in Patients with Bipolar I Disorder. *Neuropsychopharmacology,* 36**,** 1587-1592.

BURDICK, K. E., RUSSO, M., FRANGOU, S., MAHON, K., BRAGA, R. J., SHANAHAN, M. & MALHOTRA, A. K. 2014. Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. *Psychological Medicine,* 44**,** 3083-96.

CLEMENTZ, B. A., SWEENEY, J. A., HAMM, J. P., IVLEVA, E. I., ETHRIDGE, L. E., PEARLSON, G. D., KESHAVAN, M. S. & TAMMINGA, C. A. 2016. Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *The American Journal Of Psychiatry,* 173**,** 373-384.

CRADDOCK, N., O'DONOVAN, M. C. & OWEN, M. J. 2006. Genes for Schizophrenia and Bipolar Disorder? Implications for Psychiatric Nosology. *Schizophrenia Bulletin,* 32**,** 9-16.

DOUGLAS, K. M. & VAN RHEENEN, T. E. 2016. Current Treatment Options for Cognitive Impairment in Bipolar Disorder: a Review. *Current Treatment Options in Psychiatry***,** 1-26.

FIRST, M. B., SPITZER, R. L., GIBBON, M. & WILLIAMS, J. B. 1996. Structured Clinical Interview for DSM-IV Axis I Disorders Clinician Version (SCID-CV). Washington, D.C: American Psychiatric Press.

GALE, C. R., DEARY, I. J., BOYLE, S. H., BAREFOOT, J., MORTENSEN, L. H. & 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, D. D., KOVACEVIC, N., MCINTOSH, A. R. & GRADY, C. L. 2011. The importance of being variable. *The 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.

HARVEY, P. D., ASLAN, M., DU, M., ZHAO, H., SIEVER, L. J., PULVER, A., GAZIANO, J. M. & 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.

HARVEY, P. D., SIEVER, L. J., HUANG, G. D., MURALIDHAR, S., ZHAO, H., MILLER, P., ASLAN, M., MANE, S., MCNAMARA, M., GLEASON, T., BROPHY, M., PRZYGODSZKI, R., O'LEARY, T. J., 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,* 165**,** 381-389.

HATZIMANOLIS, A., BHATNAGAR, P., MOES, A., WANG, R., ROUSSOS, P., BITSIOS, P., STEFANIS, C. N., PULVER, A. E., ARKING, D. E. & SMYRNIS, N. 2015. Common genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 168**,** 392-401.

HILL, S. K., REILLY, J. L., KEEFE, R. S. E., GOLD, J. M., BISHOP, J. R., GERSHON, E. S., TAMMINGA, C. A., PEARLSON, G. D., KESHAVAN, M. S. & SWEENEY, J. A. 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, W. D., DAVIES, G., THE, C. C. W. G., LIEWALD, D. C., MCINTOSH, A. M. & DEARY, I. J. 2016. Age-Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders. *Biological Psychiatry,* 80**,** 266-273.

HOLDNACK, J. A. 2001. *Wechsler Test of Adult Reading,* San Antonio, Texas, The Psychological Corporation.

JASTAK, S. & WILKINSON, G. S. 1984. *The wide range achievement test-revised: administration manual*, Jastak Assessment Systems.

JENSEN, J. H., KNORR, U., VINBERG, M., KESSING, L. V. & MISKOWIAK, K. W. 2016. Discrete neurocognitive subgroups in fully or partially remitted bipolar disorder: Associations with functional abilities. *Journal of Affective Disorders,* 205**,** 378-386.

KÄLIN, A. M., PFLÜGER, M., GIETL, A. F., RIESE, F., JÄNCKE, L., NITSCH, R. M. & HOCK, C. 2014. Intraindividual variability across cognitive tasks as a potential marker for prodromal Alzheimer's disease. *Frontiers in Aging Neuroscience,* 6.

KAY, S. R., FLSZBEIN, A. & OPFER, L. A. 1987. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin,* 13**,** 261-276.

KERN, R. S., NUECHTERLEIN, K. H., GREEN, M. F., BAADE, L. E., FENTON, W. S., GOLD, J. G., KEEFE, R. S. E., MESHOLAM-GATELY, R., MINTZ, J., SEIDMAN, L. J., STOVER, E. & MARDER, S. R. 2008. The MATRICS Consensus Cognitive

Battery, Part 2: Co-Norming and Standardization. . *American Journal of Psychiatry,* 165**,** 214-220.
LACRITZ, L. H. & CULLUM, C. M. 1998. The hopkins verbal learning test and CVLT: a preliminary comparison. *Archives of Clinical Neuropsychology,* 13**,** 623-628.
LACRITZ, L. H., CULLUM, C. M., WEINER, M. F. & ROSENBERG, R. N. 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, D. C., STARR, J. M., DJUROVIC, S., MELLE, I., SUNDET, K. & CHRISTOFOROU, A. 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, J. M., ETSCHEL, E., KISSLING, W., HAMANN, J. & ENGEL, R. R. 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, K. E., SPERRY, S. H., COHEN, B. M. & ÖNGÜR, D. 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. *Psychological Medicine,* 44**,** 3239-3248.
MACDONALD, S. W. S., LI, S.-C. & BÄCKMAN, L. 2009. Neural underpinnings of within-person variability in cognitive functioning. *Psychology And Aging,* 24**,** 792-808.
MACDONALD, S. W. S., 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, S. A. & ASBERG, M. 1979. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry,* 134**,** 382-389.
MORGAN, E. E., WOODS, S. P., DELANO-WOOD, L., BONDI, M. W. & GRANT, I. 2011. Intraindividual variability in HIV infection: Evidence for greater neurocognitive dispersion in older HIV seropositive adults. *Neuropsychology,* 25**,** 645.
OVERALL, J. E. & GORHAM, D. R. 1962. The brief psychiatric rating scale. *Psychological reports,* 10**,** 799-812.
PEARLSON, G. D., CLEMENTZ, B. A., SWEENEY, J. A., KESHAVAN, M. S. & TAMMINGA, C. A. 2016. Does Biology Transcend the Symptom-based Boundaries of Psychosis? *Psychiatric Clinics of North America,* 39**,** 165-174.
PURCELL, S. M. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature,* 460**,** 748.
RESER, M. P., ALLOTT, K. A., KILLACKEY, E., FARHALL, J. & COTTON, S. M. 2015. Exploring cognitive heterogeneity in first-episode psychosis: What cluster analysis can reveal. *Psychiatry Research,* 229**,** 819-827.
ROSSELL, S. L. & VAN RHEENEN, T. E. 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, A. C., REILLY, J. L., RUBIN, L. H., DAROS, A. R., GERSHON, E. S., TAMMINGA, C. A., PEARLSON, G. D., HILL, S. K., KESHAVAN, M. S., GUR, R. C. & SWEENEY, J. A. 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*.

SCHRETLEN, D. J., PEÑA, J., ARETOULI, E., ORUE, I., CASCELLA, N. G., PEARLSON, G. D. & 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, D. V., LECRUBIER, Y., HARNETT SHEEHAN, K., AMORIM, P., JANAVS, J., WEILLER, E., HERGUETA, T., BAKER, R. & DUNBAR, G. C. 1998. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry,* 59**,** 22-33.

