Cognitive validation of cross-diagnostic cognitive subgroups on the schizophrenia-bipolar spectrum

James A. Karantonis [a,b*], Susan L. Rossell [b,c], Sean P. Carruthers [a,b], Philip Sumner [b], Matthew Hughes [b,] Melissa J. Green[i,j], Christos Pantelis [a,d,e,f], Katherine E. Burdick[g,h],Vanessa Cropley [a,b] and Tamsyn E. Van Rheenen [a,b]

[a] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Melbourne, Australia
[b] Centre for Mental Health, Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, Melbourne, Australia
[c] St Vincent's Mental Health, St Vincent's Hospital, VIC, Australia
[d] Florey Institute of Neuroscience and Mental Health, Australia;
[e] Department of Electrical and Electronic Engineering, University of Melbourne, VIC, Australia
[f] Centre for Neuropsychiatric Schizophrenia Research (CNSR) and Centre for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), Mental Health Centre Glostrup, Denmark
[g] Harvard Medical School, Department of Psychiatry, Boston, MA USA
[h] Brigham and Women's Hospital, Boston, MA USA
[I] School of Psychiatry, University of New South Wales (UNSW), Sydney, NSW, Australia
[j] Neuroscience Research Australia, Randwick, NSW, Australia

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

Abstract

Background: Cognitive heterogeneity in schizophrenia spectrum disorders (SSD) and bipolar disorder (BD) has been explored using clustering analyses. However, the resulting subgroups have not been cognitively validated beyond measures used as clustering variables themselves. We compared the emergent cross-diagnostic subgroups of SSD and BD patients on measures used to classify them, and also across a range of *alternative* cognitive measures assessing some of the same constructs.

Method: Domain scores from the Matrics Consensus Cognitive Battery were used in a cross-diagnostic clustering analysis of 86 patients with SSD (n = 45) and BD (n = 41). The emergent subgroups were then compared to each other and healthy controls (n = 76) on these and alternative measures of these domains, as well as on premorbid IQ, global cognition and a proxy of cognitive decline.

Results: A three-cluster solution was most appropriate, with subgroups labelled as Globally Impaired, Selectively Impaired, and Superior/Near-Normal relative to controls. With the exception of processing speed performance, the subgroups were generally differentiated on the cognitive domain scores used as clustering variables. Differences in cognitive performance among these subgroups were not always statistically significant when compared on the alternative cognitive measures. There was evidence of global cognitive impairment and putative cognitive decline in the two cognitively impaired subgroups.

Limitations: For clustering analysis, sample size was relatively small.

Conclusions: The overall pattern of findings tentatively suggest that emergent cross-diagnostic cognitive subgroups are not artefacts of the measures used to define them, but may represent the outcome of different cognitive trajectories.

Keywords: clustering; heterogeneity; bipolar disorder; schizophrenia; cognition; neuropsychology

Cognitive impairment is a prevalent feature of bipolar disorder (BD) and schizophrenia-spectrum disorders (SSD), including schizophrenia (SZ) and schizoaffective disorder (SZA)(Burdick et al., 2011; Harvey et al., 2016; Van Rheenen et al., 2016). Similarities in the patterns of cognitive performance across these disorders are apparent, suggesting that cognitive impairment does not clearly map to diagnostic labels (Krabbendam et al., 2005; Martinez-Aran and Vieta, 2015; Van Rheenen et al., 2016; Van Rheenen et al., 2017). To date, research has commonly employed a group average approach in comparing patients with these disorders to healthy controls, which has provided a good foundation for understanding the degree of cognitive impairment in each of these disorders. However, this approach tends to overlook individual variability within-groups, which is problematic given that patients with the same psychiatric diagnosis have been shown to vary extensively in terms of symptoms, functionality and illness course (Clementz et al., 2016; Dacquino et al., 2015).

Recent studies have focused on characterising heterogeneity in cognition in SSD and BD by using clustering analyses to identify underlying cognitive subgroups (Burdick et al., 2014; Carruthers et al., 2019a; Carruthers et al., 2019b). One very large study showed that patients with SSD and BD each clustered into the same number of cognitive subgroups, but despite the cognitive performance of these subgroups differing significantly from each other *within* diagnoses, the performance of each respective subgroup did not differ substantially *between* diagnoses (Van Rheenen et al., 2017). Hence, the structure and profile of cognitive heterogeneity in these disorders did not appear to differ, and subsequent analyses indicated that the cognitive subgroups derived from them may be best represented cross-diagnostically.

To date, there have been very few clustering studies of BD and SSD using the cross-diagnostic method. Of those currently published, the findings are varied, with evidence for two

(Lee et al., 2017), three (Van Rheenen et al., 2017) and four (Lewandowski et al., 2014) cluster solutions (Green et al., In Press). These are respectively defined by either i) high and low performing subgroups; ii) severely impaired, mild-moderately impaired and selectively impaired subgroups; or iii) two mixed profile subgroups alongside significantly impaired and neuropsychologically normal subgroups. Nonetheless, these cluster solutions appear to be anchored by 'near normal' and 'severely/globally impaired' subgroups in all studies, where approximately 20-30% of individuals with a SSD and between 40-50% of those with BD fall into the former category. In turn, approximately 20% of individuals across these disorders fall into the latter category.

With the range of emerging clustering solutions, replication of findings is required to determine consistency in cross-diagnostic cognitive cluster assignment of patients across samples. Indeed, whilst the evidence for heterogeneity in SSD and BD is steadily growing, what remains controversial is how this heterogeneity is best represented. Therefore, in this study we aimed to build on research using clustering approaches to characterise cognitive heterogeneity in these disorders, by attempting to replicate cross-diagnostic subgroups in a new sample of patients with SSD and BD. We also aimed to overcome a relatively neglected component of work in this area to date, by determining the extent to which the emergent subgroup classifications may be an artefact of the particular measures used to create them.

Indeed, in past research there has been limited independent cognitive validation of emerging subgroups, and studies have typically only examined subgroup differences in cognitive performance on cognitive tests that have themselves been used as clustering variables. Not surprisingly, this approach often produces clear differences between subgroups, but does not necessarily validate the cognitive profile when using measures beyond those used to create them

(Heinrichs et al., 2015). One recent study by Lewandowski et al. (2018) used clustering analysis to examine the replicability of cross-diagnostic cognitive subgroups arising in their previous work (Lewandowski et al., 2014). Importantly, the authors were able to replicate the four cluster solution previously identified using a similar cognitive battery. However, in their replication they still tested for group differences on the measures used *in the clustering procedure*. This method is inherently biased toward producing group differences given the very premise of clustering is to find similarities between certain cases that differentiate from others. In the current study we aimed to build on this work, in the interests of internally validating cross-diagnostic subgroups across cognitive domains *in the same sample*. Indeed, here we not only clustered a cross-diagnostic sample of SSD and BD patients based on their performance on a battery of cognitive tasks, but also compared the resulting cross-diagnostic subgroups on a range of *alternative* cognitive measures assessing some of the same constructs. As a further means of validating the subgroups, we also examined the utility of scores of estimated premorbid cognition, current global cognition, and the discrepancy between the two—a proxy of cognitive decline—in differentiating them. These scores may reflect underlying differences in the cognitive course of illness between patients, and there is already some preliminary evidence that they differ between subgroups of BD and/or SSD (Jensen et al., 2016; Reser et al., 2015; Van Rheenen et al., 2017; Woodward and Heckers, 2015).

We predicted that the cognitive subgroups emerging from a cross-diagnostic sample of SSD and BD cases would be differentiated on the measures used to classify cognitive subgroups. In the context of evidence that similar subgroups emerge from subgrouping studies of BD and/or SSD using from different clustering methodologies and/or cognitive batteries, we also predicted that the subgroups would retain their characteristic profile of cognitive performance when

compared on alternative cognitive measures independent to those from which their classification was derived.

Method

This study was approved by the relevant Human Ethics Review Board and abided by the Declaration of Helsinki.

Participants

The sample comprised $n = 86$ outpatients with a confirmed DSM-IV-TR diagnosis of SZ ($n = 30$), SZA ($n = 15$), or BD type I ($n = 41$), and $n = 76$ healthy controls. All participants provided informed consent. Psychiatric diagnosis (including BD subtype) and healthy control eligibility was confirmed using the Mini-International Neuropsychiatric Interview (MINI)(Lecrubier et al., 1997). Positive, negative and general symptomology was assessed with the Positive and Negative Syndrome Scale (PANSS)(Müller et al., 1998; Van den Oord et al., 2006), whilst depression and mania symptoms were assessed with the Montgomery and Asberg Depression Rating Scale (MADRS)(Davidson et al., 1986) and Young Mania Rating Scale (YMRS)(Youngstrom et al., 2003), respectively. All participants were proficient in English and aged between 18 and 65 years. Those who were currently abusing or were dependent on alcohol or substances in the last three months, had significant visual or verbal impairments, or a known neurological disorder were excluded. Healthy controls with a first degree relative with a mental illness were not included.

Cognitive Measures

All cognitive test scores were transformed to z-scores based on the healthy control sample means and standard deviations.

Primary clustering measures. All participants completed the MATRICS Consensus Cognitive Battery (MCCB) (Nuechterlein et al., 2008), which has been validated in both SSD and BD samples (Burdick et al., 2011; Sperry et al., 2015; Tan and Rossell, 2014; Van Rheenen and Rossell, 2014). In order to replicate past cross-diagnostic cognitive clustering studies of BD and/or SSD (Burdick et al., 2011; Lewandowski et al., 2018; Van Rheenen et al., 2017), the following MCCB domain scores (standardised sum of relevant subtests) were used as clustering variables: *speed of processing*, comprising the Trail Making Test-A (TMT-A), Animal Naming, and the Brief Assessment of Cognition in Schizophrenia: Symbol Coding (BACS-SC); *working memory,* comprising the Letter Number Span (LNS) and Wechsler Memory Scale 3rd Edition Spatial Span task (WMS-SS); *verbal learning*, comprising the Hopkins Verbal Learning Test-Revised (HVLT-R); *visual learning*, comprising the Brief Visual Memory Test (BVMT); and *reasoning and problem solving* (a component of executive functioning), comprising the Neuropsychological Assessment Battery: Mazes (NAB-Mazes). Although past work also included the Continuous Performance Task-Independent Pairs (CPT-IP) d'prime score from the MCCB to measure *attention*, we did not include this domain in the primary clustering analyses here, as HCA is not robust to missing data and there were missing cases on this measure (n = 21). Further, the Mayer-Salovey Caruso Emotional Intelligence Test from the MCCB was not included in the clustering analysis given our focus on *non-social* cognition.

