Title: Characterising the structure of cognitive heterogeneity in schizophrenia spectrum disorders. A systematic review and narrative synthesis.

Authors: *Sean P. Carruthers[a], Tamsyn E. Van Rheenen[a,b], Caroline Gurvich[c], Philip J. Sumner[a], , Susan L. Rossell[a,d]

*Corresponding author: Sean P. Carruthers; scarruthers@swin.edu.au

Affiliations:

[a]Centre for Mental Health, Faculty of Health, Arts and Design, Swinburne University of Technology, Victoria, 3122, Australia.

[b]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Carlton South, Victoria 3053, Australia.

[c]Monash Alfred Psychiatry research centre (MAPrc), Monash University Central Clinical School and The Alfred Hospital, Melbourne 3004 Australia.

[d]Psychiatry, St Vincent's Hospital, Melbourne, Victoria 3065, Australia.

Abstract

The aim of the present review was to systematically summarise our current understanding of the structure of the cognitive heterogeneity that exists within schizophrenia spectrum disorder (SSD). Fifty-two relevant studies were identified from January 1980 to March 2019 that investigated cognitive subgroups within SSD. Twenty-five studies employed classification criteria based on current neuropsychological function, 14 studies employed various data-driven subgrouping methodologies and 13 studies investigated putative cognitive symptom trajectories. Despite considerable methodological variability, three distinct cognitive subgroups reliability emerged; a relatively intact cognitive subgroup characterised by high cognitive performance, an intermediate cognitive subgroup defined by mixed or moderate levels of cognitive function/dysfunction and a globally impaired subgroup characterised by severe cognitive deficits. Whilst preliminary evidence suggests that these subgroups may have further investigative relevance in and of themselves, additional research is required and discussed. A set of reporting guidelines are also presented to overcome the methodological inconsistencies identified in the reviewed literature.

Keywords: Cognition, neuropsychology, intelligence, schizophrenia, schizoaffective disorder, psychosis, subgroup, cluster, cognitive trajectory, heterogeneity, global impairment, relatively intact, near-normal,

1. Introduction

Cognitive dysfunction is highly prevalent in schizophrenia spectrum disorder (SSD); with pervasive and widespread impairments relative to healthy controls (HC) well documented in the literature. Diminished intellectual and cognitive function at an early age, in addition to scholastic underperformance, are known risk factors for developing a SSD (Dickson et al., 2012; MacCabe et al., 2008); with meta-analytic evidence indicating that moderate-to-large cognitive deficits are present before the first psychotic episode. These deficits persist throughout the fluctuating course of psychotic symptoms experienced across the lifelong illness (Heinrichs, 2005; Heinrichs and Zakzanis, 1998; Mesholam-Gately et al., 2009), affect both overall and domain-specific functioning (Fioravanti et al., 2005; Heinrichs and Zakzanis, 1998; Schaefer et al., 2013; Tandon et al., 2009), and are said to be universally experienced, to some extent, by all with the disorder (Keefe et al., 2005; Tandon et al., 2009). Some have even advocated for the classificaiton of SSD as an illness of cognition (Kahn and Keefe, 2013).

More severe cognitive impairments have been associated with impoverished quality of life, compromised independent living skills and poorer occupational success in SSD (Bell et al., 2008; Dickerson et al., 2007, 2008; Harvey et al., 2012; Harvey and Strassnig, 2012). Evidence suggests that it is the cognitive symptoms that contribute to greater disability amongst individuals with SSD than the more recognisable and manageable positive symptoms (Harvey et al., 2010; Insel, 2010). Cognitive deficits in SSD are generally considered to be treatment-refractory pharmacologically, with only modest improvements in cognitive function, at best, observed during the course of antipsychotic treatment (Heinrichs, 2007; Keefe et al., 2007; Klingberg et al., 2008). Various remediation therapies have been

developed to treat cognitive dysfunction afflicting individuals with SSD, with only medium sized therapeutic effects reliably reported (Bowie et al., 2012; Fett et al., 2011; Wykes et al., 2011). The mitigation of cognitive impairment through more effective treatments over the next two decades has been predicted as an important step towards alleviating the debilitating burden of the illness for many individuals with SSD (Insel, 2010).

Akin to the clinical symptoms, considerable cognitive heterogeneity exists within the SSD population however. Growing evidence has indicated that within this heterogeneity, identifiable subgroups exist that are characterised by more homogenous cognitive profiles (Goldstein, 1990; Heinrichs and Awad, 1993; Sauvé et al., 2018). Numerous cognitive subgroups have been presented and characterised using a wide-variety of classification techniques and neuropsychological assessments (Green et al., 2013; Heinrichs et al., 1997; Vaskinn et al., 2014; Weickert et al., 2000). Preliminary evidence suggests that such cognitive subgroups may represent a meaningful differentiation within the SSD population, with overlapping subgroups independently shown to exhibit different clinical and functional outcomes, as well as distinctive brain structure (Gilbert et al., 2014; Van Rheenen et al., 2017; Weinberg et al., 2016; Wells et al., 2015). Despite the large body of evidence and general consensus supporting the notion that severe and widespread cognitive decline and/or dysfunction is universally experienced amongst those with the disorder, it is becoming increasingly apparent that several cognitive subgroups may exist within the SSD population.