SOLÉ, B., JIMÉNEZ, E., TORRENT, C., MAR BONNIN, C., TORRES, I., REINARES, M., PRIEGO, Á., SALAMERO, M., COLOM, F. & VARO, C. 2016. Cognitive variability in bipolar II disorder: who is cognitively impaired and who is preserved. *Bipolar disorders,* 18**,** 288-299.

SPERRY, S. H., O'CONNOR, L. K., ÖNGÜR, D., COHEN, B. M., KESHAVAN, M. S. & LEWANDOWSKI, K. E. 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, C. A., IVLEVA, E. I., KESHAVAN, M. S., PEARLSON, G. D., CLEMENTZ, B. A., WITTE, B., MORRIS, D. W., BISHOP, J., THAKER, G. K. & SWEENEY, J. A. 2013. Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *American Journal of Psychiatry*.

TAMMINGA, C. A., 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, E. J. & ROSSELL, S. L. 2014. Building a neurocognitive profile of thought disorder in schizophrenia using a standardized test battery. *Schizophrenia Research,* 152**,** 242-245.

VAN RHEENEN, T. E., BRYCE, S., TAN, E. J., NEILL, E., GURVICH, C., LOUISE, S. & ROSSELL, S. L. 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, T. E., CROPLEY V, WELLS R, BRUGGEMANN J, SWAMINATHAN V, SUNDRAM S, WEINBERG W, JACOMB I, LENROOT R, PEREIRA AM, ZALESKY A, BOUSMAN C, SHANNON WEICKERT C, WEICKERT TW & P, P. Under review. Widespread volumetric reductions in schizophrenia and schizoaffective patients displaying compromised cognitive abilities.

VAN RHEENEN, T. E. & ROSSELL, S. L. 2014. An Empirical Evaluation of the MATRICS Consensus Cognitive Battery in Bipolar Disorder. *Bipolar Disorders,* 16**,** 318-325.

WEICKERT, T. W., GOLDBERG, T. E., GOLD, J. M., BIGELOW, L. B., EGAN, M. F. & WEINBERGER, D. R. 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, A. M. & ZALESKY, A. 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. *The British Journal of Psychiatry,* 133**,** 429-435.

Table 1. Descriptive summary for HC group and clustering outcomes in BD and SZ groups.

Diagnosis/Cluster	**Mood state**	***Gender***	***Age***		***BPRS***		***Premorbid IQ****		***Diagnostic subtype***	***Psychosis History***
a) HC (n = 575)	***% Euthymic***	***% Male***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***% BD I***	***%Yes***
	NA	46.60	38.15	15.98	NA	NA	.00	1.00	NA	NA
b) SZ (n=564)	***% Euthymic***	***% Male***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***% BD I***	***%Yes***
Relatively Intact (n=75)	NA	58.67	42.09	10.82	29.30	6.62	-.01	1.00	NA	NA
Mild-Moderate (n= 262)	NA	74.04	42.17	10.77	31.37	7.79	-.65	1.14	NA	NA
Relatively Severe (n=227)	NA	76.65	43.77	9.86	32.40	8.36	-1.56	1.35	NA	NA
Group comparisons	NA	$\chi^2(2)=9.51$, p=.01	F(2,561) =1.65, p=.19		F(2,522) =4.14, p=.02; Severe Global >Intact/Selective (p=.01, d= -0.41) Severe Global >Moderate Global (p=.07, d= 0.12)		F(2,559) =58.78, p<.001; Mild-Moderate <Intact/Selective (p<.001, d=0.60) Relatively Severe < Relatively Intact p<.001, d =1.30) Relatively Severe < Mild-Moderate (p<.001, d=0.73)		NA	NA
c) BD (n=402)	***% Euthymic***	***% Male***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***% BD I***	***%Yes***
Relatively Intact (n=200)	46.02	49.00	40.70	12.07	24.68	5.69	-.04	1.17	79.50	59.38
Mild-Moderate (n=128)	36.45	52.34	39.14	11.49	25.68	6.01	-.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	χ^2 (2)=4.53, p=.10	χ^2 (2)=.69, p=.71	F(2,398) =.73, p=.48		F(2,375) =.2.92, p=.06		F(2,399) =.25.81, p<.001 Mild-Moderate <Relatively Intact (p<.001, d=0.56) Relatively Severe < Relatively Intact (p<.001, d= 0.83) Relatively Severe < Mild-Moderate (trend p=.07, d=0.33)		χ^2 (2)=2.70, p=.26	χ^2 (2)=2.74, p=.26

Post-hoc p values are Games-Howell corrected for unequal samples sizes/unequal variances. * Note that premorbid IQ values represent residual scores (from which site differences have been regressed out) that are standardized against HC values (M=0, SD=1). Euthymic criteria = $\leq$ 8 mania and depression rating scale (see supplementary material for further details).

Table 2. Cognitive domain comparisons between HC and BD/SZ clusters.

SZ	**Relatively Intact**		**Mild-Moderate**		**Severe**			**Cluster comparisons**
Domain	***M***	***SD***	***M***	***SD***	***M***	***SD***	***Omnibus / Main effect***	***Post-hoc comparisons***
							F(21,3393)=61.31, p<.001	
SOP	-0.42	0.82	-1.43	0.89	-2.69	0.99	F(3,1135)=461.67, p<.001	SZ Relatively Intact vs HC p≤.001 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
AV	0.03	1.07	-1.17	1.00	-2.16	1.09	F(3,1135)=275.33, p<.001	SZ Relatively Intact vs HC p=.996 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
WM	0.31	0.88	-0.81	0.82	-2.55	0.94	F(3,1135)=432.90, p<.001	SZ Relatively Intact vs HC p=.030 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
VerL	-0.15	0.79	-0.87	0.76	-1.62	0.64	F(3,1135)=208.55, p<.001	SZ Relatively Intact vs HC p=.436 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
VisL	0.30	0.71	-0.82	0.90	-1.92	0.79	F(3,1135)=264.53, p<.001	SZ Relatively Intact vs HC p=.009 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
RPS	0.13	0.88	-0.77	0.78	-1.15	0.79	F(3,1135)=112.07, p<.001	SZ Relatively Intact vs HC p=.641 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001

SC	-0.35	0.99	-0.95	1.02	-1.56	0.92	F(3,1135)=154.42, p<.001	SZ Relatively Intact vs HC p=.022 SZ Mild-Moderate vs HC p<.001 SZ Relatively Severe vs HC p<.001 SZ Relatively Intact vs SZ Mild-Moderate p<.001 SZ Relatively Intact vs SZ Relatively Severe p<.001 SZ Mild-Moderate vs SZ Relatively Severe p<.001
BD	**Relatively Intact**		**Mild-Moderate**		**Severe**			**Cluster comparisons**
Domain	***M***	***SD***	***M***	***SD***	***M***	***SD***	***Omnibus / Main effect***	***Post-hoc comparisons***
							F(21,2907)=34.62, p<.001	
SOP	-0.38	0.88	-1.13	0.70	-2.41	0.84	F(3,973)=177.73, p<.001	BD Relatively Intact vs HC p<.001 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p<.001
AV	-0.29	1.08	-0.90	0.91	-2.112	0.91	F(3,973)=113.51, p<.001	BD Relatively Intact vs HC p=.005 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p<.001
WM	0.12	0.87	-1.24	0.83	-2.0847	1.08	F(3,973)=159.46, p<.001	BD Relatively Intact vs HC p=.401 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p<.001
VerL	-0.11	1.07	-0.73	0.89	-1.4348	0.77	F(3,973)=59.44, p<.001	BD Relatively Intact vs HC p=.564 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p<.001
VisL	0.06	0.90	-1.25	0.93	-1.2669	0.90	F(3,973),=93.31, p<.001	BD Relatively Intact vs HC p=.882 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Severe p=.1.00
RPS	0.09	0.91	-0.94	0.69	-1.1581	0.65	F(3,973)=69.41, p<.001	BD Relatively Intact vs HC p=.637 BD Mild-Moderate vs HC p<.001

								BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p<.001 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p=.107
SC	-0.19	0.99	-0.41	1.09	-0.8993	0.93	F(3,973)=21.45, p<.001	BD Relatively Intact vs HC p=.079 BD Mild-Moderate vs HC p<.001 BD Relatively Severe vs HC p<.001 BD Relatively Intact vs BD Mild-Moderate p=.222 BD Relatively Intact vs BD Relatively Severe p<.001 BD Mild-Moderate vs BD Relatively Severe p=.006
BD vs SZ								**Cluster comparisons**
Domain							***Omnibus / Main effect***	***Post-hoc comparisons***
SOP							F(5,960)=189.47, p<.001	*SZ Relatively Intact vs BD Relatively Intact p>.999* SZ Relatively Intact vs BD Mild-Moderate p<.001 SZ Relatively Intact vs vs BD Relatively Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p=.006* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Relatively Severe vs BD Severe p=.178*
AV							F(5,960)=106.43, p<.001	*SZ Relatively Intact vs BD Relatively Intact p=.244* SZ Relatively Intact vs BD Mild-Moderate p<.001 SZ Relatively Intact vs vs BD Relatively Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p=.091* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Relatively Severe vs BD* Relatively *Severe p>.999*
WM							F(5,960)=260.58, p<.001	*SZ Relatively Intact vs BD Relatively Intact p=.593* SZ Relatively Intact vs BD Mild-Moderate p<.001 SZ Relatively Intact vs vs BD Relatively Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p<.001* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Severe vs BD Relatively Severe p=.014*
VerL							F(5,960)=88.73, p<.001	*SZ Relatively Intact vs BD Relatively Intact p≥.999* SZ Relatively Intact vs BD Mild-Moderate p<.001

		SZ Relatively Intact vs vs BD Relatively Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p=.643* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Relatively Severe vs BD Relatively Severe p=.438*
VisL	F(5,960)=147.86, p<.001	*SZ Relatively Intact vs BD Relatively Intact p=.191* SZ Relatively Intact vs BD Mild-Moderate p<.001 SZ Relatively Intact vs vs BD Relatively Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p=.004* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Relatively Severe vs BD Relatively Severe p<.001*
RPS	F(5,960)=76.81, p<.001	*SZ Relatively Intact vs BD Relatively Intact p>.999* SZ Relatively Intact vs BD Mild-Moderate p<.001 SZ Relatively Intact vs vs BD Severe p<.001 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p=.244* SZ Mild-Moderate vs BD Relatively Severe p<.001 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p=.088 *SZ Relatively Severe vs BD Relatively Severe p>.999*
SC	F(5,960)=50.66, p<.001	*SZ Relatively Intact vs BD Relatively Intact p=.822* SZ Relatively Intact vs BD Mild-Moderate p=.998 SZ Relatively Intact vs vs BD *Relatively* Severe p=.008 SZ Mild-Moderate vs BD Relatively Intact p<.001 *SZ Mild-Moderate vs BD Mild-Moderate p<.001* SZ Mild-Moderate vs BD *Relatively* Severe p>.999 SZ Relatively Severe vs BD Relatively Intact p<.001 SZ Relatively Severe vs BD Mild-Moderate p<.001 *SZ Relatively Severe vs BD Relatively Severe p<.001*

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 Games-Howell test for unequal variances/unequal sample sizes. Post-hoc comparisons between corresponding SZ and BD clusters are italicized for readability. SOP = speed of processing; AV = attention/vigilance; WM = working memory; VerL = verbal learning; VisL =visual learning; RPS = reasoning and problem solving; SC = social cognition