Alternative 'non-clustering' cognitive measures. Participants completed several alternative cognitive measures that were used to create secondary domain scores of the following: *speed of processing*, comprising the Delis-Kaplan Executive Function Systems (D-

KEFS) Colour-Word Interference Test (CWIT) (Delis, 2001) colour naming and word reading trials; *working memory* comprising the Wechsler Memory Scale Digit Span forwards and backwards trials; and *executive function,* comprising the Trail Making Test-B (TMT-B), CWIT inhibition[1] contrast score and Wisconsin Card Sorting Test (WCST) categories completed and perseverative errors. Despite missing data, the MCCB CPT-IP d'Prime score was included to assess *attention* as an additional means of ascertaining cluster differentiation. All tests are described in detail elsewhere (Delis, 2001; Elwood, 1991; Heaton, 1993; Nuechterlein et al., 2008).

Further differentiating cognitive variables (premorbid and current cognition, and estimated decline scores). The Wechsler Test of Adult Reading (WTAR)(Holdnack, 2001) was used to estimate *premorbid IQ*. To measure *current global cognition*, a composite score was calculated from the standardised sum of the *z*-scores of the MCCB cognitive domain scores used as clustering variables (speed of processing, working memory, verbal and visual learning, reasoning and problem solving). As per previous studies (Russo et al., 2017; Van Rheenen et al., 2017), a discrepancy score was calculated by subtracting estimated premorbid IQ scores from the current global cognition composite score to estimate cognitive *'decline'.* Both the premorbid IQ and current global cognition composites scores were standardised to healthy control means and SDs. An additional composite score comprising the non-clustering cognitive domain measures (speed of processing, working memory, executive functioning and attention) was also calculated using the above method, as a means of comparing global cognitive performance using different component measures.

[1] An inhibition contrast score was computed from the average score of the first two trials of the test (colour naming and word reading), subtracted from the score of the third trial (inhibition).

Statistical Analysis

All data was analysed using the Statistical Package for the Social Sciences version 25. In a preliminary data analysis, demographic, clinical and cognitive performance was compared between controls and the traditional diagnostic categories of BD and SSD using Chi-square tests, analysis of variance (ANOVA) or multivariate analysis of (co)variance (MANCOVA; see supplementary material for more detail). The speed of processing, working memory, verbal and visual learning, reasoning and problem solving domain scores from the MCCB were then entered as clustering variables in a hierarchical clustering analysis (HCA; Ward's method, squared Euclidean distance) using the cross-diagnostic patient sample.

In line with our previous work (Van Rheenen et al., 2017), a combination approach was used to determine the final cluster solution. We reasoned that this approach would limit the influence of the subjectivity involved in deciding clustering outcomes. This method included visual inspection of the dendrogram and scree plot (supplementary Figure S1), in conjunction with profile analysis to identify cluster solutions with the greatest cluster differentiation. Inspection of the cluster agglomeration schedule coefficients alongside discriminant functions analyses (leave-one-out classification) were also used to ascertain the predictive power of the clustering variables in differentiating newly emergent subgroups (supplementary Figure S2). The final cluster solution was based on consensus between authors TVR and JK on the number of clusters visible in the dendrogram, taking into consideration the information gathered from the scree plot, results of the discriminant functions analysis, agglomeration schedule, and profile analysis.

The emergent cross-diagnostic clusters[2] were compared to each other and to healthy controls on demographic and clinical variables using chi square tests or analyses of variance (ANOVA) with *post-hoc* correction for unequal variances/sample sizes (Games-Howell). Age differences were evident between subgroups and controls, therefore age was included as a covariate in the three subsequent MANCOVAs comparing these groups on: (1) the primary clustering variables (select MCCB domain scores), (2) the alternative 'non-clustering' cognitive domain scores, and (3) the premorbid IQ, current global cognition, and (3a) estimated cognitive decline scores[3].

For each separate cognitive analysis, a false discovery rate (FDR) of 5% was set to correct for multiple comparisons at the omnibus level. Whenever the null hypothesis of equality across the emergent subgroups and controls was rejected, pair-wise post-hoc tests were performed and corrected using an FDR rate of 5% to determine where the group difference lay.

Results

Descriptive statistics and details for group comparisons on the cognitive clustering variables between SSD, BD and control groups are provided in supplementary Tables S1 and S2, as well as supplementary Figure S3. As per previous whole group findings, on the MCCB the BD patients generally performed intermediate to SSD patients and controls, with reduced performance relative to controls on the domains of speed of processing and verbal learning. SSD patients performed significantly worse than controls on all domains, and worse than BD patients

[2] We conducted a series of analyses entering diagnostic category (SSD, BD) and cluster membership as fixed factors to ascertain whether the cognitive performance of patients within each emergent cluster differed by diagnosis. No diagnosis * cluster interactions were evident, indicating the same performance across diagnostic categories within clusters. For brevity, these analyses are not reported.

[3] As estimated cognitive decline scores for each cross-diagnostic cluster were calculated from data standardised to healthy control means/SDs, we were unable to calculate meaningful discrepancy scores for healthy controls, and thus, they were not included in the analysis. Analysis 3a was conducted using an analysis of (co)variance.

on the domains of visual learning and working memory. On the alternative 'non-clustering' domains, BD patients performed significantly better on the domain of working memory compared to SSD patients. Compared with healthy controls, there was no significant difference in BD patient performance, whilst SSD patients performed worse on working memory, executive function, and attention domains.

(1) Cross-diagnostic cluster analysis using primary clustering measures.

A three-cluster solution provided the most appropriate fit for the data (supplementary material). A three-cluster solution provided the most appropriate fit for the data (supplementary material). Cluster 1 was characterised by significant impairments across all domains used in the cluster analysis, to the order of 1.2–2.0 SDs below the healthy control mean. This cluster was labelled Globally Impaired and had significantly worse performance on all cognitive variables relative to the other two emergent clusters. Cluster 2 was characterised by significant deficits relative to healthy controls (0.3–0.9 SDs) on the speed of processing, working memory and reasoning and problem solving domains, but showed no differences on the verbal and visual learning domains. Cluster 2 was labelled Selectively Impaired. Cluster 3 performed significantly better on the domains of visual learning, working memory, and reasoning and problem solving relative to healthy controls, in the order of 0.4–0.5 SDs. However, Cluster 3 had equivalent performance on the domains of speed of processing and verbal learning. Thus, Cluster 3 was labelled Superior/Near-Normal. Cluster 3 also performed significantly better than Cluster 2 on all domains except verbal learning. Cognitive performance comparisons for the clusters are provided in Figure 1a, and Table 1.

(2) Cluster performance on the alternative 'non-clustering' cognitive domain measures.

Small to moderate, highly significant correlations ($p < .001$) were evident between the clustering and alternative 'non-clustering' cognitive domains of speed of processing ($r = 0.37$), working memory ($r = 0.56$), reasoning and problem solving/executive function ($r = 0.52$), and the composite cognitive scores ($r = 0.50$). Table 2a and Figure 1b shows cross-diagnostic cluster and control performance on the alternative 'non-clustering' cognitive domain measures. The Globally Impaired cluster performed significantly worse than the Superior/Near-Normal cluster on the domains of working memory, executive function, and attention, with large effects ($d > 1.00$). The *Globally Impaired* cluster also performed worse than the Selectively Impaired cluster on the domains of working memory ($d = 0.76$) and executive function ($d = 1.68$), and worse than the healthy control group on all additional cognitive domains except speed of processing ($d = 1.15–2.07$).

The Selectively Impaired cluster performed significantly worse on the attention domain compared to controls ($d > 0.58$), although medium (non-significant) effects were evident for the other domains. This cluster did not significantly differ from the Superior/Near-Normal cluster on any of the additional domains, although qualitatively worse performance of medium effect was evident. Similarly, there were no significant differences between the Superior/Near-Normal cluster and healthy controls.

(3) Cluster differences in premorbid IQ, current global functioning, and estimated cognitive decline.

Table 2b and Figure 2a show cross-diagnostic cluster and control performance on premorbid IQ and composite scores. The Globally Impaired cluster had significantly lower

premorbid IQ scores than healthy controls (d = 1.20), the Selectively Impaired cluster (d = 0.77) and the Superior/Near-Normal cluster (d = 0.99). No other cluster differences in premorbid IQ were evident. In contrast, composite scores reflecting current global cognition were significantly lower in the Globally Impaired cluster compared to the other patient clusters and controls. The Selectively Impaired cluster score was also significantly lower than that of the Superior/Near-Normal cluster and controls. The Superior/Near-Normal cluster had significantly greater current global cognition than healthy controls. A similar pattern of findings was evident for the composite score of the 'non-clustering' cognitive measures (Table 2b).

Discrepancy score analysis indicated negative discrepancy scores for the Globally Impaired and the Selectively Impaired clusters, thereby suggesting evidence of 'decline' as indicated by worse 'current' cognition relative to estimated premorbid cognition. In contrast, the discrepancy score of the Superior/Near-Normal cluster was positive, indicating better 'current' cognition compared to estimated premorbid cognition. The discrepancy scores of all clusters differed from each other, with the largest discrepancy evident for the Globally Impaired cluster followed by the Selectively Impaired cluster. A graphical depiction of the cognitive discrepancy scores for the cross-diagnostic clusters are presented in Figure 2b, whilst values are presented in Table 2b.