The use of the subgrouping approach is beginning to foster a deeper understanding of the cognitive symptoms of SSD, which could ultimately lead to more effective treatment strategies. However, this emerging body of literature is limited by a high degree of methodological variability. Indeed, a broad range of study-specific *a priori*-defined classification criteria has been applied, with limited independent replication performed. Similarly, inherent limitations within data-driven subgrouping techniques restrict the cross-

study generalizability of the subgroups identified in each independent investigation. It is becoming clear that a comprehensive appraisal of the literature is required to establish a definitive and generalizable representation of the structure of cognitive subgroups within the SSD population. In this systematic review we aim to address this, by providing a synthesis of existing cognitive subgrouping studies in the SSD population as a means of characterising the structure of cognitive heterogeneity. Here, we seek to identify limitations in extant literature, with a view to provide methodological guidelines for research in this area and perspectives on ways in which our understanding of cognitive subgroups can be progressed.

2. Method

2.1. Search protocol

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009), where appropriate; and is registered at the International Prospective Register of Systematic Reviews (PROSPERO ID: CRD42018108473). The databases PubMed, Scopus and Web of Science were used to locate relevant literature, with searches limited to full-length empirical articles published between the 1st of January 1980 and 1st of March 2019. The search syntax employed was optimized for each database, with search terms based upon three concepts: schizophrenia spectrum disorders, cognition, and cognitive subgroups (see Appendix 1 for full search syntax). Reference lists of published reviews related to the topic and personal reference libraries of authors were checked for additional relevant works.

2.2. Study selection

Titles, abstracts and keywords of each record were initially scanned by a single reviewer (SPC) at stage one and were included if they met the following criteria: the article

was 1) written in English; 2) investigated cognitive subgroups as either a primary aim or auxiliary methodological aspect; and 3) included SSD or first episode psychosis (FEP) participants, with SSD being defined as schizophrenia or schizoaffective disorder. At stage two, two reviewers (SPC, SLR) independently screened full-text articles. A record was eligible if: 1) subgroups were based on non-social cognition; 2) cognitive performance was compared between subgroups; and 3) subgroups presented were comprised of only SSD participants and/or FEP participants who went on to develop SSD; that is, only retrospective FEP investigations which confirmed participant transition into SSD were eligible. Studies that included demographic or clinical variables into the subgrouping methodology were excluded. Studies employing single summary metrics of cognitive function that were comprised in-part by a measure of social cognition or a comprehensive neuropsychological test battery that included a social cognitive domain (e.g. the MATRICS Consensus Cognitive Battery [MCCB]; Nuechterlein et al., 2008) were eligible. The full-text articles of relevant cross-diagnostic investigations (e.g. schizophrenia-bipolar disorder) were screened for the reporting of any SSD-only results. Studies with overlapping samples were assessed on the extent of sample and methodological overlap; if studies with overlapping samples presented novel methodologies or findings (e.g. employing an alternative subgrouping method to identify different subgroups; providing additional subgroup characterisation above what was previously published) then both studies were deemed as eligible. To ensure that all methodologies and data was comprehensively available, only in-text results were considered eligible. Any discrepancies in the study selection between the two reviewers were discussed, with a third reviewer (PJS) included as an external adjudicator.

2.3. Data extraction and synthesis.

Data extraction was performed by a single reviewer. Extracted data for each study included: meta-study information (e.g. title, authors, aims, sample size); cognitive measures/variables assessed; subgrouping details (e.g. classification criteria, type of data-driven analysis, pre-processing); and subgroup characteristics (e.g. % of sample N, significant/non-significant pairwise comparisons on cognitive measures plus effect sizes). When possible, effect sizes were either extracted or calculated, if not presented, for all subgroup pairwise comparisons. The authors of three eligible studies provided unpublished data upon request (Guimond et al., 2016; Heinrichs et al., 2015; Sauvé et al., 2018).

3. Results

3.1. Search summary

The search strategy yielded a total of 2469 records, from which 52 studies met the inclusion criteria (see Figure 1 for PRISMA flow diagram; see Appendix 2). Included studies were divided into four collections based on the overarching subgrouping methodology employed: 1) classification criteria based on current cognitive function (n = 25 studies); 2) *a priori* defined/*k*-means iterative clustering (n = 3 studies); 3) exploratory data-driven techniques (n = 11 studies); and 4) putative cognitive symptom trajectories (n = 13 studies). Due to the considerable methodological variability, meta-analyses were not performed, and a narrative synthesis of the results is provided below.

3.2. Classification criteria based on current cognitive function

Twenty-five of the eligible studies were identified as using *a priori* defined classification criteria to subgroup participants based on their current neuropsychological function. These studies have been grouped together and reviewed based on the

neuropsychological domain(s) used to subgroup participants. Details about the individual studies are reported below, with a results summary presented in Table 1.

3.2.1. Intellectual functioning

Five studies classified participants into cognitive subgroups based on current intellectual functioning. Overall, 22% of the SSD participants met the various criteria to be classified as having high IQ and were reported to exhibit control-equivalent performance on the majority of cognitive measures assessed. Although, deficits in memory and executive function were prevalent across multiple studies. High IQ SSD participants were consistently reported to outperform the remaining SSD subgroups on the majority of measures assessed. Details about the individual studies are reported below.

Utilising the Wechsler Abbreviated Scale of Intelligence (WASI) II to estimate current IQ, Vaskinn et al. (2014) classified 239 SSD and 228 HC participants as being either intellectually low (IQ: 80-95), normal (IQ: 100-115) or superior (IQ $\geq$ 120) and compared diagnostic differences in neuropsychological profiles at each IQ level. Their analyses revealed that the performance profile of SSD participants was consistently below that of the HCs at all three IQ levels. For the low (d_{μ} = 0.67; $d_{\text{range:}}$ 0.10 to 1.21), normal (d_{μ} = 0.71; $d_{\text{range:}}$ 0.34 to 1.09) and superior (d_{μ} = 0.77; $d_{\text{range:}}$ 0.61 to 1.11) IQ levels, moderate-to-large effects were calculated between the SSD and HC subgroups on the eight measures of executive function, processing speed, verbal memory, fluency and working memory included in the profile analysis. Whilst the cognitive profile of the superior IQ SSD subgroup was below that of the control-equivalent, Vaskinn et al. (2014) reported that their performance on all eight measures fell within the normal range (z-score within one standard deviation (s.d.) of a sizeable reference HC sample).