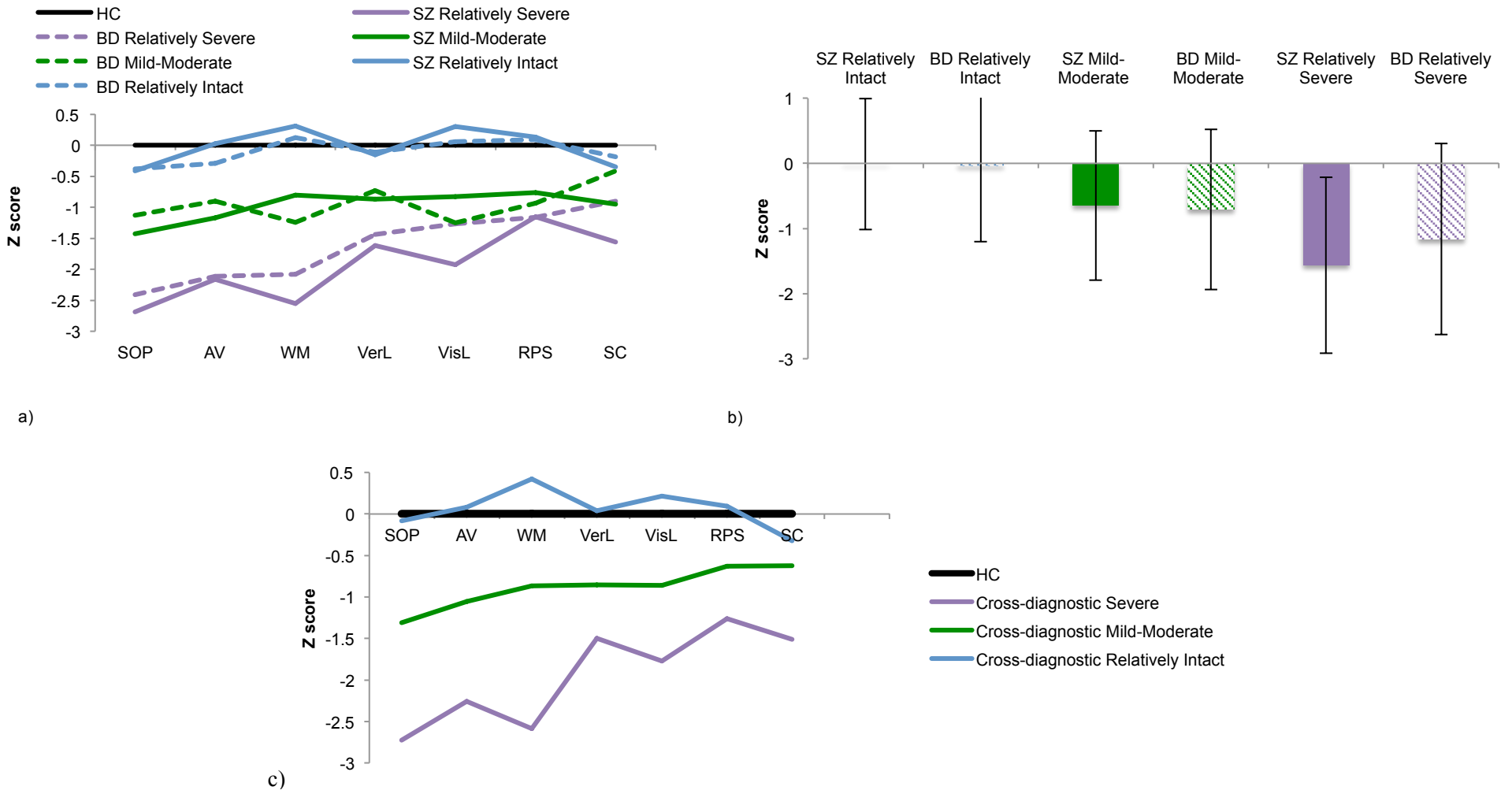


Figure 1. Cognitive profile and premorbid IQ comparisons of newly derived SZ, BD and cross-diagnostic clusters standardized against HC values (M = 0, SD = 1); a) cognitive domain scores comparing SZ clusters to BD clusters - Mild-Moderate SZ/BD significant differences: SOP d = 0.37, WM d = 0.52, VisL d =0.20, SC d =0.50 - Relatively Severely SZ/BD significant differences: WM d =0.46, VisL d = 0.77, SC d = 0.71; b) premorbid IQ scores; c) Cross-diagnostic clustering; SOP = speed of processing; AV = attention/vigilance; WM = working memory; VerL = verbal learning; VisL =visual learning; RPS = reasoning and problem solving; SC = social cognition. Error bars represent standard deviations

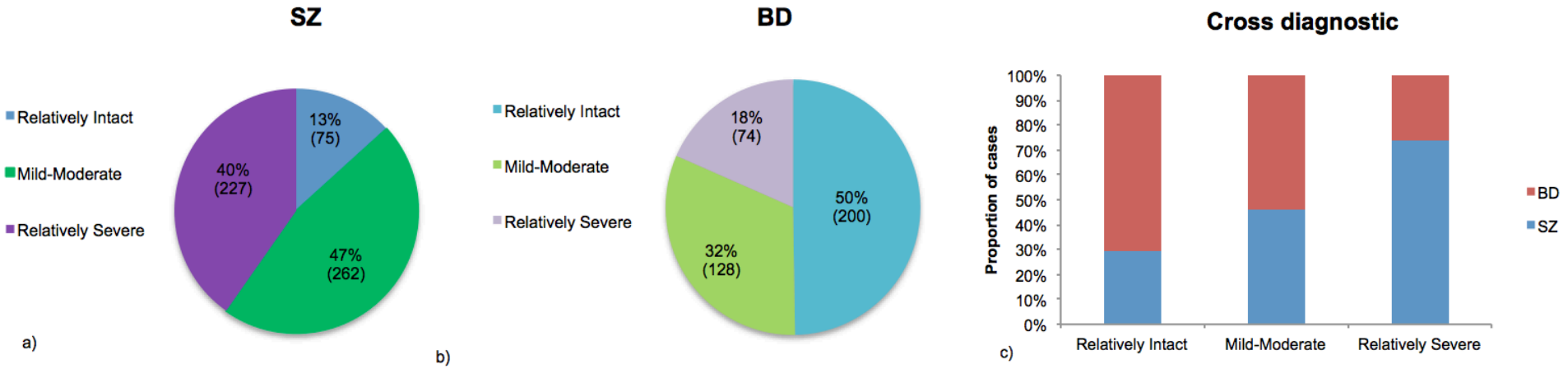
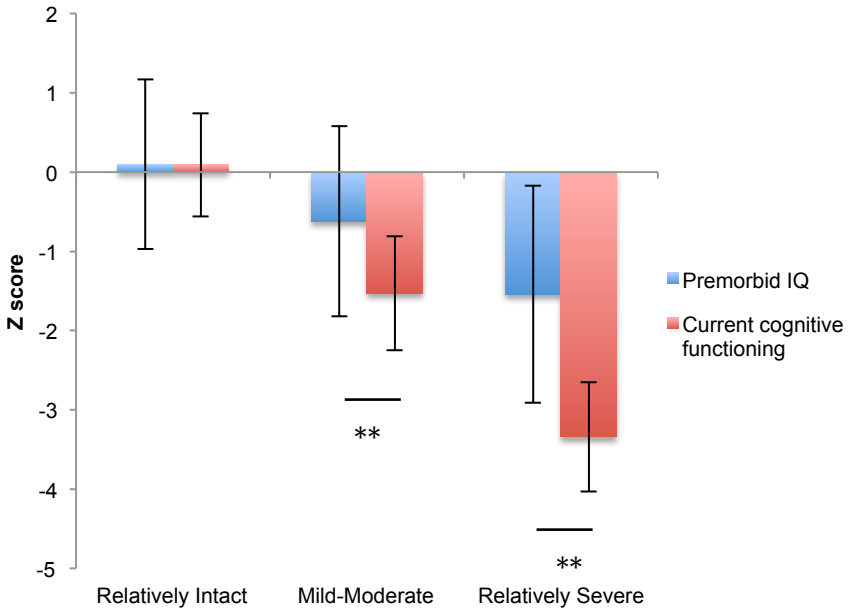
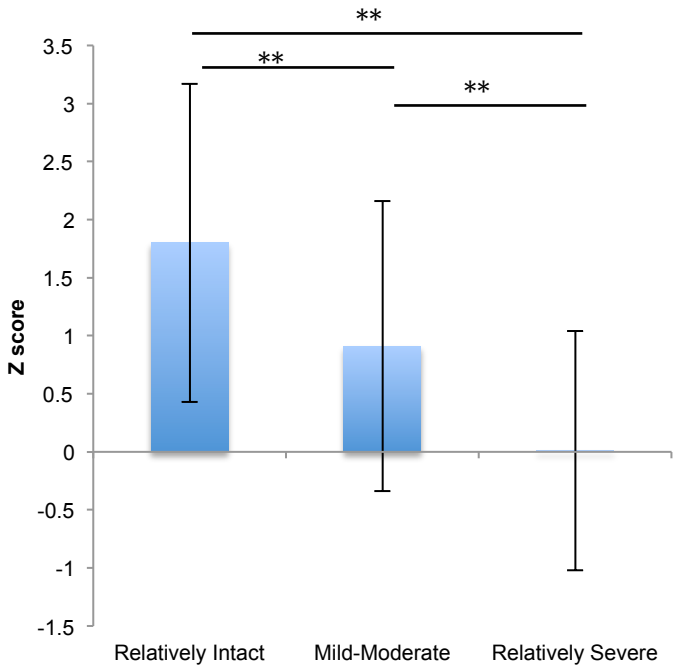


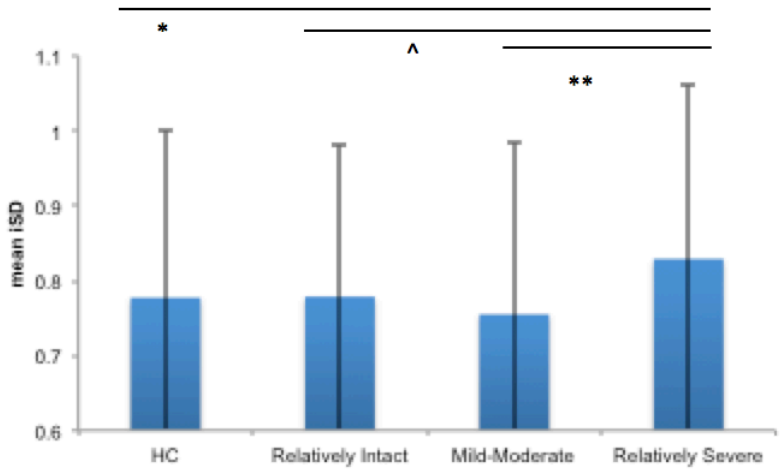
Figure 2. Percentage distribution of cases falling into each cluster when classified within categorical diagnoses and a cross-diagnostic approach a) SZ clusters; b) BD clusters; c) Cross-diagnostic clusters. Numbers in brackets in a and b represent the sample number per cluster. Figure c represents the proportion of BD or SZ participants (relative to the total sample number for each cohort) falling into each cluster.