Clinical characteristics of the cross-diagnostic clusters

The clinical characteristics of each cluster are detailed in Table 3. There were no significant differences between the emergent clusters in terms of age, gender, PANSS, MADRS, or YMRS scores. However, both the Globally Impaired and Selectively Impaired clusters were older than the control group. Consistent with previous cross-diagnostic cognitive clustering

studies, there was an over-representation of SSD relative to BD patients in the Globally Impaired cluster and an overrepresentation of BD relative to SSD patients in the Superior/Near-Normal cluster. Notably, the Selectively Impaired cluster had a relatively even diagnostic distribution (56/44% BD/SSD). There were no differences in the number of BD patients with a psychosis history between clusters and no differences in age of illness onset between clusters. However, illness duration was longer in the Globally Impaired compared to the Selectively Impaired cluster. More patients in the Globally Impaired cluster used atypical antipsychotics than in the Selectively Impaired cluster, while more patients in the Selectively Impaired cluster used atypical antipsychotics than patients in the Superior/Near-Normal cluster. There were no significant differences between clusters with regards to other medication types or total number of medications.

Discussion

This study extends previous cognitive subgrouping studies of BD and SSD by providing more detailed characterisation of cognitive variability in a cross-diagnostic sample of patients with these disorders. A hierarchical clustering analysis revealed evidence for three underlying cross-diagnostic subgroups, consistent with past work on the topic using the same method (Burdick et al., 2014; Van Rheenen et al., 2017). Relative to controls, two of these subgroups — named Globally Impaired and Selectively Impaired — showed deficits on the cognitive clustering variables used in the analysis. Deficits in the former subgroup were relatively severe and generalised across the clustering variables. In contrast, the latter subgroup showed equivalent verbal and visual memory performance to controls, but mild impairments in speed of processing, working memory, and reasoning and problem solving. The third subgroup outperformed controls

on the domains of visual learning, working memory, and reasoning and problem solving. This subgroup was labelled Superior/Near-Normal on account of equivalent performance on the other clustering variables. The cognitive performance of each subgroup rendered them largely statistically distinct in the context of the cognitive variables used to classify them, with the exception of the Selectively Impaired and Superior/Near-Normal subgroups on the verbal learning domain.

Although there were significant small-to-moderate correlations between the clustering and alternative 'non-clustering' domain measures, our prediction that the emergent subgroups would perform consistently across these measures was only partially met. That is, unlike the clear statistically significant subgroup differences observed for the majority of the cognitive measures used to classify the subgroups, the greatest separation of subgroups on the alternative 'non-clustering' cognitive measures was observed for working memory and executive function, in addition to attention. In turn, performance on the speed of processing domain did not significantly distinguish these subgroups from each other. There were also no significant subgroup differences between the Selectively Impaired and Superior/Near-Normal subgroups on any of the domains, contrary to initial primary clustering measure findings.

Importantly, the pattern of relatively severe working memory and executive function deficits across measures that were not used for clustering, and those that were, held in the Globally Impaired subgroup and extended to attention. However, this subgroup performed markedly better on the alternative cognitive domain measure of speed of processing that was not used in the initial clustering step, compared to its clustering domain measure counterpart. Further, the Selectively Impaired subgroup maintained a qualitatively intermediate level of performance similar to the previous analysis, but this subgroup did not significantly differ from

healthy controls on alternative cognitive domains of speed of processing, working memory and executive function (that were not used for the initial clustering).

Of relevance to these findings is the smaller correlation between the speed of processing domain measures (clustering vs alternative 'non-clustering) compared to the domain measures of executive function or working memory. Two of the three subtests used to create the *clustering* speed of processing domain measure (TMT-A and BACS-SC) also tap into motor behaviour more so than the subtests used to create the 'non-clustering' domain measure (colour and word naming trials of the CWIT). Thus, it is possible that the divergence in speed of processing findings could reflect the presence of a deficit in these subgroups that is more heavily weighted to *motor speed* dysfunction. However, an alternative, and perhaps more likely explanation is that the subtests encompassed within the alternative 'non-clustering' speed of processing domain — word reading and colour naming — are not sensitive enough to elicit the same magnitude of deficit as the clustering domain measures. This is because they draw on relatively automatic and overlearnt processes, while all of the other cognitive measures employed here do not. Indeed, the better separation of the subgroups on the 'non-clustering' working memory, executive function and attention domains may well be because these represent standard and sensitive alternative tests of these processes that are commonly used in the literature.

Importantly, the size and significance of deficits in *global cognition* as measured by a composite index based on measures that were used for clustering and those that were not, was consistent across subgroups. Although the significance levels in the analyses of indices not used in the initial clustering did not always support the primary clustering analysis findings, the general pattern regarding the direction of performance relative to controls on these cognitive domain measures (not used in clustering) was similar to that observed for those that were used in

the initial clustering step. That is, with the exception of speed of processing, the Superior/Near-Normal subgroup generally showed performance around a half a standard deviation of controls on all domains, and the Globally Impaired subgroup again showed the most severe decrements in performance of all subgroups. Thus, the consistency of subgroup performance did not appear to markedly differ on measures that were reasonably equivalent, but performance consistency did vary in the extent to which differences were statistically significant—a finding that may be at least partially related to reduced statistical power in the sample.

Premorbid IQ and estimated decline

As expected, analysis of the discrepancy between estimated premorbid and 'current' global cognition indicated that the Globally Impaired and Selectively Impaired subgroups potentially underwent cognitive 'decline' from premorbid levels, while the Superior/Near-Normal subgroup had a degree of improvement. This latter finding is very curious, and may be explained as a factor of our data standardisation to the control group. That is, any natural 'decline' in the controls might give the illusion that the Superior/Near-Normal subgroup had improvements in cognition where in reality they experienced minimal change. In light of this, the discrepancy findings in this subgroup may actually represent a relative *preservation* of cognitive functioning rather than a true improvement. However, this statement is very speculative given we were unable to statistically disentangle whether this was the case with the measures available to us. Thus, it should be considered with caution.

The difference in the patterns of estimated premorbid and 'current' cognitive functioning in our data for the other subgroups suggests that they may follow putatively divergent cognitive trajectories, akin to those evident in previous studies of SSD that incorporated premorbid IQ

measures into their clustering methodologies (Van Rheenen et al., 2018; Weickert et al., 2000; Weinberg et al., 2016; Woodward and Heckers, 2015). Indeed, the Globally Impaired subgroup had significantly lower premorbid IQ scores (>0.5 SDs) than healthy controls, and current cognitive performance approximately 2 SDs below controls. This subgroup may thus be one that is cognitively disadvantaged prior to illness onset with ongoing cognitive deterioration — akin to what has been considered to be a 'compromised' subgroup in past studies. The Selectively Impaired subgroup on the other hand, with a near normal level of estimated premorbid IQ but a significant, mild-moderate degree of 'current' cognitive impairment of approximately 0.75 SD's below healthy controls, may be one that is cognitively well-functioning at the outset of illness but declines after its onset — akin to what has been previously called a 'deteriorated' subgroup. Given the cross-diagnostic nature of our sample, the presence of such subgroups in our data appear to somewhat challenge common conceptualisations of all patients with SSD as evidencing a static course of cognitive impairment, and cognitive impairment in all patients with BD as following a neuroprogressive course secondary to the manifestation of the illness itself (Van Rheenen et al., in press; Woodward, 2016).

Clinical characterisation

As per previous research, the clustering procedure resulted in more BD participants assigned to the Superior/Near-Normal subgroup and more SSD participants to the Globally Impaired subgroup (Lewandowski et al., 2018; Lewandowski et al., 2014; Van Rheenen et al., 2017). Nonetheless, all diagnoses were represented in each subgroup. No significant differences were evident between the cross-diagnostic clusters with regards to age, gender distribution, current symptomology, psychosis history and age of illness onset. Thus, it seems a reasonable

inference that subgroup assignment in this data was largely driven by cognitive ability, rather than clinical characteristics (Green et al., In Press).

Notably, atypical antipsychotic use differed amongst the cognitive subgroups, with greater antipsychotic use among patients in the Globally Impaired (83%) than the Selectively Impaired (63%) and Superior/Near-Normal subgroups (50%). Although some meta-analytic evidence indicates that antipsychotic medication may negatively impact cognition, this appears to be specific to domains such as verbal learning rather than affecting cognition universally (Bourne et al., 2013). Relevantly, the subgroup difference in antipsychotic use was evident despite the absence of differences in symptom severity. The SSD cohort also had a higher rate of atypical antipsychotic use than the whole BD cohort. Thus, it is possible that the difference in atypical antipsychotic use between the subgroups is a function of the greater proportion of SSD patients assigned to the Globally Impaired subgroup than to the Superior/Near-Normal subgroup, who instead had the larger proportion of BD patients. The absence of differences in the proportion of patients in the different subgroups taking other medications or in terms of their total number of medications, further supports the notion that these subgroups are driven by general cognitive performance rather than other clinical factors.

Of further relevance, illness duration differences were evident between the subgroups, in that the patients in the Globally Impaired subgroup had experienced their illness for longer than those in the Selectively Impaired and Superior/Near-Normal subgroups. This difference may explain the larger degree of 'decline' seen in the Globally Impaired subgroup as a function of illness progression. There were no illness duration differences between the Selectively Impaired and Superior/Near-Normal subgroups, despite the 'decline' seen in the former. This pattern suggests that cognitive 'decline' may be exaggerated in patients who start out cognitively

'lower', a notion consistent with recent findings showing that SSD patients with low cognitive reserve—typically proxied by level of performance on premorbid IQ measures—are less able to buffer the adverse effects of illness or brain pathology on cognitive outcomes (de la Serna et al., 2013; Rolstad et al., 2016; Van Rheenen et al., (In Press)).