Ayesa-Arriola et al. (2018) employed the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) III as an estimate of premorbid IQ to classify 292 FEP and 199 HC participants as being either intellectually low (premorbid IQ < 90), normal (premorbid IQ: 90-100) or high (premorbid IQ > 110). No significant FEP-HC differences between the two high premorbid IQ, as well as the two normal premorbid IQ subgroups on eight measures of attention, executive functioning, motor dexterity, processing speed, verbal and visual learning, working memory and overall cognitive ability. In contrast, the low premorbid IQ FEP subgroup performed significantly worse than the low premorbid IQ HC subgroup on six of the eight domains assessed; with the two subgroups not significantly differing in visual learning and working memory performance. The three FEP subgroups did not significantly differ in their attentional capacity or motor dexterity, with the high and normal premorbid IQ subgroups also exhibiting paralleled performance on the remaining six domains. The low FEP subgroup exhibited significantly impaired performance on the remaining six domains relative to both the high and normal FEP subgroups.

Focussing on current IQ, both Kremen et al. (2001) and Ruiz et al. (2007) administered a shortened four-subtest version of the WAIS-R (block design, digit span, digit symbol and vocabulary) and WAIS-III (arithmetic, block design, digit symbol, information) to their respective samples, whilst Wilk et al. (2005) administered the WAIS-III in its entirety. Kremen et al. (2001) classified 75 SSD participants and 91 HCs into one of two subgroups representing average (95-110) or low-average (81-94) IQ, with Ruiz et al. (2007) categorising 44 SSD participants as having either high or low IQ based on a cut-off of 85 points or more. Whereas Wilk et al. (2005) examined the WAIS-III performance of 97 SSD and 95 HC participants to identify 13 from each group with high-average IQ (≥ 100).

Kremen et al. (2001) reported that 20 SSD participants with average IQ performed as well as 21 IQ-matched HCs in their executive motor and general visuospatial ability, mental

encoding capacity and perceptual-motor speed. Significant deficits in executive function, sustained attention and verbal declarative memory were reported for the SSD subgroup. Sixteen low IQ SSD participants were reported to exhibit significantly impaired executive function, sustained attention, perceptual-motor speed and verbal declarative learning ability relative to 15 IQ-matched HCs. The two low IQ subgroups did not significantly differ in their executive motor control, problem-solving and general visuospatial ability.

Ruiz et al. (2007) documented that the 21 SSD participants classified as having high IQ displayed intact performance on the arithmetic, information and picture completion subtests of the WAIS-III, as well as overall success on the Wisconsin Card Sorting Test (WCST), all Rivermead Behavioural Memory Test (RBMT) indices except for the overall memory profile score and two novel measures of implicit memory relative to 35 HCs. Significant impairments in executive function, implicit memory and processing speed were also reported amongst the high IQ SSD subgroup. In contrast, 23 low IQ SSD participants performed significantly worse than controls on 15 out of the 20 variables assessed; with control-equivalent memory, verbal and visual learning detected. Wilk et al. (2005) reported that their high-average IQ SSD subgroup exhibited control-equivalent performance on the general memory, processing speed, verbal comprehension and working memory outcome indices of the WAIS-III and Wechsler Memory Scale (WMS) III. The SSD participants outperformed the IQ-matched HCs on the perceptual organization index, however displayed significantly impaired immediate memory.

Only Ruiz et al. (2007) compared cognitive performance across SSD subgroups, with the high IQ participants found to exhibit significantly better performance on the arithmetic, digit span, digit symbol, information and picture completion subtests of the WAIS-III and better overall performance on the WCST compared to the low IQ participants. The two

subgroups did not significantly differ in cognitive flexibility, verbal fluency and on several indices of everyday and implicit memory/learning ability.

3.2.2. Neuropsychological functioning

Eleven studies classified participants into cognitive subgroups based on their performance on a test battery comprised of multiple cognitive domains. In seven of the 10 studies, SSD subgroups were defined based on participant performances on individual cognitive domains, though the particular domain used were different across studies. In two of the 10 studies, SSD subgroups were defined based on single summative indices across multiple domains. The remaining two studies, SSD subgroups were defined based on both performance within individual domains and overall performance Furthermore, in nine studies, SSD participants were dichotomized into normal and impaired subgroups, whereas in the remaining two studies further differentiating cognitively impaired SSD participants into a generalized impairment subgroup and a moderately impaired or study-specific domain-deficit subgroups. On average, 30% of SSD participants met the various criteria to be classified as having normal cognitive function or falling within normal limits. The normal/within normal limits subgroups were consistently reported to outperform the remaining SSD participants and as exhibiting control-equivalent performance on the majority of cognitive measures assessed. Impaired processing speed was commonly reported amongst the normal/within normal limits subgroups relative to HCs. Details about the individual studies are reported below.

Ortiz-Gil et al. (2011) classified preserved cognitive ability based on scores above the fifth percentile for normal adults on the RBMT screening score and the Behavioural Assessment of the Dysexecutive Syndrome (BADS) profile score. Twenty-three SSD participants met the criteria for preserved cognitive function and significantly outperformed

the remaining 26 impaired SSD participants on both the RBMT screening score and BADS profile score used to classify the participants, with no other significant differences reported. Both SSD subgroups exhibited the same profile of performance relative to 39 HCs; that of being significantly impaired in IQ and reasoning ability, with control-equivalent premorbid IQ and verbal comprehension capacity.