2
1
0
-1
-2
-3
-4
-5
Z score
Premorbid IQ
Current cognitive functioning
**
**
Relatively Intact
Mild-Moderate
Relatively Severe

a)


3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Z score
**
**
**
Relatively Intact
Mild-Moderate
Relatively Severe

b)


*
^
**
1.1
1
0.9
0.8
0.7
0.6
mean iSD
HC
Relatively Intact
Mild-Moderate
Relatively Severe

c)

Figure 3. Analysis of estimated cognitive course in the cross-diagnostic clusters; a) estimated premorbid IQ vs. current cognitive functioning (z scores are standardised against HC mean/SD); b) discrepancy score indicating estimated improvement or decline in cognition (z scores are standardised against HC mean/SD). Larger scores indicate greater discrepancy between premorbid IQ and current cognitive functioning; c) dispersion (mean individual standard deviation) of cognitive domain scores. **p = .000; * p=.01; ^p = .06. Error bars represent standard deviations.

Characterizing cognitive heterogeneity on the schizophrenia – bipolar disorder spectrum

Supplementary Material

Tamsyn E. Van Rheenen*[a,b,c], PhD; Kathryn E. Lewandowski[e,f], PhD; Eric J. Tan[b,c], PhD; Luz H Ospina[g], PhD; Dost Ongur[e,f], MD, PhD; Erica Neill[b,d], PhD; Caroline Gurvich[c], DPsych; Christos Pantelis[a,j,k], MD, MRCPsych; Anil K Malhotra[h], MD; Susan L. Rossell[b,c,d], PhD and Katherine E Burdick[g,i], PhD.

[a] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, VIC, Australia
[b] Brain and Psychological Sciences Research Centre, Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, VIC, Australia
[c] Cognitive Neuropsychiatry Laboratory, Monash Alfred Psychiatry Research Centre, The Alfred Hospital and Central Clinical School, Monash University, Melbourne, Australia.
[d] Department of Psychiatry, St Vincent's Hospital, VIC, Australia
[e] Schizophrenia and Bipolar Disorder Program, McLean Hospital, Belmont, MA, USA
[f] Harvard Medical School, Department of Psychiatry, Boston, MA USA
[g] Icahn School of Medicine at Mount Sinai, NY, USA.
[h] Hofstra Northwell School of Medicine, Hempstead, NY, USA
[i] James J Peters VA Hospital, NY, USA
[j] Florey Institute for Neuroscience and Mental Health, VIC, Australia
[k] Centre for Neural Engineering (CfNE), Department of Electrical and Electronic Engineering, University of Melbourne, VIC, Australia

Participants

Data was drawn from schizophrenia (SZ), bipolar disorder (BD) and control (HC) participants from four sites: Monash Alfred Psychiatry Research Centre, Melbourne Australia (SZ n= 77, BD n = 50, HC n = 161); 2) Zucker Hillside Hospital, New York, USA (SZ n = 446, BD n =116, HC n = 312), 3) Icahn School of Medicine at Mount Sinai, New York, USA (SZ n = 22, BD n = 164, HC n = 35); 4) McLean Hospital, Belmont, MA, USA (SZ n = 19, BD n = 72, HC n = 67). All clinical participants were outpatients, and included those recruited for several different studies assessing cognitive functioning generally, genotype-phenotype relationships or efficacy of cognitive remediation treatment. In the case of the latter, only baseline scores were used.

Statistical analysis

To negate the influence of site differences associated with the potential for slight differences in administration of cognitive tests, we regressed site from the age and gender corrected T scores for all domains of cognition. The residual scores were then used in several hierarchical clustering analyses (HCA) with the aim of identifying more homogenous subgroups in the data. To compare cognitive profiles of newly emergent clusters across groups, HCAs were conducted independently in the HC, SZ and BD patients. A HCA was then conducted on the SZ and BD patients in cross-diagnostic manner. HCAs comprising patients were conducted using the residual cognitive domain scores standardized against HC means and standard deviations. In each HCA, case similarity was computed with the squared Euclidean Distance and Wards Linkage was specified as the agglomeration procedure. Visual inspection of the dendogram (with agreement between TVR and KB) and the cluster profiles of groups, as well as discriminant functions analyses (DFA) to assess the predictive power of each of the seven domains of cognition in differentiating participants into the subgroups resulting from the HCA, were used as criteria to establish the appropriate number of clusters to retain in each analysis. Stability of the final cluster solutions were determined by repeating each HCA in a random split half of each sample. Further information about the selection of clusters can be found below.

Multivariate analyses of variance (MANOVA) with post hoc tests were used to determine cluster differences in the cognitive profiles of the newly emergent cognitive clusters. In the case of the SZ, BD and cross-diagnostic cluster analyses, the MANOVAs were conducted including the HC group as the comparison group. Cognitive impairment was considered according to the following criteria: individuals with a z score less than 0.5 below HCs were cognitively intact, those with z-scores between 0.5 and 1.5 below HCs had mild-moderate impairment, and those with z-scores greater than 1.5 below HCs had severe impairment. Clusters were labeled according to these criteria. A MANOVA comparing the cognitive profiles of the newly emergent clusters across SZ and BD patients was also performed. This was followed by DFAs to determine classification accuracy for distinguishing between the new SZ and BD clusters that corresponded to each other in terms of similarity in cognitive profiles.

Descriptive analyses comparing the emergent clusters in terms of age, gender, estimated premorbid IQ, BPRS scores, psychosis history (BD only: yes/no), mood status (BD only: symptomatic/euthymic), diagnostic subtype (BD only: subtype I and II) or diagnosis itself (cross-diagnostic analysis only) were conducted using one-way analyses of variance (ANOVA), MANOVA or chi-square tests where appropriate. 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 (12, 13), 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 HRSD >8) or euthymic (CARS-M or YMRS score $\leq$ 8 and MADRS or HRSD $\leq$ 8) at the time of assessment, to compare performance across these subgroups within the new clusters.

To determine the extent to which cross-diagnostic clustering was more meaningful in reducing error variance in cognitive performance than clustering within each diagnosis separately, the following procedure was followed: i) the error sums of squares for the separate BD and SZ cluster comparisons of cognitive performance were adjusted in order to allow for the grand mean for the cross-diagnostic sample for each measure. This was completed as the sums of squares of the separate BD and SZ analyses were already corrected for sample means, which differed for the BD and SZ samples. ii) the adjusted error sums of squares for BD and SZ samples were added and one degree of freedom was added to the sum of the error degrees of freedom to allow for the grand mean adjustment, iii) an F-test was conducted using the quotient of the adjusted Mean Square Errors for the separate BD and SZ clusters and the Mean Square Error for the combined sample clustering.