Limitations and conclusions

The findings of this study should be considered in the context of several limitations. First, our cross-diagnostic sample was relatively small for clustering analysis. Nonetheless, our data supported the emergence of three cross-diagnostic subgroups that were generally similar to those seen in previous work with larger samples. Second, our healthy control group was significantly younger than both SSD and BD cohorts. Although we covaried for age in our analyses, we cannot entirely exclude its potential effects on performance. Third, we did not have secondary measures of verbal or visual learning, and some of the alternative 'non-clustering' cognitive domains comprised only one measure. This may have reduced reliability compared to domains comprising multiple measures. Future research in this area should endeavour to more carefully equate measures of cognitive domains used for clustering with those not used in clustering, in terms of sensitivity and the number and breadth of component subtests. Fourth, inferences regarding cognitive 'decline' should be interpreted with caution, given there was less variability in premorbid IQ versus cognitive composite scores. Thus, it is possible that the properties of the measures themselves may be influencing the apparent subgroup differences or 'decline'. Finally, although we included data on the proportion of patients taking medications from different classes, we did not have data on dosage for all patients, which may have had an effect. Future

research should take into consideration medication dosage in cognitive subgrouping studies in these psychiatric populations.

In sum, the strengths of the study include the use of a data-driven approach to analysis, and the use of a detailed cognitive battery. To our knowledge, this is the first study to attempt to internally validate cross-diagnostic cognitive subgroups *in the same sample,* using alternative cognitive measures of the same domains that were used in the initial clustering steps. The general pattern of cognitive performance, in terms of the direction of impairment on individual domains and the composite measures indexed by both sets of cognitive measures (i.e., those used for the initial clustering and those that were not) was somewhat similar. However, the pattern of *statistically significant* differences (both to controls and each other) varied, particularly on the speed of processing domain where the Globally Impaired group showed substantially better performance on the 'non-clustering' compared to the clustering measure. While general differences in statistical significance are likely to related to issues of power, the speed of processing findings do raise the possibility that there could be *subtle* differences in patient assignment to subgroups depending on the types of speed of processing measures used in clustering analyses. However, we believe that discrepancies in performance are more likely due to the limited comparability between the different measures of speed of processing (as indicated by a correlation of just $r = 0.3$). Indeed, higher correlations between the global composites and domain scores of other cognitive domains were evident, with generally similar direction of performance deficits for all measures that were used in clustering, and those that were not. Taken in combination, these findings suggest that the subgroups evident here do not represent an artefact of the cognitive indices used to define them. Rather the subgroups may represent the

outcome of divergent cognitive trajectories, as indicated by differences in premorbid IQ and discrepancy scores. Further studies are required to determine if these results replicate.

References

Bourne, C., Aydemir, Ö., Balanzá-Martínez, V., Bora, E., Brissos, S., Cavanagh, J.T.O., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., Ferrier, I.N., Fleck, D.E., Frangou, S., Gallagher, P., Jones, L., Kieseppä, T., Martínez-Aran, A., Melle, I., Moore, P.B., Mur, M., Pfennig, A., Raust, A., Senturk, V., Simonsen, C., Smith, D.J., Bio, D.S., Soeiro-de-Souza, M.G., Stoddart, S.D.R., Sundet, K., Szöke, A., Thompson, J.M., Torrent, C., Zalla, T., Craddock, N., Andreassen, O.A., Leboyer, M., Vieta, E., Bauer, M., Worhunsky, P.D., Tzagarakis, C., Rogers, R.D., Geddes, J.R., Goodwin, G.M., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatrica Scandinavica 128, 149-162.

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. Psychol. Med. 44, 3083-3096.

Carruthers, S.P., Gurvich, C., Sumner, P.J., Rheenen, T.E.V., Rossell, S.L., 2019a. Characterising cognitive heterogeneity in schizophrenia spectrum disorders—A systematic review and narrative synthesis. Neuroscience & Biobehavioural Reviews.

Carruthers, S.P., Gurvich, C.T., Meyer, D., Bousman, C., Everall, I.P., Neill, E., Pantelis, C., Sumner, P.J., Tan, E.J., Thomas, E.H., Van Rheenen, T., Rossell, S., 2019b. Exploring Heterogeneity on the Wisconsin Card Sorting Test in Schizophrenia Spectrum Disorders: A Cluster Analytical Investigation. Journal of the International Neuropsychological Society, 1-11.

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. American Journal of Psychiatry 173, 373-384.

Dacquino, C., De Rossi, P., Spalletta, G., 2015. Schizophrenia and bipolar disorder: The road from similarities and clinical heterogeneity to neurobiological types. Clinica Chimica Acta 449, 49-59.

Davidson, J., Turnbull, C.D., Strickland, R., Miller, R., Graves, K., 1986. The Montgomery-Åsberg Depression Scale: reliability and validity. Acta Psychiatrica Scandinavica 73.

de la Serna, E., Andrés-Perpiñá, S., Puig, O., Baeza, I., Bombin, I., Bartrés-Faz, D., Arango, C., Gonzalez-Pinto, A., Parellada, M., Mayoral, M., Graell, M., Otero, S., Guardia, J., Castro-Fornieles, J., 2013. Cognitive reserve as a predictor of two year neuropsychological performance in early onset first-episode schizophrenia. Schizophrenia Research 143, 125-131.

Delis, D.C., Kaplan, E., & Kramer, J. H. , 2001. Delis Kaplan Executive Function System. The Psychological Corporation, San Antonio, TX.

Elwood, R.W., 1991. The Wechsler Memory Scale - Revised: Psychometric Characteristics and Clinical Application. Neuropsychology Review 2.

Green, M.J., Girshkin, L., Kremerskothen, K., Watkeys, O., Quidé, Y., In Press. A systematic review of studies reporting data-driven cognitive subtypes across the psychosis spectrum. Neuropsychology Review.

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.

Heaton, K.R., Chelune, G.J., Talley, J.L., Kay, G.G., Curtiss, G., 1993. Wisconsin Card Sorting Test Manual. Psychological Assessment Resources, Odessa, FL.

Heinrichs, R.W., Pinnock, F., Muharib, E., Hartman, L., Goldberg, J., McDermid Vaz, S., 2015. Neurocognitive normality in schizophrenia revisited. Schizophrenia Research 2, 227-232.

Holdnack, H.A., 2001. Wechsler Test of Adult Reading: WTAR. The Psychological Corporation, San Antonio.

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.

Krabbendam, L., Arts, B., van Os, J., Aleman, A., 2005. Cognitive functioning in patients with schizophrenia and bipolar disorder: a quantitative review. Schizophrenia Research 80, 137-149.

Lecrubier, Y., Sheehan, D.V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., Janavs, J., Dunbar, G.C., 1997. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 12, 224-231.

Lee, J., Rizzo, S., Altshuler, L., Glahn, D.C., Miklowitz, D.J., Sugar, C.A., Wynn, J.K., Green, M.F., 2017. Deconstructing Bipolar Disorder and Schizophrenia: A cross-diagnostic cluster analysis of cognitive phenotypes. Journal of Affective Disorders 209, 71-79.

Lewandowski, K.E., Baker, J.T., McCarthy, J.M., Norris, L.A., Ongur, D., 2018. Reproducibility of Cognitive Profiles in Psychosis Using Cluster Analysis. Journal of International Neuropsychological Society 24, 382-390.

Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, 3239-3248.

Martinez-Aran, A., Vieta, E., 2015. Cognition as a target in schizophrenia, bipolar disorder and depression. European Neuropsychopharmacology 25, 151-157.

Müller, M.J., Rossbach, W., Dannigkeit, P., Müller-Siecheneder, F., Szegedi, A., Wetzel, H., 1998. Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). Schizophrenia Research 32, 151-160.

Nuechterlein, K.H., Green, M.F., Kern, R.S., Baade, L.E., Barch, D.M., Cohen, J.D., Essock, S., Fenton, W.S., Frese, F.J., Gold, J.M., Goldberg, T., Heaton, R.K., Keefe, R.S.E., Kraemer, H., Mesholam-Gately, R., Seidman, L.J., Stover, E., Weinberger, D.R., Young, A.S., Zalcman, S., Marder, S.R., 2008. The MATRICS Consensus Cognitive Battery, Part 1: Test Selection, Reliability, and Validity. American Journal of Psychiatry 165, 203-213.

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.

Rolstad, S., Abe, C., Olsson, E., Eckerstrom, C., Landen, M., 2016. Cognitive reserve lessens the burden of white matter lesions on executive functions in bipolar disorder. Psychological Medicine 46, 3095-3104.

Russo, M., Van Rheenen, T.E., Shanahan, M., Mahon, K., Perez-Rodriguez, M.M., Cuesta-Diaz, A., Larsen, E., Malhotra, A.K., Burdick, K.E., 2017. Neurocognitive subtypes in patients with bipolar disorder and their unaffected siblings. Psychological Medicine 47, 2892-2905.

Sperry, S.H., O'Connor, L.K., Ongur, 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 International Neuropsychological Society 21, 468-472.

Tan, E.J., Rossell, S.L., 2014. Building a neurocognitive profile of thought disorder in schizophrenia using a standardized test battery. Schizophr. Res. 152, 242-245.

Van den Oord, E.J., Rujescu, D., Robles, J.R., Giegling, I., Birrell, C., Bukszar, J., Murrelle, L., Moller, H.J., Middleton, L., Muglia, P., 2006. Factor structure and external validity of the PANSS revisited. Schizophr Res 82, 213-223.

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., Fagerlund, B., Wannan, C., Bruggemann, J., , Lenroot, R.K., Pantelis, C., (In Press). Cognitive reserve attenuates age-related cognitive decline in the context of putatively accelerated brain ageing in schizophrenia-spectrum disorders. Psychological Medicine

Van Rheenen, T.E., Cropley, V., Zalesky, A., Bousman, C., Wells, R., Bruggemann, J., Sundram, S., Weinberg, D., Lenroot, R.K., Pereira, A., Shannon Weickert, C., Weickert, T.W., Pantelis, C., 2018. Widespread Volumetric Reductions in Schizophrenia and Schizoaffective Patients Displaying Compromised Cognitive Abilities. Schizophrenia Bulletin 44, 560-574.

Van Rheenen, T.E., Lewandowski, K.E., Bauer, I.E., Kapczinski, F., Miskowiak, K., Burdick, K.E., Balanzá-Martínez, V., in press. Current understandings of the trajectory and emerging correlates of cognitive impairment in bipolar disorder: an overview of evidence.