Heinrichs et al. (2008) used the vocabulary subtest of the WAIS III to classify 151 SSD and 72 HC participants as having either superior verbal ability (vocabulary score ≥ 14) or not (vocabulary score < 14). No significant differences were reported between 25 SSD and 22 HC verbally superior participants on nine measures of attention, current IQ, fluency, premorbid IQ, processing speed, reasoning, verbal learning and working memory. Similarly, no significant differences on seven of the nine measures were reported between the verbally superior SSD participants and 50 comparison HCs; with the verbally superior SSD subgroup exhibiting significantly better premorbid IQ and WAIS vocabulary performance. One hundred and twenty-six comparison SSD participants were reported to exhibit deficits on eight of the nine measures assessed in comparison to the verbally superior HC subgroup, with a non-significant difference in attention detected. Conversely, the comparison SSD subgroup performed as well as the comparison HCs on seven measures, with significant deficits in verbal learning and processing speed reported. In comparing cognitive performance amongst the two SSD subgroups, Heinrichs et al. (2008) reported that the verbally superior SSD participants significantly outperformed the comparison participants on eight of the nine measures administered, with both subgroups accumulating a similar number of errors on a measure of sustained attention.

Extending beyond verbal ability, Ammari et al. (2010) assigned 154 SSD participants into one of three cognitive subgroups based on their performance on the vocabulary and matrix reasoning subtests of the WAIS-III and total immediate recall on the California Verbal

Learning Test (CVLT). Using previously published standards, Ammari et al. (2010) defined cognitive normality as scoring between 8-13 on the WAIS-III subtests and 43-60 on the CVLT; impaired verbal-memory only as scoring between 8-13 on the WAIS-III subtests and less than 43 on the CVLT; and generalized impairment as scoring less than 8 on the WAIS-III subtests and 43 on the CVLT. Twenty-four SSD participants were classified as cognitively normal and exhibited equivalent performance on nine measures of attention, fluency, processing speed and verbal learning to that of 18 HCs. Twenty-six SSD participants were classified as verbal memory-impaired and exhibited an HC-equivalent performance on seven of the nine measures; with significant impairments in verbal learning and working memory present. Twenty-three SSD participants were classified as having generalized cognitive impairment and exhibited significant deficits relative to the HC group on eight measures, with a HC-equivalent performance on an index of attention detected. In comparing the SSD subgroups, the cognitively normal and verbal memory impaired participants only significantly differed in total immediate recall on the CVLT. In contrast, the generalized impairment subgroup exhibited significant deficits in fluency, reasoning and problem solving, verbal ability, as well as working memory compared to both remaining SSD subgroups, with an additional verbal learning deficit relative to the cognitively normal SSD participants present.

Four of the included studies classified participants into cognitive subgroups based on their performance on each of the individual domains comprising more comprehensive cognitive batteries than that employed above. Moses (1983) assessed 100 SSD participants on all 14 subscales of the Lubria-Nebraska Neuropsychological Battery (LNNB) to classify them as having either normal (0-2 subscales > critical level), borderline (3-4 subscales > critical level) or abnormal cognitive function (≥ 5 subscales > critical level) based on an unspecified LNNB critical level. Kremen et al. (2000) classified 75 SSD participants as being either

neuropsychologically within normal limits or neuropsychologically abnormal following their performance on a battery comprised of eight cognitive domains. Within normal limits was defined as having less than two domain scores less than or equal to two s.d. below the mean performance of 91 HCs; whilst the neuropsychologically abnormal subgroup was characterised by having two or more domain scores below the two s.d. threshold or one or more domain scores greater than three s.d. below the HC mean. In a follow-up study, Kremen et al. (2004) reclassified the 58 SSD participants labelled as neuropsychologically abnormal into one of four distinct impairment profile types based on patterns typically associated with common neuropsychological syndromes; verbal memory, abstraction, diffuse or other. And finally, Holthausen et al. (2002) assessed 118 FEP participants on a composite battery comprised of nine cognitive domains to consequently classify participants as being either cognitively normal or impaired. Participants were labelled as cognitively normal after having no domain sores greater than two s.d. below the mean performance of 45 HCs; with the impaired subgroup characterised by having one or more domain scores above the threshold.

Whilst Moses (1983) did not have a HC comparison group, the 17 SSD participants classified by Kremen et al. (2000) as neuropsychologically within normal limits exhibited control-equivalent performance on six measures of attention, motor-control, verbal comprehension, verbal learning, visuo-spatial ability and working memory; however significant impairments in executive function and processing speed were reported. In comparison, the 58 neuropsychologically abnormal SSD participants performed significantly worse than HCs on all eight domains assessed. According to Holthausen et al. (2002), 23 cognitively normal FEP participants exhibited control-equivalent performance on seven measures of attention, fluency, visual learning and visuoconstruction; however significant processing speed and verbal learning deficits were reported. The 95 cognitively impaired FEP

participants were found to exhibit significant impairments on all nine cognitive domains relative to HCs.