To understand differences in premorbid intellectual functioning and estimated cognitive course between HCs and the clusters generated by the most meaningful clustering method (i.e., clustering BD or SZ independently or cross-diagnostically, as determined by the above procedure), estimated premorbid IQ (as assessed by the collated, site-corrected residual scores on the WTAR, WRAT and NAART) was assessed against a global estimate of current cognitive functioning (measured by a composite of all residual cognitive domains scores) using paired samples t-tests to determine evidence of stability or decline in cognition. Discrepancy scores calculated by subtracting current cognitive performance from premorbid IQ were compared across clusters to assess magnitude differences in decline, using one-way ANOVA. A measure of subtest dispersion (i.e., the standard deviation of variance across each cognitive domain) was also calculated to compare clusters on an indirect estimate of cognitive ageing.

Finally, to determine whether the cross-diagnostic cluster solution was uni or multi-dimensional 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 post-hoc tests were corrected using the Games-Howell method for unequal variance and unequal sample size.

Selection of cluster solutions

Although the dendograms of all hierarchical clustering analyses (HCAs) provided evidence for a three-cluster model, in some cases two and four-cluster models were also evident. Thus, we chose to investigate these alternate models to determine the most meaningful clustering solution based on the data. We ran discriminant functions analysis (DFA) on all three clustering solutions (two-cluster, three-cluster and four-cluster) for each group. These were conducted in the interests of ascertaining the relative accuracy of the cognitive domains in predicting membership into the new clusters. A leave one out classification was used to assess the reliability of the original classification model generated by each DFA. This leave-one-out cross-validation works by deleting each case in turn, and reclassifying the remaining observations by means of the classification rule established in the original model.

In all cases, the DFAs indicated slightly better predictive power for the two-cluster models, followed by the three-cluster models and then the four-cluster models; however differences in overall classification accuracy between models were small (<8%), except in the case of the healthy controls (HC; ~13%). The overall cross-validated classification accuracy for all cluster solutions was medium - high (>65%) for each analysis (BD 4 cluster solution =67.7%; 3 cluster solution = 76.4%; 2 cluster solution - 83.1%; SZ 4 cluster solution = 75.4%; 3 cluster solution = 78.4%; 2 cluster solution = 86.9%) except in the case of HCs where there was poor classification accuracy (3 cluster solution = 62% and 4 cluster solution = 59% respectively) for the three and four-cluster models. As a result, a two-cluster solution was decided on for the HCs (accuracy =75%).

Inspection of the cognitive profiles of the three different cluster solutions in the whole group and patient group analyses showed that the two-cluster model provided evidence for generically ‘better’ and ‘worse’ cognitive functioning, while the four cluster solution further separated one of the clusters from the three-cluster model. This resulted in reduced cluster cell sizes and smaller cluster separating effect sizes in the four-cluster solution. Thus, a three-cluster solution for the whole sample, SZ, BD and cross-diagnostic group analyses were selected as providing good differentiation amongst the clusters and meaningful groupings in terms of cognitive profiles in these clusters.

The DFA results for the final cluster solutions in each group are reported in Supplemental Table S1. Supplemental Figure S1 shows the graphs of the canonical discriminant functions for the whole sample, SZ, BD and cross-diagnostic clustering analyses.

Table S1. *Discriminant Functions Analysis results for final cluster solutions in each group*

Group	DFA results	Classification model type		Cluster (%)		
HC	Function 1: (χ^2 (7) = 220.39, p <.001, canonical correlation =.57): Highest loading domain; Processing Speed (r=.63)			1	2	-
		Original	1	82.6	17.4	
			2	34.9	65.1	
		Cross-validated	1	81.8	18.2	
			2	36.2	63.8	
		75% of cross-validated grouped cases correctly classified				
SZ	Function 1: (χ^2 (14) = 625.02, p <.001, canonical correlation =.80): Highest loading domain; Working Memory (r=.75) Function 2 (χ^2 (6) = 34.28, p <.001, canonical correlation =.24): Highest loading domain; Reasoning and Problem Solving (r=.63)			1	2	3
		Original	1	81.5	18.5	.0
			2	9.9	83.6	6.5
			3	.0	38.7	61.3
				1	2	3
		Cross-validated	1	81.1	18.9	.0
			2	11.5	81.7	6.9
			3	.0	41.3	58.7
		78.4% of cross-validated grouped cases correctly classified				
BD	Function 1: (χ^2 (14) = 406.86, p <.001, canonical correlation =.76): Highest loading domain; Working Memory (r=.71) Function 2 (χ^2 (6) = 72.80, p <.001, canonical correlation =.41): Highest loading domain; Visual Learning (r=-.56)			1	2	3
		Original	1	68.0	24.2	7.8
			2	11.0	88.0	1.0
			3	17.6	14.9	67.6
				1	2	3
		Cross-validated	1	68.0	24.2	7.8
			2	11.5	86.5	2.0
			3	21.6	14.9	63.5
		76.4% of cross-validated grouped cases correctly classified				
Cross-diagnostic	Function 1: (χ^2 (14) = 1159.959, p <.001, canonical correlation =.83): Highest loading domain; Working Memory (r=.75) Function 2 (χ^2 (6) = 38.794, p <.001, canonical correlation =.20): Highest loading domain; Social Cognition (r= -.57)			1	2	3
		Original	1	80.9	19.1	.0
			2	6.2	85.4	8.4
			3	.0	34.7	65.3
				1	2	3
		Cross-validated	1	80.5	19.5	.0
			2	6.6	84.6	8.8
			3	.0	35.2	64.8
		79.5% of cross-validated grouped cases correctly classified				