Van Rheenen, T.E., Lewandowski, K.E., Tan, E.J., Ospina, L.H., Ongur, D., Neill, E., Gurvich, C., Pantelis, C., Malhotra, A.K., Rossell, S.L., Burdick, K.E., 2017. Characterizing cognitive heterogeneity on the schizophrenia-bipolar disorder spectrum. Psychological Medicine, 1-17.

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.

Weinberg, D., Lenroot, R., Jacomb, I., Allen, K., Bruggemann, J., Wells, R., Balzan, R., Liu, D., Galletly, C., Catts, S.V., Weickert, C.S., Weickert, T.W., 2016. Cognitive Subtypes of Schizophrenia Characterized by Differential Brain Volumetric Reductions and Cognitive Decline. JAMA Psychiatry 73, 1251-1259.

Woodward, N.D., 2016. The course of neuropsychological impairment and brain structure abnormalities in psychotic disorders. Neuroscience Research 102, 39-46.

Woodward, N.D., Heckers, S., 2015. Brain Structure in Neuropsychologically Defined Subgroups of Schizophrenia and Psychotic Bipolar Disorder. Schizophrenia Bulletin 41, 1349-1359.

Youngstrom, E.A., Gracious, B.L., Danielson, C.K., Findling, R.L., Calabrese, J., 2003. Toward an integration of parent and clinician report on the Young Mania Rating Scale. Journal of Affective Disorders 77, 179-190.

Table 1. *Cross-diagnostic cognitive cluster and control performance on the clustering variables*

	Globally Impaired, *n*=30		**Selectively Impaired, n=32**		**Superior/Near Normal, n=24**		*f*	*df*	*p*	*Post-hoc (p, Cohen's d)*
a) Clustering variables	*M*	*SD*	*M*	*SD*	*M*	*SD*				
Speed of Processing	-1.64	0.92	-0.89	0.88	0.28	0.94	22.075	3	<.001	**Globally Impaired vs Selectively Impaired (.004, *d* = 0.74)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = 1.89)** **Selectively Impaired vs Superior/Near-Normal (<.001, *d* = 1.16)** **HC vs Globally Impaired (<.001, *d* = 1.53)** **HC vs Selectively Impaired (<.001, *d* = 0.81)** HC vs Superior/Near-Normal (.167, *d* = -0.32)
Working Memory	-1.33	0.77	-0.44	0.74	0.48	0.79	22.713	3	<.001	**Globally Impaired vs Selectively Impaired (<.001, *d* = 1.02)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = 2.11)** **Selectively Impaired vs Superior/Near-Normal (<.001, *d* = 1.08)** **HC vs Globally Impaired (<.001, *d* = 1.48)** **HC vs Selectively Impaired (.031, *d* = 0.47)** **HC vs Superior/Near-Normal (.011, *d* = -0.6)**
Verbal Learning	-1.17	0.81	-0.25	0.77	-0.13	0.82	11.446	3	<.001	**Globally Impaired vs Selectively Impaired (<.001, *d* = 1.03)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = 1.16)** Selectively Impaired vs Superior/Near-Normal (.633, *d* = 0.13) **HC vs Globally Impaired (<.001, *d* = 1.25)** HC vs Selectively Impaired (.264, *d* = 0.24) HC vs Superior/Near-Normal (.621, *d* = 0.11)
Visual Learning	-2.00	0.76	-0.40	0.73	0.48	0.77	47.705	3	<.001	**Globally Impaired vs Selectively Impaired (<.001, *d* = 1.9)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = 2.96)** **Selectively Impaired vs Superior/Near-Normal (<.001, *d* = 1.06)** **HC vs Globally Impaired (<.001, *d* = 2.26)** HC vs Selectively Impaired (.077, *d* = 0.38) **HC vs Superior/Near-Normal (.005, *d* = -0.66)**
Reasoning and Problem Solving	-1.29	0.80	-0.71	0.76	0.43	0.81	20.807	3	<.001	**Globally Impaired vs Selectively Impaired (.009, *d* = 0.66)**

	Globally Impaired, *n*=30	**Selectively Impaired, n=32**	**Superior/Near Normal, n=24**	*f*	*df*	*p*	*Post-hoc (p, Cohen's d)*
							Globally Impaired vs Superior/Near-Normal (<.001, *d* = 1.96) **Selectively Impaired vs Superior/Near-Normal (<.001, *d* = 1.3)** **HC vs Globally Impaired (<.001, *d* = 1.34)** **HC vs Selectively Impaired (.001, *d* = 0.7)** **HC vs Superior/Near-Normal (.013, *d* = -0.58)**

Note. a) Clustering domain scores from the MATRICS Consensus Cognitive Battery (MCCB); b) Non-clustering composite score. **Bolded** items are significant at $p < .05$ FDR corrected. *d*, Cohen's *d*. Means and SDs represent *z* scores standardised to healthy control values. All values are adjusted for age.

Table 2. *Cross-diagnostic cognitive cluster and control performance on non-clustering cognitive measures*

			Globally Impaired		Selectively Impaired		Superior/ Near Normal					Group comparisons
			M	*SD*	*M*	*SD*	*M*	*SD*	*f*	*df*	*p*	*Post-hoc (p, Cohen's d)*
a)	*Domain*	*Tests*										
	Speed of Processing	Stroop colour/word naming	-0.41	1.22	-0.54	1.21	0.07	1.20	1.542	3	0.207	
	Working Memory	Digit span forwards/backwards	-1.32	1.03	-0.53	1.02	0.08	1.01	9.681	3	<.001	**Globally Impaired vs Selectively Impaired (0.008, *d* = -0.76)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = -1.37)** Selectively Impaired vs Superior/Near-Normal (0.044*, *d* = -0.61) **HC vs Globally Impaired (<.001, *d* = 1.23)** HC vs Selectively Impaired (0.055, *d* = 0.48) HC vs Superior/Near-Normal (0.662, *d* = -0.11)
	Executive Function	Stroop CWIT Inhibition contrast score, TMT-B, WCST categories completed and perseverative errors	-1.93	0.95	-0.34	0.94	0.22	0.93	26.822	3	<.001	**Globally Impaired vs Selectively Impaired (<.001, *d* = -1.68)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = -2.29)** Selectively Impaired vs Superior/Near-Normal (0.044*, *d* = -0.60) **HC vs Globally Impaired (<.001, *d* = 2.07)** HC vs Selectively Impaired (0.103, *d* = 0.41) HC vs Superior/Near-Normal (0.487, *d* = -0.18)
	Attention^	CPT d'prime	-1.08	1.07	-0.47	1.06	0.06	1.06	7.639	3	<.001	Globally Impaired vs Selectively Impaired (0.044*, *d* = -0.57) **Globally Impaired vs Superior/Near-Normal (<.001, *d* = -1.07)** Selectively Impaired vs Superior/Near-Normal (0.092, *d* = -0.50) **HC vs Globally Impaired (<.001, *d* = 1.15)** **HC vs Selectively Impaired (0.021, *d* = 0.58)** HC vs Superior/Near-Normal (0.726, *d* = 0.09)
b)												
	Estimated premorbid IQ	WTAR	-0.67	0.60	-0.19	0.11	-0.07	0.12	7.512	3	<.001	**Globally Impaired vs Selectively Impaired (0.001, *d* = -0.81)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = -1.02)** Selectively Impaired vs Superior/Near-Normal (0.45, *d* = -0.21) **HC vs Globally Impaired (<.001, *d* = 1.00)** HC vs Selectively Impaired (0.359, *d* = 0.20) HC vs Superior/Near-Normal (0.985, *d* = -0.01)

		Globally Impaired		Selectively Impaired		Superior/ Near Normal					Group comparisons
		M	*SD*	*M*	*SD*	*M*	*SD*	*f*	*df*	*p*	*Post-hoc (p, Cohen's d)*
Composite scores	Clustering Variables Composite Score	-2.08	0.78	-0.75	0.78	0.43	0.76	58.395	3	<.001	**Globally Impaired vs Selectively Impaired (<.001, *d* = -1.70)** **Globally Impaired vs Superior/Near-Normal (<.001, *d* = -3.24)** **Selectively Impaired vs Superior/Near-Normal (<.001, *d* = -1.53)** **HC vs Globally Impaired (<.001, *d* = 2.54)** **HC vs Selectively Impaired (<.001, *d* = 0.86)** **HC vs Superior/Near-Normal (0.006, *d* = -0.65)**
	Non-Clustering Variables Composite Score	-1.07	1.10	-0.59	1.09	-0.13	1.08	6.605	3	<.001	Globally Impaired vs Selectively Impaired (<.122, *d* = -0.44) **Globally Impaired vs Superior/Near-Normal (<.001) , *d* = -1.10)** **Selectively Impaired vs Superior/Near-Normal (0.026, *d* = -0.67)** **HC vs Globally Impaired (<.001, *d* = 0.98)** **HC vs Selectively Impaired (0.03, *d* = 0.55)** HC vs Superior/Near-Normal (0.689, *d* = 0.10)
IQ Decline Discrepancy Score		-1.51	0.76	-0.66	0.76	0.41	0.7	41.197	2	<.001	**Globally Impaired vs Selectively Impaired (<0.01, *d* = -1.12)** **Globally Impaired vs Superior/Near-Normal (<.001 , *d* = -2.51)** **Selectively Impaired vs Superior/Near-Normal (<0.001, *d* = -1.40)**

Note: Table 2a) Cross-diagnostic cognitive cluster and control performance on non-clustering cognitive measures; Table 2b) Cross-diagnostic cognitive cluster and control performance on premorbid IQ, composite and declines scores (patients only). CPT, Continuous Performance Test; CWIT, colour-word interference test; WCST, Wisconsin Card Sorting Test; WTAR, Wechsler Test of Adult Reading. *Note*: Means and SDs represent *z* scores standardised to healthy control values. All HC means = 0 and standard deviations = 1. **Bolded** items are significant at $p < .05$ FDR corrected. * indicates items that were significant at an uncorrected level but did not survive FDR correction. All values are adjusted for age. ^ please note that data was missing for $n = 21$ participants.