Moses (1983) did report that the 69 normal SSD particpants outperformed the 16 borderline and 15 abnormal SSD participants on nine and all 14 LNNB subscales respectively. The borderline subgroup exhibited arithmetic, memory and reading abilities that were equivalent to the normal subgroup, whilst the abnormal subgroup exhibited more pronounced impairments on the arithmetic, expressive and receptive speech, right hemisphere and writing LNNB subscales relative to the borderline subgroup. The 17 SSD participants classified by Kremen et al. (2000) as being neuropsychologically within normal limits exhibited significantly better performance on all eight cognitive domains assessed relative to the 58 neuropsychologically abnormal SSD participants. Following a re-classification of the neuropsychologically abnormal subgroup, Kremen et al. (2004) reported significant subgroup-by-performance interactions across the four impairment types for each of the eight cognitive domains assessed. Whilst no pairwise comparisons amongst the four subgroups were reported, considerable cognitive profile overlap is depicted graphically for the verbal memory, abstraction and other subgroups, with a distinctive deficit visually apparent for the diffuse subgroup. In contrast, Holthausen et al. (2002) reported that their cognitively normal FEP participants exhibited significantly better performance on all nine of the composite cognitive domains assessed relative to the cognitive impaired participants

In contrast to the above studies, Muharib et al. (2014) and Heinrichs et al. (2015) used only a global summary index to subgroup two overlapping samples. Normal range cognition was defined as having an MCCB overall composite score equal to or greater than 40, with an average of 16% of the combined SSD sample meeting this criterion. The subgroups were consistently reported to exhibit intact premorbid IQ and performance on seven of the eight MCCB domains relative to two overlapping age, gender and overall composite score matched

HC comparison groups; a non-significant difference in current IQ was also detected by Muharib et al. (2014). Unlike their collaborators, Heinrichs et al. (2015) documented a significant processing speed impairment amongst their SSD subgroup, with the two groups also reported not to significantly differ in performance on a task of probabilistic reasoning and two adjunct measures of social cognition. Sixteen percent of the combined sample was, on average classified as having below normal ranged/impaired cognition. Muharib et al. (2014) reported that the cognitively impaired subgroup exhibited significant deficts on all 10 variables assessed relative to the same HC group matched to the normal range cognition SSD subgroup. In contrast, Heinrichs et al. (2015) compared the performance of the cognitively below normal ranged SSD subgroup against an age, gender and overall composite score matched HC comparison group and reported no-significant performance differences on all 12 variables assessed.

Muharib et al. (2014) reported that the normal ranged SSD subgroup outperformed the cognitively impaired SSD subgroup in premorbid and current IQ estimates, as well as on the seven non-social cognitive domains of the MCCB. Whilst Heinrichs et al. (2015) did not report any significant testing for comparisons on the MCCB domains in this instance, large effects in favour of the cognitively normal subgroup were calculated for all seven MCCB cognitive domains (d_{μ} = 2.15; $d_{\text{range:}}$ 0.64 to 4.20). Heinrichs et al. (2015) did report that the normal ranged SSD participants exhibited significantly better performance on all three of the four adjunct measures compared to their SSD counterparts.

Palmer et al. (1997) and González-Blanch et al. (2010) classified participants into cognitive subgroups using both individual domain and global deficit summary score performance. Palmer et al. (1997) employed two neuropsychologists to blindly rate the performance of 171 SSD participants on a composite battery comprised of eight cognitive domains, in addition to a global deficit summary index. Both neuropsychologists rated each

participant on the eight cognitive domains and global deficit index using a 9-point scale from 1 (above average) to 9 (severe impairment) based on raw test performance and demographically adjusted *t*-scores. Neuropsychological normality was defined as being rated to have both a global deficit index score of no greater than five, and no more than one domain rating greater than five; with neuropsychological impairment based on the global deficit index and two or more domain ratings greater than five. Comparatively, González-Blanch et al. (2010) examined the performance of 111 FEP particpants on a composite battery of six cognitive domains, in addition to a global deficit score. The global deficit score in this instance was reported as an approach used for summarizing overall performance and was computed by converting test scores into an overall standardized deficit score ranging from 0 (no impairment) to 5 (severe impairment). Cognitive normality was defined as exhibiting no more than one cognitive domain score one s.d. below the mean performance of 28 HCs and having a global deficit score either less than or equal to 0.5. In contrast, cognitively impaired was defined as having more than one domain score one s.d. below the HC mean and a global deficit score greater than 0.5.

Relative to 63 HCs, Palmer et al. (1997) reported that the 44 neuropsychologically normal SSD participants exhibited intact performance in attention, executive function, information retention, psychomotor ability, sensory perception, verbal comprehension and global functioning; with only a significant memory deficiency detected. González-Blanch et al. (2010) stated that the 25 cognitive normal FEP participants paralleled the performance of 28 HCs on the six domains of attention, executive function, processing speed, verbal comprehension, verbal and visual learning, as well as the global deficit score; with a significant impairment in motor dexterity detected. Both Palmer et al. (1997) and González-Blanch et al. (2010) respectively found that the 124 neuropsychologically impaired SSD and 86 cognitively impaired FEP participants performed significantly worse than HCs on all

domains assessed. Similarly, both studies reported that the neuropsychologically normal/cognitive normal subgroups significantly outperformed the neuropsychologically impaired/cognitive impaired subgroups on all domains assessed.

3.2.3. Memory specific

Seven studies classified participants into cognitive subgroups based on memory. Two studies employed a composite memory-based battery to subgroup participants, two studies assessed list learning ability, whilst the three remaining studies used various CVLT indices to subgroup participants. On average, 20% of SSD participants were labelled as having normal/unimpaired memory ability, whilst the remaining SSD participants were classified as either impaired/below normal or having a specific memory deficit. Overall, the SSD participants with apparent normal memory exhibited intact performance in verbal learning ability relative to HCs. Mixed patterns of significant and non-significant performance differences between SSD subgroups were evident, with the majority of comparisons on non-memory-based measures reported as non-significant. Details about the individual studies are reported below.