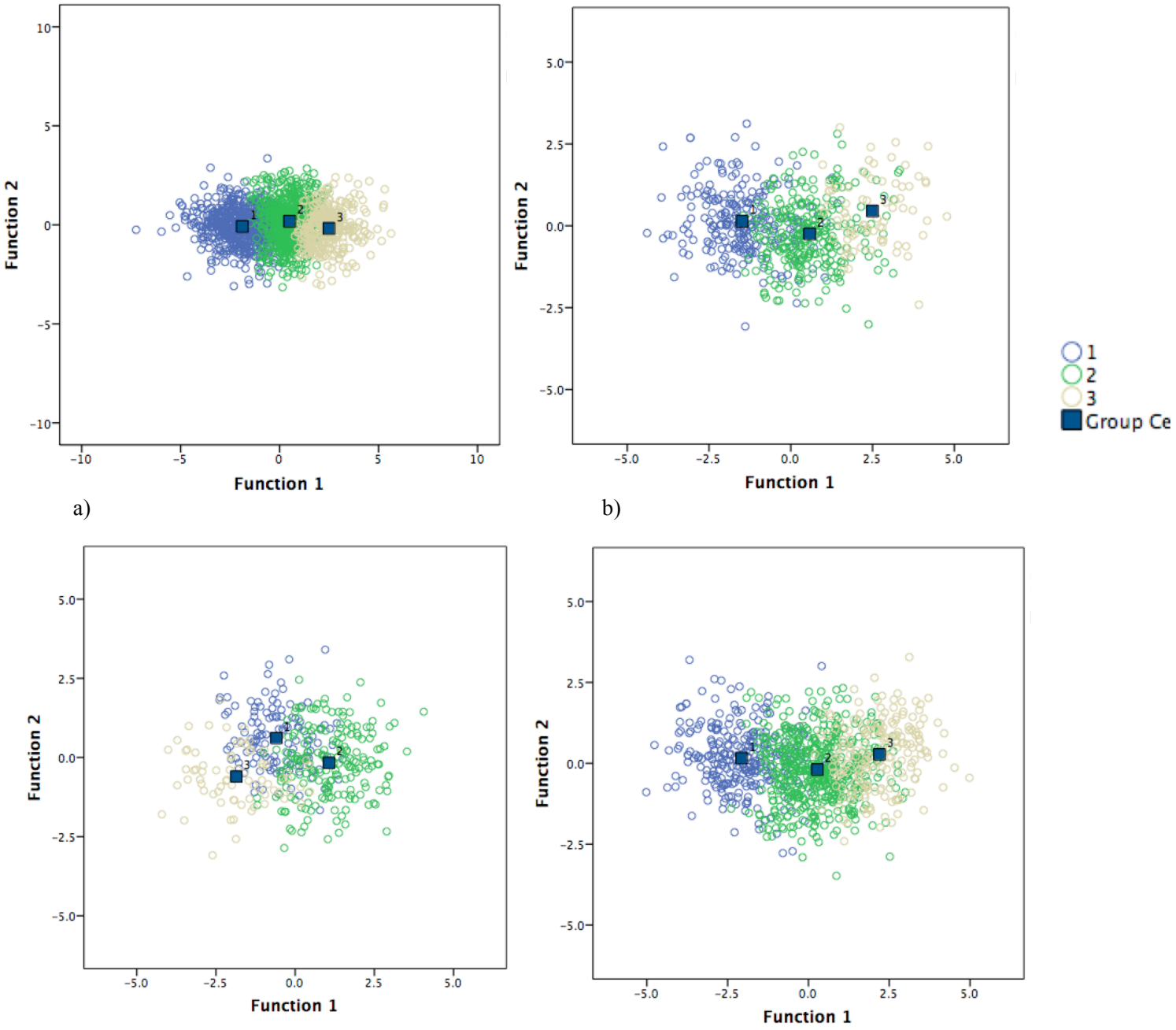


Figure S1. Graphs of Canonical Discriminant Functions; a)Whole group; b) SZ, c) BD, d) Cross-diagnostic. Note that no graph is produced for the 2-cluster solution in HCs

Table S2. *Descriptives for the sample prior to clustering*

	SZ			BD			Comparisons
	% male	**M**	**SD**	**% male**	**M**	**SD**	
Gender	73			51			$\chi^2(1)=49.47$, $p<.001$
Age		42.80	10.43		40.22	12.26	$F(1,963)=12.38$, $p<.001$
BPRS		31.51	7.93		24.84	5.82	$F(1,901)=192.27$, $p<.001$
Premorbid IQ		-0.93	1.33		-0.46	1.32	$F(1,962)=29.66$, $p<.001$

Note that premorbid IQ scores represent z scores standardized against healthy control means and SD's (M=0, SD=1).

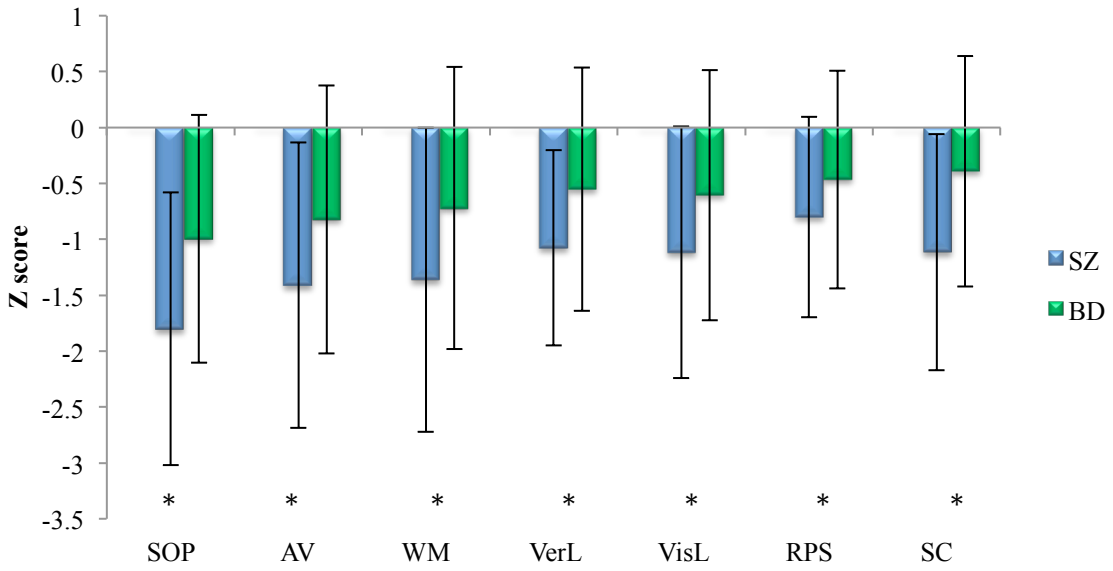


Figure S2. Cognitive performance for SZ and BD participants prior to clustering. Clinical group scores are standardised against healthy control means and standard deviations. *$p<.001$; SOP = speed of processing; AV = attention/vigilance; WM = working memory; VerL = verbal learning; VisL =visual learning; RPS = reasoning and problem solving; SC = social cognition

Table S3. *Chi Square comparisons of cluster assignment distributions between diagnoses*

Subgroup	**Diagnosis (n)**		**Comparison**
	SZ	BD	
Relatively Intact	75	200	χ^2 (1)= 275, p<.001
Mild-Moderate	262	128	χ^2 (1)=390, p<.001
Relatively Severe	227	74	χ^2 (1)=301, p<.001

Table S4. *Cross-validated classification results for the categorical cluster (SZ or BD) predicting discriminant functions analyses*

Cluster comparison	Model type	Diagnostic category		
			SZ Relatively Intact	BD Relatively Intact
Relatively Intact	Original	SZ Relatively Intact	**2.7%**	97.3%
		BD Relatively Intact	1.0%	**99.0%**
	Cross-validated	SZ Relatively Intact	**2.7%**	97.3%
		BD Relatively Intact	1.0%	**99.0%**
			SZ Mild-Moderate	BD Mild-Moderate
Mild-Moderate	Original	SZ Mild-Moderate	**89.3%**	10.7%
		BD Mild-Moderate	59.4%	**40.6%**
	Cross-validated	SZ Mild-Moderate	**88.5%**	11.5%
		BD Mild-Moderate	64.1%	**35.9%**
			SZ Relatively Severe	BD Relatively Severe
Relatively Severe	Original	SZ Relatively Severe	**95.6%**	4.4%
		BD Severe	62.2%	**37.8%**
	Cross-validated	SZ Relatively Severe	**95.2%**	4.9%
		BD Relatively Severe	66.2%	**33.8%**