Table 3. *Clinical descriptives and comparison between the cross-diagnostic cognitive clusters and healthy controls.*

	Globally Impaired, *n*=30			**Selectively Impaired, n=32**			**Superior/Near Normal, n=24**			**Healthy Controls (HC), n=76)**			**Group comparisons**			
	n (%)	*M*	*SD*	*n (%)*	*M*	*SD*	*n (%)*	*M*	*SD*	*n (%)*	*M*	*SD*	*f /χ2*	*df*	*p*	*Post-hoc (p)*
Age (years)		43.57	11.141		42	10.595		38.38	10.834		32.62	11.87	9.263	3, 158	<.001	Globally Impaired vs Selectively Impaired (.942) Globally Impaired vs Superior/Near-Normal (.320) Selectively Impaired vs Superior/Near-Normal (.598) **HC vs Globally Impaired (<.001)** **HC vs Selectively Impaired (.001)** HC vs Superior/Near-Normal (.135)
Gender (M/F)	12 (40)/ 18 (60)			16 (50)/ 16 (50)			15 (62)/ 9 (38)			39 (51.32)/ 37 (48.68)			2.7	2	0.259	
Diagnosis (BD/SSD)	8 (27)/ 22 (73)			18 (56)/ 14 (44)			15 (63)/ 9 (37)						8.365	2	0.015	
Proportion of diagnosis (BD%/SDD%)	(19)/(49)			(44)/(31)			(37)/(20)									
Age of onset (years)		21.28	9.32		24.44	11.69		21.21	6.75				0.936	2, 71	0.397	
Illness duration (years)^		24.19	12.35		15.60	10.71		17.17	11.98				3.904	2, 72	0.025	**Globally Impaired vs Selectively Impaired (.028)** Globally Impaired vs Superior/Near-Normal (.113) Selectively Impaired vs Superior/Near-Normal (.880)
PANSS Negative		13.73	5.76		12.22	4.90		11.08	3.76				1.958	2, 83	0.148	
PANSS Positive		16.80	5.39		14.56	6.39		13.46	5.76				2.315	2, 83	0.148	
PANSS General		33.37	10.58		31.22	9.75		30.46	9.79				0.629	2, 83	0.535	
PANSS Total		63.43	18.19		58	18.87		55	17.02		1.71	3.02	1.528	2, 83	0.223	
MADRS		10.87	11.33		9.42	9.22		8.25	7.30		1.24	1.83	0.508	2, 82	0.603	
YMRS		5.17	5.53		6.45	7.54		3.50	<.001				1.657	2, 82	0.197	

BD with psychosis history	6 (75)		12 (66.67)		8 (53.33)		32.62	11.87	4.748	2	0.093	
Medication												
Typical AP%	4 (13)		2 (6)		2 (8)				0.958	2	0.619	
Atypical AP%	25 (83)		20 (63)		12 (50)				6.954	2	0.031	Globally Impaired > Selectively Impaired > Superior/Near-Normal
Mood Stabiliser %	11 (36)		12 (38)		9 (38)				0.006	2	0.997	
Antidepressant %	11 (36)		8 (25)		6 (25)				1.29	2	0.525	
Benzodiazepine %	3 (10)		4 (13)		3 (13)				0.119	2	0.942	
Total number of medications		2 1		2 1		2 1			2.673	2, 83	0.075	

Note. **Bolded** items are significant at $p < .05$. PANSS = Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale; MADRS, Montgomery Asberg Depression Rating Scale; AP, antipsychotic; BD, bipolar disorder; SSD, schizophrenia spectrum disorder. ^ *Note* illness duration data was missing for $n = 13$ patients.

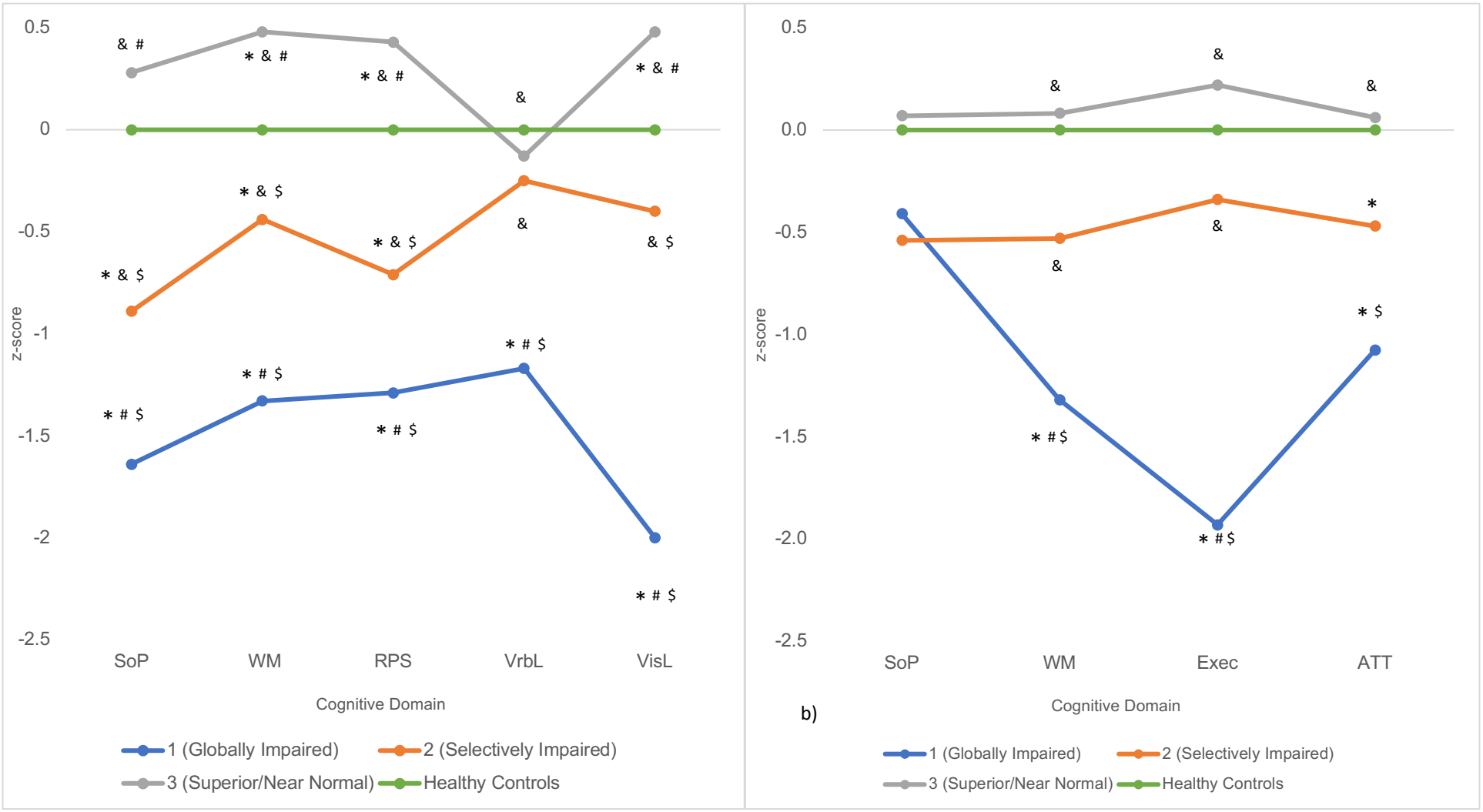


Figure 1. Cross-diagnostic cluster performance on the cognitive measures. a) performance on clustering variables (domain scores of MATRICS Consensus Cognitive Battery), including speed of processing (SoP), working memory (WM), reasoning and problem solving (RPS), verbal learning (VrbL), and visual learning (VisL); b) performance on non-clustering domains scores of speed of processing (SoP), working memory (WM), executive function (Exec), and attention (ATT). For significant differences on these measures refer to Table 1a. Data represents Z scores standardised to healthy control mean/SDs. Significant difference (after FDR correction) vs * = HC; $^{\&}$ = Globally Impaired ; $^{\#}$ = Selectively Impaired ; $^{\$}$ = Superior/Near-Normal; HC means = 0 and standard deviations = 1.

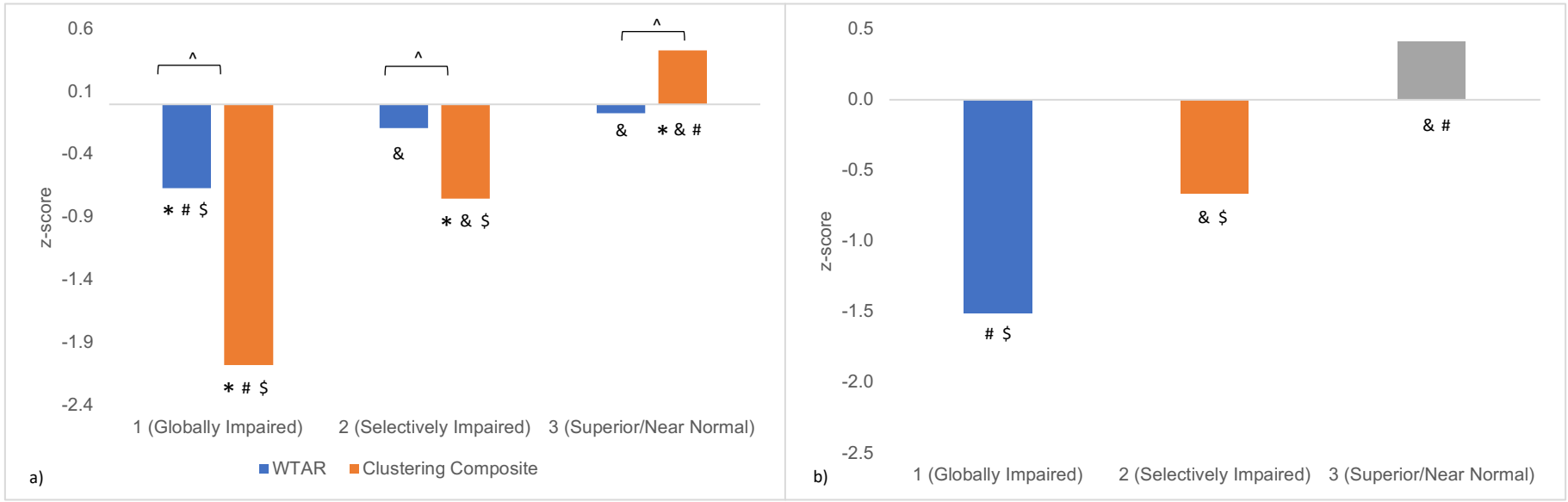


Figure 2. Cross-diagnostic cluster performance on measures of premorbid IQ, clustering composite and IQ decline. a) performance on premorbid IQ measure (Wechsler Test of Adult Reading) and performance on the *clustering composite* variable, consisting of the standardised sum of all five clustering variables; b) *discrepancy score*. Significant difference (after FDR correction) vs * = HC; $^{\&}$ = Globally Impaired; $^{\#}$ = Selectively Impaired; $^{\$}$ = Superior/Near-Normal; $^{\wedge}$ = significant difference between premorbid IQ and clustering composite score (supplementary material); HC means = 0 and standard deviations = 1.