After administering a battery of memory-based tests to a sample of 90 SSD participants, Torres et al. (1997) simply compared the 10 best and worst performers. The high memory subgroup was reported to have significantly higher IQ and outperform the low memory subgroup on seven measures of verbal and non-verbal memory. Subgroups did not significantly differ in verbal fluency. Wexler et al. (2009) classified 81 SSD participants as being either neuropsychologically near-normal or neuropsychologically impaired based on their performance on two verbal and two non-verbal serial position working memory tasks. Twenty-one SSD participants were initially assigned to the near-normal subgroup, however only a subsample of 14 participants were selected to provide an apparent validation of the

discriminate ability of the classification method devised by the authors. The limited subsample of near-normal SSD participants exhibited sustained attention and verbal learning abilities that paralleled 22 HCs. The near-normal subgroup exhibited significantly better immediate verbal learning compared to 36 neuropsychologically impaired SSD participants. The neuropsychologically impaired subgroup performed significantly worse than HCs on both of the two measures assessed.

Guimond et al. (2016) classified 50 SSD participants into cognitive subgroups based on their performance on the International Shopping List Task (ISLT). Twenty-three SSD participants were labelled as having low-to-mild verbal memory impairment after exhibiting a z-standardized performance on the ISLT that was no more than 1.4 s.d. below that of 23 HCs; with the remaining 27 SSD participants classified as having moderate-to-severe verbal memory impairments (ISLT z-score ≤ -1.4 HC mean). Both SSD subgroups were reported as having significant current IQ deficits relative to the HCs, whilst also exhibiting intact verbal learning performance. The moderate-to-severe SSD subgroup were significantly impaired in their executive function and verbal memory ability in comparison to the low-to-mild verbal memory impairment subgroup; with the two SSD subgroups not significantly differing on the remaining five domains of the Cogstate battery.

Using a comparable word list learning task from the CERAD Neuropsychological Battery, Bowie et al. (2004) classified 589 SSD participants as having either unimpaired, subcortical or cortical learning patterns based on their immediate recall, delayed recall and recognition abilities. The three cognitive impairment profiles were based on earlier research into dementia and which had previously been applied to schizophrenia research (see Paulsen et al., 1995; below). Twenty-three SSD participants were labelled as having an unimpaired learning pattern after their z-standardized performance on each of the three verbal learning measures fell within one s.d. of normative HC data; 156 participants were classified as

having a subcortical learning pattern after their z-standardized immediate and delayed recall performance was more than one s.d. below normative HC data, in addition to a recognition performance within one s.d. of normative HC data; and 238 participants were classified as having a cortical learning pattern after their performance on all three measures fell below one s.d. of the normative HC data. The unimpaired subgroup was reported to significantly outperform both the subcortical and cortical subgroups on seven measures of fluency, memory, verbal comprehension and verbal learning. Similarly, the subcortical subgroup significantly outperformed the cortical subgroup on all seven domains.

Three studies classified participants into cognitive subgroups based upon their performance on various indices of the CVLT. McDermid Vaz and Heinrichs (2006) and Heinrichs et al. (2017) both used the CVLT to assess verbal learning, whereas Vaskinn et al. (2008) classified SSD participants into subgroups based on participant learning potential. From a sample of 55 SSD participants, McDermid Vaz and Heinrichs (2006) classified 34 as being memory unimpaired after their cumulative immediate recall performance fell within two s.d. of normative HC data; with the remaining 68 SSD participants labelled as memory impaired. Similarly, Heinrichs et al. (2017) classified 26 SSD participants as having normal-range verbal-declarative memory after performing within one s.d. of the z-standardized average of 51 HCs on each of the eight CVLT outcome indices assessed. The remaining 129 SSD participants who failed to meet this criterion were classified as having below normal range verbal-declarative memory. Vaskinn et al. (2008) classified SSD participants into subgroups based on their learning potential using age- and gender-corrected z-scores calculated from normative HC data. Twenty-four SSD participants were labelled as high-achievers after exhibiting a z-score greater than zero on the first immediate recall trial of the CVLT; 61 participants were labelled as learners after exhibiting a z-score less than zero on the first immediate recall trial combined with a z-standardised CVLT learning slope greater

than zero; and 25 participants were labelled as non-learners due to a z-score less than zero on both CVLT indices.

McDermid Vaz and Heinrichs (2006) did not initially characterise their cognitive subgroups, however 55 of the original participants were re-examined three years later, with a moderate level of classification agreement between the two time points reported. At follow-up, 12 unimpaired SSD participants were found to outperform 43 memory-impaired SSD participants on a proactive inhibition task of secondary memory. The two subgroups did not significantly differ on three measures of executive function and motor dexterity.

Heinrichs et al. (2017) reported that the unimpaired memory subgroup exhibited control equivalent performance on all eight CVLT subgrouping indices and five adjunct measures of intellect and memory. Similarly, the below-normal ranged SSD subgroup did not significantly differ on all eight CVLT indices used to subgroup participants, in addition to a measure of working memory. Significant impairments in premorbid IQ, verbal comprehension and verbal learning were detected. Compared to the below normal range subgroup, the normal ranged SSD participants exhibited significantly better performance in premorbid IQ, verbal comprehension and verbal learning, with a non-significant difference on a measure of working memory detected. Statistical comparison of the two SSD subgroups on the CVLT was not performed, however effect sizes were reported. Large effects were presented for seven comparisons in favour of the normal subgroup (d_{μ} = 1.04; $d_{\text{range:}}$ 0.91 to 1.29), with a small-to-moderate difference present for CVLT recognition hits (d = 0.44).