Bolded values indicate % correct classification for each diagnostic category

Table S5. *Classification and misclassification of patients into corresponding cross-diagnostic clusters*

Cluster	**Corresponding cluster classification for cross-diagnostic clustering (n)**		**Classification of misclassified patients into non-corresponding cross-diagnostic clusters (n)**
	Corresponding	**Not corresponding**	
SZ Relatively Intact	56/19	19	19 into Mild-Moderate
SZ Mild-Moderate	231/31	31	15 into Relatively Intact; 16 into Relatively Severe
SZ Relatively Severe	202/25	25	25 into Mild-Moderate
BD Relatively Intact	122/78	78	78 into Mild-Moderate
BD Mild-Moderate	124/4	4	4 into Relatively Severe
BD Relatively Severe	50/24	24	24 into Mild-Moderate

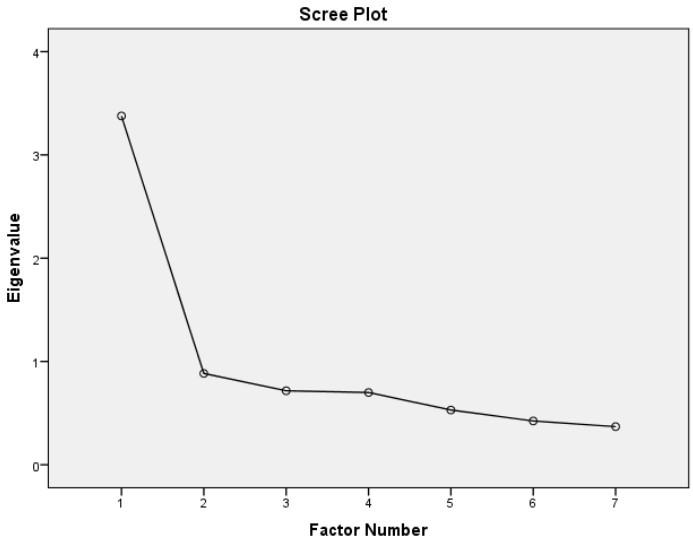


Figure S3. Scree plot graphing eigenvalues in the exploratory factor analysis. Factor 1 explained 40.55% of the variance in the model (eigenvalue 2.84). It was the only factor with an eigenvalue over 1 and was retained as evidence of a unfactorial model for cognition in SZ and BD. Note that the Kaiser-Meyer-Olkin measure of sampling adequancy =.86 and Bartlett's Test of Sphericity was significant at $p<.001$

Table S6. *Communality values from the exploratory factor analysis*

Cognitive domain	**Extraction**
Speed of processing	.61
Attention/Vigilance	.35
Working Memory	.63
Verbal Learning	.36
Visual Learning	.41
Reasoning and Problem Solving	.29
Social Cognition	.18

The values in the 'Extraction' column indicate the proportion of each cognitive domains' variance that can be explained by the retained factor.

Table S7. *Factor loadings from the exploratory factor analysis*

Cognitive domain	Factor 1
Working Memory	.79
Speed of processing	.78
Visual Learning	.64
Verbal Learning	.60
Attention/Vigilance	.60
Reasoning and Problem Solving	.54
Social Cognition	.42

Note: The factor loadings indicate the extent to which each domain is correlated with the factor.

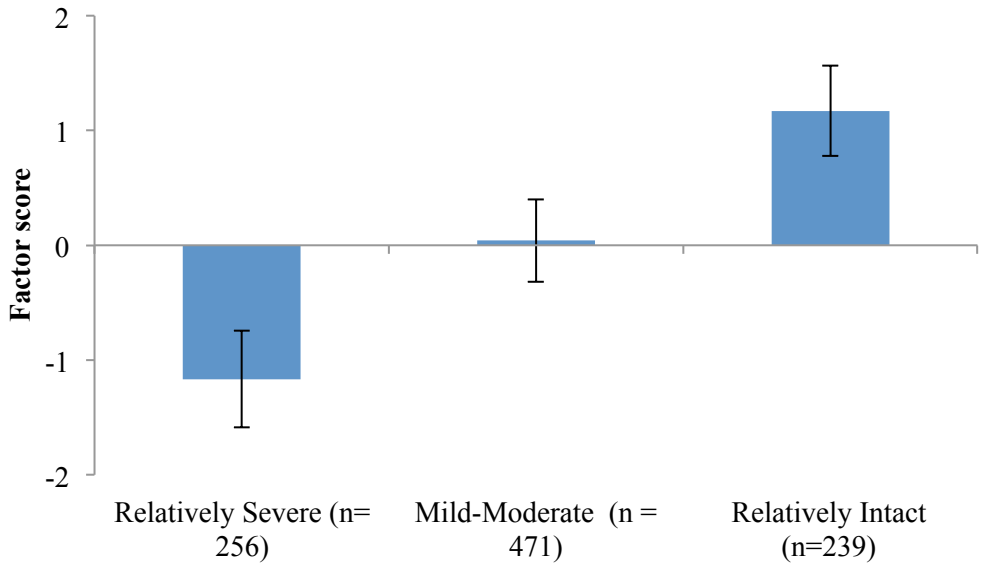


Figure S4. Graphical representation of the 3 emergent subgroups that were clustered on the basis of the unidimensional factor score. All groups differed significantly from each other (all p's <.001). A discriminant functions analysis indicated excellent predictive power for the general factor score in predicting membership into each of the clusters (cross-validated classification accuracy = 99.3%). There was 88.8% cross-over between patients falling into the original cross-diagnostic clusters generated on the basis of all MCCB domains, and the unifactorial clusters.

References

1. Sheehan DV, Lecrubier Y, Harnett Sheehan K, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of Clinical Psychiatry. 1998;59:22-33.
2. First MB, Spitzer RL, Gibbon M, Williams JB: Structured Clinical Interview for DSM-IV Axis I Disorders Clinician Version (SCID-CV). Washington, D.C, American Psychiatric Press; 1996.
3. Overall JE, Gorham DR. The brief psychiatric rating scale. Psychol Rep. 1962;10:799-812.
4. Kay SR, Flszbein A, Opfer LA. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. Schizophr Bull. 1987;13:261-276.
5. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. British Journal of Psychiatry. 1979;134:382-389.
6. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23.
7. Young R, Biggs J, Ziegler V, Meyer D. A rating scale for mania: reliability, validity and sensitivity. The British Journal of Psychiatry. 1978;133:429-435.
8. Altman EG, Hedeker DR, Janicak PG, Peterson JL, Davis JM. The clinician-administered rating scale for mania (CARS-M): Development, reliability, and validity. Biol Psychiatry. 1994;36:124-134.
9. Holdnack JA: Wechsler Test of Adult Reading. San Antonio, Texas, The Psychological Corporation; 2001.
10. Blair JR, Spreen O. Predicting premorbid IQ: a revision of the National Adult Reading Test. The Clinical Neuropsychologist. 1989;3:129-136.
11. Jastak S, Wilkinson GS: The wide range achievement test-revised: administration manual, Jastak Assessment Systems; 1984.
12. Leucht S, Rothe P, Davis J, Engel R. Equipercentile linking of the BPRS and the PANSS. Eur Neuropsychopharmacol. 2013;23:956-959.
13. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. Neuropsychopharmacology. 2006;31:2318-2325.