Cognitive validation of cross-diagnostic cognitive subgroups on the schizophrenia-bipolar spectrum

Supplementary Material

James A. Karantonis [a,b], Susan L. Rossell [b,c], Sean P. Carruthers [a,b], Philip Sumner [b], Matthew Hughes [b,] Melissa J. Green[i,j], Christos Pantelis [a,d,e,f], Katherine E. Burdick[g,h],Vanessa Cropley [a,b] and Tamsyn E. Van Rheenen [a,b*]

[a] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Melbourne, Australia
[b] Centre for Mental Health, Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, Melbourne, Australia
[c] St Vincent's Mental Health, St Vincent's Hospital, VIC, Australia
[d] Florey Institute of Neuroscience and Mental Health, Australia;
[e] Department of Electrical and Electronic Engineering, University of Melbourne, VIC, Australia
[f] Centre for Neuropsychiatric Schizophrenia Research (CNSR) and Centre for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), Mental Health Centre Glostrup, Denmark
[g] Harvard Medical School, Department of Psychiatry, Boston, MA USA
[h] Brigham and Women's Hospital, Boston, MA USA
[I] School of Psychiatry, University of New South Wales (UNSW), Sydney, NSW, Australia
[j] Neuroscience Research Australia, Randwick, NSW, Australia

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

Additional information about methods and statistics.

Estimated cognitive decline scores

In addition to the main analysis, a further within-cross diagnostic cluster comparison was conducted to identify differences between estimated premorbid and 'current' cognition using paired sample t-tests.

Selection of cluster solutions

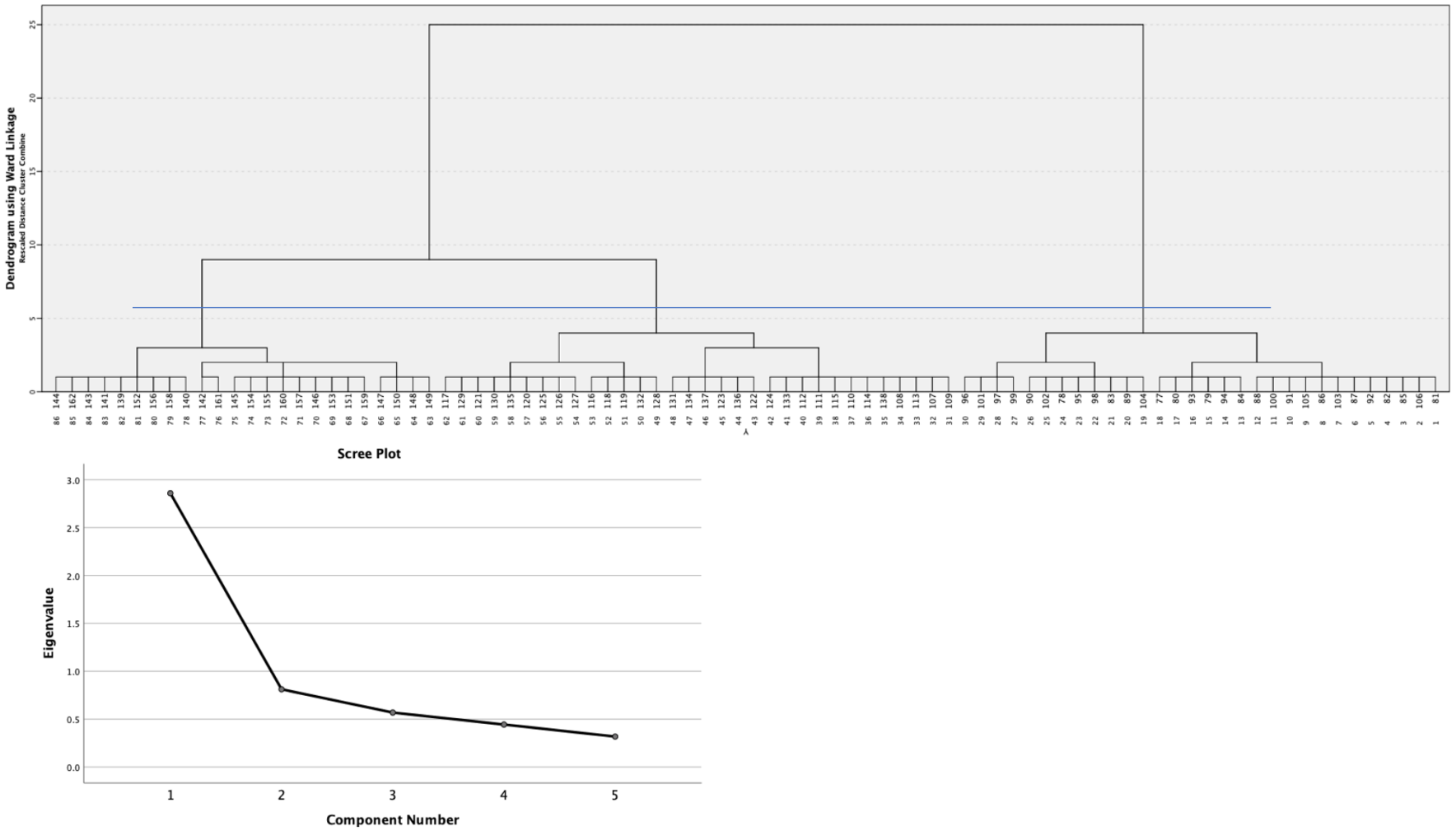
In the process of determining the cluster solution for the patient group, the analyses produced two, three and four-cluster models. Each model was further investigated in order to settle on the most meaningful clustering solution. Discriminant functions analyses (DFA) were conducted on all solutions, using a 'leave-one-out' classification.

The DFA suggested that the three-cluster model had the highest overall cross-validated classification accuracy (95.3%), followed by two-cluster (91.9%), and four-cluster (88.4%) models. Dendrograms (Figure S1) and scree plots from patient HCA were visually inspected by JK and TVR, and in combination with the DFA results (Figure S2), it was agreed unanimously that a three-cluster solution for the patient group was the best solutions for the data.

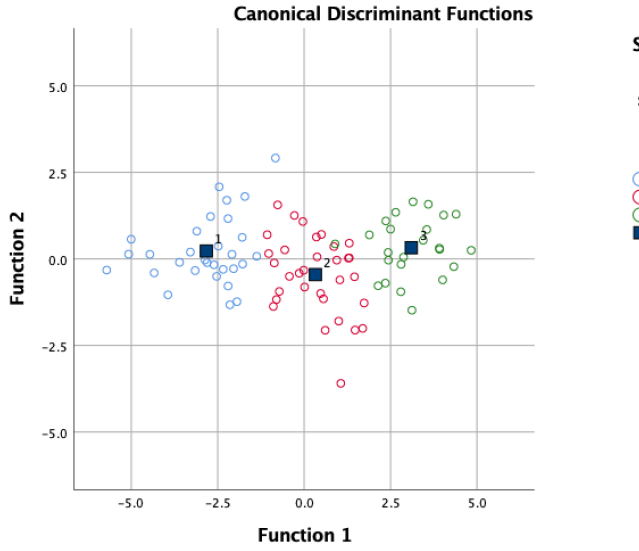
Preliminary analysis of traditional diagnostic categories

As with the primary analysis of cognitive subgroups, cognitive performance of BD, SSD, and healthy control groups were compared in a multivariate analysis of (co)variance (MANCOVA), controlling for age. For each separate cognitive analysis, a false

discovery rate (FDR) of 5% was set to correct for multiple comparisons at the omnibus level. Whenever the null hypothesis of equality across the emergent subgroups and controls was rejected, pair-wise post-hoc tests were performed and corrected using an FDR rate of 5% to determine where the group difference lay.



Supplementary Figure S1. Scree plot and dendrogram from cross-diagnostic hierarchical cluster analysis. The blue line indicates the agreed number of clusters by authors JK and TVR.



SoP, WM, Vrbl, Visl & RPS. Clustering settled on for Paper 1. HC are labelled cluster 4.