Vaskinn et al. (2008) reported that the high-achievers significantly outperformed the learners and non-learners on six and 10 CVLT outcome indices respectively. The learners registered a significantly better learning slope in comparison to the high-achievers and outperformed the non-learners on eight of the CVLT outcome variables, with equal performances detected for the first immediate recall trial and recognition false alarms. The

high-achievers were reportedly using more semantic clustering than both subgroups, with both the high-achievers and learners being more consistent than the non-learners. No significant SSD subgroup differences in serial position effects and serial and subjective clustering were detected. Similarly, no significant SSD subgroup effects were detected for nine adjunct measures of executive function, fluency, motor dexterity and working memory.

3.2.4. Executive function specific

Two studies classified participants into cognitive subgroups based on executive function. On average, 52% of the SSD participants were labelled as having preserved/non-impaired executive ability and outperformed the remaining participants on a select number of adjunct measures. Details about the individual studies are reported below.

By exploring the nature of performance deficits on the WCST, Butler et al. (1992) classified 44 SSD participants into two subgroups of impaired or unimpaired WCST performance based on a cut-off of 30 perseverative responses. Twenty participants met the unimpaired criterion and outperformed the remaining impaired participants on four measures of executive function, fluency, reaction time and verbal comprehension. Comparably, Rüsch et al. (2007) classified 30 participants as unimpaired on the WCST based on successfully completing five or more categories and were reported to register significantly less perseverative errors on the WCST, in addition to a significantly reduced score on the Mini-Mental State Examination compared to 21 impaired performers. The two subgroups did not significantly differ on four indices of working memory.

3.3. Data-driven research

3.3.1. Methodological considerations

Cluster analysis is a data reduction technique that identifies subgroups within a data-set which share distinct similarities (see Clatworthy et al., 2005; Jain, 2010; Morissette and Chartier, 2013; Yim and Ramdeen, 2015). Hierarchical cluster analysis (HCA) combines individual participants who are identified as having the most similar profile of performance across the measures examined into a single cluster or subgroup. This exploratory sequential process continues until all participants have been clustered together into a single cluster comprised of the whole sample. Iterative partitioning techniques, such as *k*-means clustering divide participants into distinct subgroups based on an initial set of cluster centroids or means that represent the variables entered into the analysis. The user predefines the number of clusters to retain and can enter in previously identified group centroids to guide the process. Iterative partitioning can be employed to optimize the cluster solutions identified using an exploratory HCA. Out of the 18 eligible studies identified as employing data-driven techniques, 12 used an exploratory hierarchical cluster analysis (HCA), with four using *k*-means clustering. Gilbert et al. (2014) employed a gaussian mixture model cluster analysis to subgroup participants, with Green et al. (2013) using a Grade of Membership (GoM) analysis (see Woodbury et al., 1978, for further details).

For this section, further details and the study references being cited to can be found in Table 2. As the number of clusters that an exploratory HCA ranges from the number of participants entered into the analysis to one, it is therefore important for authors to detail the basis for how the number of clusters retained was determined. Twelve studies reported conducting a visual inspection of the dendrogram (a hierarchical tree diagram) which at times was combined with either a discriminant function plot or an inverse scree-plot of the agglomeration schedule coefficients to establish the number of clusters to retain. Of these 12 studies, only three presented the plots used to establish the number of clusters to retain in-text. One study reported defining the number of subgroups *a priori* despite the exploratory

nature of HCA, with an additional study not stating how the number of clusters retained was determined. All four studies using *k*-means clustering defined the number of clusters to retain *a priori* based on cited previous research.

As emergent clusters are determined by the combination of methodological specifications (e.g. pre-process standardization, clustering algorithm), the characteristics of the sample examined, and the measures entered into the analysis, validation of the clustering solution is preferred. Twelve studies did not provide any form of validation for their clustering solution. Four studies assessed concordance between the classification from the final clustering solution with one or multiple solutions identified using a different combination of methodological specifications. Two studies combined classification concordance with an assessment of classification strength via a discriminant function analysis, with one requiring concordance with expert neuropsychologists' ratings and face-value consistencies with previous research. The remaining study used consistencies with previous research as proof of subgroup validation.

Thirteen studies provided statistical characterisation of the emergent clusters via *post-hoc* pairwise comparisons across subgroups of the variables entered into the analysis; with six studies comparing subgroups on variables not entered into the subgrouping analysis. Seven studies compared subgroup cognitive performance against that of a HC comparison group, with four studies providing subjective descriptions of the emergent subgroup's cognitive profiles. Van Rheenen et al. (2017) was the only included data-driven study to investigate if cognitive subgroups existed within their HC group.

3.3.2. *K*-means iterative clustering

Three studies were identified that employed *k*-means iterative clustering to examine an *a priori* defined number of subgroups based on current neuropsychological function. Two

studies entered participant performance on the CVLT into the clustering analysis to identify three verbal learning subgroups, whilst the remaining study examined the four cognitive subgroups that emerged based on participant performance on a composite cognitive battery. Details about the individual studies are reported below, with a results summary presented in Table 3.

Both Turetsky et al. (2002) and McDermid Vaz and Heinrichs (2006) used *k*-means clustering to subdivide their respective samples into memory subgroups based on participant total immediate recall, intrusion errors, discriminability-recall difference performance on the CVLT. The three memory subgroups and CVLT indices were predefined based on the early works of Paulsen et al. (1995) who classified schizophrenia participants into distinct learning profiles reflective of various clinical diseases (i.e., Alzheimer's disease). From a sample of 116 SSD particpants, Turetsky et al. (2002) reported that a subgroup of 59 emergered that had significantly better free recall ability compared to the remaining participants and were subsequently labelled as having an unimpaired verbal learning ability. The unimpaired subgroup also exhibited an equivalent cued-intrusion error rate and discriminability-recall difference score to that of 129 HCs. The 36 SSD participants labelled as having a subcortical memory impairment profile exhibited equivalent free-recall to that of the remaining 21 participants characterised by a cortical memory impairment profile, however performed significantly better than the cortical subgroup on both cued-intrusion error rate and discriminability-recall difference score. No further statistical characterisation was conducted, however distinct performance profiles on the primary CVLT indices were presented graphically. A profile of close-to-control level of performance was evident for the unimpaired subgroup, with profiles of similar-to-parallel levels of impairment evident for the subcortical and cortical subgroups.