1
2
3
Group Centroid

Classification Results[a,c]

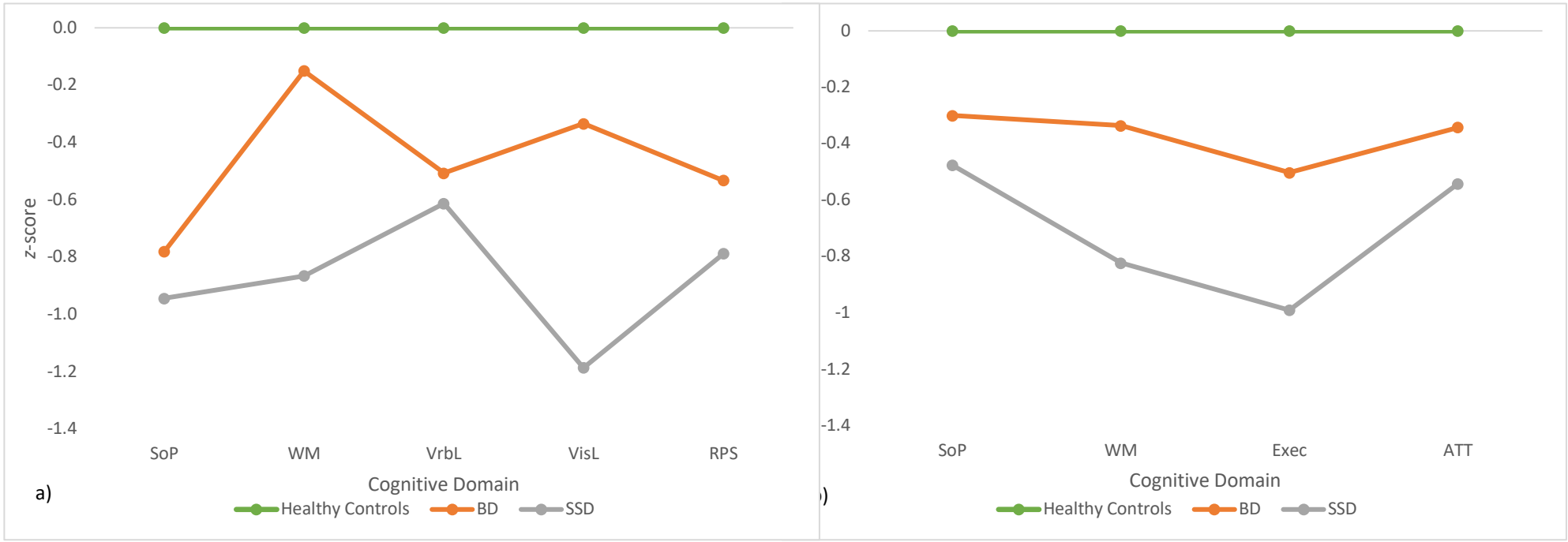
			Predicted Group Membership			
		MCCB_domains_3	1	2	3	Total
Original	Count	1	30	0	0	30
		2	0	32	0	32
		3	0	1	23	24
	%	1	100.0	.0	.0	100.0
		2	.0	100.0	.0	100.0
		3	.0	4.2	95.8	100.0
Cross-validated[b]	Count	1	29	1	0	30
		2	2	30	0	32
		3	0	1	23	24
	%	1	96.7	3.3	.0	100.0
		2	6.3	93.8	.0	100.0
		3	.0	4.2	95.8	100.0

a. 98.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 95.3% of cross-validated grouped cases correctly classified.

Supplementary Figure S2. Graph of canonical discriminant functions and classification results for the cross-diagnostic patient group (three-cluster solution).



Supplementary Figure S3. Cognitive performance of the BD, SSD, and healthy control sample on MCCB domains. *a)* Clustering domains*; b)* Non-clustering domains. *SoP*, speed of processing; *VrbL,* verbal learning; *VisL*, visual learning; *WM,* working memory; *RPS,* reasoning and problem solving; *Exec*, executive function; *Att*, attention; BD, bipolar disorder; SSD, schizophrenia spectrum disorders.

Supplementary Table S1. *Descriptives for the BD, SDD, and healthy control groups, prior to clustering analysis*

	BD (*n*=41)			**SSD (n=45)**			**Healthy Controls (*n*=76)**			**Group comparisons between BD, SSD and HC**			
	n (%)	*M*	*SD*	*n (%)*	*M*	*SD*	*n (%)*	*M*	*SD*	*f /χ2*	*df*	*p*	*Post-hoc*
Age (years)		39.49	11.24		43.40	10.41		32.62	11.87	13.790	2, 159	<.000	BD & SSD > HC
Gender (m/f)	24 (58.54)/ 17 (41.46)			19 (42.22)/ 26 (58.78)			39 (51.32)/ 37 (48.68)			2.312	2	0.315	
Age of onset (years)		23.84	10.00		20.81	8.83							
Illness duration (years)		15.24	12.33		22.82	10.89							
PANSS Negative		10.54	4.24		14.16	5.07							SSD > BD
PANSS Positive		11.61	4.52		18.16	5.43							SSD > BD
PANSS General		29.00	8.97		34.27	10.36							SSD > BD
PANSS Total		50.9	15.22		66.49	17.72				18.953	1, 84		SSD > BD
MADRS		6.00	4.62		12.95	11.52		1.71	3.02	12.992	1, 83	0.001	SSD > BD
YMRS		3.59	3.46		6.64	7.40		1.24	1.83	5.782	1, 83	0.018	SSD > BD
Psychosis history (BD)	26 (63.41)									19.942	1	<.000	SSD > BD
Medication													
Typical AP%	2 (4.88)			6 (13.33)						1.818	1	0.178	
Atypical AP%	22 (53.66)			35 (77.78)						5.584	1	0.018	SSD > BD
Mood Stabiliser %	22 (53.66)			10 (22.22)						9.074	1	0.003	BD > SSD
Antidepressant %	13 (31.7)			12 (26.67)						0.264	1	0.607	
Benzodiazepine %	4 (9.76)			6 (13.33)						0.267	1	0.605	
Total number of medications		2	1		2	1				0.548	1	0.461	

Note: PANSS, Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale; MADRS, Montgomery Asberg Depression Rating Scale; AP, antipsychotic; BD, bipolar disorder; SSD, schizophrenia spectrum disorder.

Supplementary Table S2. *BD, SSD, and healthy control cognitive performance and comparison*

	BD (*n*=41)		**SSD (n=45)**		**Healthy Controls (*n*=76)**		**Group comparisons between BD, SSD and HC**			
	M	*SD*	*M*	*SD*	*M*	*SD*	*f*	*df*	*p*	*Post-hoc (p, Cohen's d)*
a) Clustering variables										
Speed of Processing	-0.78	1.27	-0.94	1.28	0	1	6.956	2	0.001	**HC vs BD (0.004, *d* = 0.57)** **HC vs SSD (<.001, *d* = 0.65)** BD: vs SSD (0.689, *d* = -0.09)
Working Memory	-0.15	0.83	-0.87	1.01	0	1	8.275	2	<.001	HC vs BD (0.694, *d* = 0.08) **HC vs SSD (<.001, *d* = 0.75)** **BD: vs SSD (0.002, *d* = -0.69)**
Verbal Learning	-0.51	0.75	-0.61	0.99	0	1	4.614	2	0.011	**HC vs BD (0.019, *d* = 0.46)** **HC vs SSD (0.008, *d* = 0.53)** BD: vs SSD (0.733, *d* = -0.07)
Visual Learning	-0.33	1.12	-1.19	1.20	0	1	10.467	2	<.001	HC vs BD (0.368, *d* = 0.18) **HC vs SSD (<.001, *d* = 0.87)** **BD: vs SSD (0.001, *d* = -0.71)**
Reasoning and Problem Solving	-0.53	0.98	-0.79	1.11	0	1	3.961	2	0.021	HC vs BD (0.061, *d* = 0.37) **HC vs SSD (0.008, *d* = 0.52)** BD: vs SSD (0.445, *d* = -0.16)
b) Non-clustering domains and variables										
Processing Speed	-0.26	1.21	-0.37	1.25	0	1	0.793	2	0.455	
Attention^	-0.36	1.10	-0.71	1.13	0	1	5.377	2	0.006	HC vs BD (0.039*, *d* = 0.45) **HC vs SSD (0.002, *d* = 0.77)** BD: vs SSD (0.185, *d* = 0.32)
Working Memory	-0.36	1.07	-0.93	1.11	0	1	5.756	2	0.004	HC vs BD (0.187, *d* = 0.29) **HC vs SSD (0.001, *d* = 0.81)** **BD: vs SSD (0.029, *d* = 0.53)**
Executive Function	-0.48	1.14	-0.98	1.18	0	1	6.527	2	<.001	HC vs BD (0.046*, *d* = 0.44) **HC vs SSD (<.001, *d* = 0.86)** BD: vs SSD (0.076, *d* = 0.43)
Premorbid IQ										
WTAR	-0.32	0.91	-0.37	0.95	0	1	4.602	2	0.012	**HC vs BD (0.012, *d* = 0.55)** **HC vs SSD (0.009, *d* = 0.64)** BD: vs SSD (0.722, *d* = 0.09)
Composite scores										
Clustering Variables Composite Score	-0.66	1.04	-1.07	1.07	0	1	11.036	2	<.001	**HC vs BD (0.002**, *d* = 0.70**)** **HC vs SSD (<.001**, *d* = 1.08**)** BD: vs SSD (0.110, *d* = 0.38)
Non-Clustering Variables Composite Score	-0.39	1.13	-0.72	1.16	0	1	3.447	2	0.035	**HC vs BD (0.119, *d* = 0.34)** **HC vs SSD (0.011, *d* = 0.62)** BD: vs SSD (0.238, *d* = 0.28)
IQ Decline Discrepancy Score	-0.37	1.05	-0.92	1.11	-	-	3.986	1	0.059	

Note. a) Cognitive performance on domains from the MATRICS Consensus Cognitive Battery (MCCB); b) cognitive performance on non-clustering measures. **Bolded** items are significant at p <.05 FDR corrected. * indicates item did not survive FDR correction. Means and SDs represent z scores standardised to healthy control values. CPT, Continuous Performance Test; CWIT, colour-word interference test; WCST, Wisconsin Card Sorting Test; WTAR, Wechsler Test of Adult Reading; *d*, Cohen's *d*. All HC means = 0 and standard deviations = 1. All values are adjusted for age. ^ missing data n = 21.

References

Lewandowski, K. E., Baker, J. T., Mccarthy, J. M., Norris, L. A. & Ongur, D. (2018). Reproducibility of Cognitive Profiles in Psychosis Using Cluster Analysis. *Journal of International Neuropsychological Society,* 24**,** 382-390.

Van Rheenen, T. E., Lewandowski, K. E., Tan, E. J., Ospina, L. H., Ongur, D., Neill, E., Gurvich, C., Pantelis, C., Malhotra, A. K., Rossell, S. L. & Burdick, K. E. (2017). Characterizing cognitive heterogeneity on the schizophrenia-bipolar disorder spectrum. *Psychological Medicine***,** 1-17.