McDermid Vaz and Heinrichs (2006) reported that 40 SSD participants were identified as having significantly better total immediate recall than the remaining participants and were described to exhibit an intrusion error rate and discriminability-recall difference score that fell within average normative limits; consequently, described as representing an unimpaired memory profile. The two remaining subgroups of 42 and 20 SSD participants were described as representing subcortical and cortical memory profiles respectively and did not significantly differ in severity of free recall impairment. The cortical subgroup was reported to make more cued intrusion errors and have a lower discriminability-recall difference score compared to the subcortical subgroup; with a significantly higher discriminability-recall difference score compared to the unimpaired subgroup unexpectedly reported. A moderate level of classification agreement was reported after reassessing 55 participants three years later. At follow-up, subgroups were externally characterised by comparing their performance on the WCST, Purdue Pegboard Test and a proactive inhibition task of secondary memory. The unimpaired subgroup was reported to exhibit significantly better performance on the index of secondary memory compared to the remaining participants, with significantly better current IQ compared to the cortical subgroup

Based on the findings of Hill et al. (2002), Geisler et al. (2015) entered the performance of 129 SSD participants on a composite cognitive battery into a *k*-means cluster analysis to classify the sample into one of four cognitive subgroups. Unlike Hill et al. (2002) however, participant performance on the cognitive battery was initially entered into a principle component analysis to derive eight cognitive dimensions, with the resulting principle component scores entered into the cluster analysis. Whilst the emergent subgroups were characterised, no statistical comparisons were conducted between subgroups on the cognitive measures used to subgroup participants. The 38 participants of cluster one were described to exhibit diminished verbal fluency paired with impaired processing speed; the 26

participants of cluster two were characterised by diminished verbal episodic memory combined with poor fine motor control and signal detection; the 21 participants of cluster three were reported to have above average verbal fluency and impaired episodic face memory and slowed processing speed; with the remaining 44 participants reportedly characterised by a deficit in general intellectual function. Cluster four was reported to have significantly lower premorbid IQ relative to clusters one and three, with clusters two and four exhibiting significantly lower premorbid IQ compared to a sample of 165 HCs.

3.3.3. Exploratory data-driven subgrouping

Eleven studies were identified that employed exploratory data-driven techniques to identify cognitive subgroups within their respective SSD samples, with a results summary presented in Table 3. Both Cobia et al. (2011) and Green et al. (2013) reported that two subgroups emerged following exploratory data-driven subgrouping using study specific batteries of current cognitive function. Cobia et al. (2011) entered the cognitive performance of 79 SSD participants into a HCA with *k*-means optimization and identified two clusters subsequently labelled as neuropsychologically near-normal (n = 45) and neuropsychologically impaired (n = 34). Whilst Green et al. (2013) entered the cognitive performance of 617 SSD participants into a GoM analysis and identified two subgroups which were subsequently labelled as cognitively spared (n = 323) and cognitive deficit (n = 294). The near-normal subgroup presented by Cobia et al. (2011) exhibited significant impairments on nine of the 10 measures assessing executive function, fluency, memory, processing speed and visuospatial perception relative to 65 HCs; a control-equivalent performance on the matrix reasoning subtest of the WAIS-III was detected. Similarly, Green et al. (2013) reported that the cognitively spared SSD subgroup was significantly impaired on eight of the 10 measures assessing attention, current IQ, fluency, memory verbal

comprehension, working memory and global cognitive functioning compared to 764 HCs; whilst displaying intact premorbid IQ and visuospatial perception. In both studies, the neuropsychologically impaired/cognitive deficit subgroups performed significantly worse than HCs on all measures assessed. The neuropsychologically near normal/cognitively spared subgroups performed significantly better than their respective SSD counterparts in both studies in all but one comparison. Cobia et al. (2011) reported that their SSD subgroups did not significantly differ in spatial working memory ability.

Four studies were identified that reported the emergence of a three-subgroup solution following various exploratory data-driven methodologies. Whilst the labels applied to each subgroup differed across studies, all four reported that a relatively intact subgroup, a moderately impaired subgroup and a severely impaired subgroup emerged. On average, 26% of SSD participants emerged as having relatively intact cognitive function, with 40% characterised by moderate impairments and 32% by severe impairments. SSD participants with relatively intact cognitive function were consistently reported to outperform the remaining SSD participants on the majority of measures assessed, whilst exhibiting near-normal functioning relative to HC comparisons. Overall, the moderately impaired subgroup outperformed the severely impaired subgroup for the majority of contrasts. Details of the studies are provided below.

Entering the performance of 210 SSD on the WCST into a HCA with *k*-means optimization, Carruthers et al. (2019) identified three cognitive subgroups which were based exclusively on executive function. A subgroup of 72 SSD participants emerged as having relatively intact executive function after exhibiting equivalent performance on four out of a possible seven WCST outcome indices relative to 192 HCs. The SSD subgroup performed significantly worse than controls on the WCST in terms of the total number of trials administered, non-perseverative errors, total correct and times failing to maintain set. A

second subgroup of 114 SSD participants emerged which was characterised by a moderate level of impairment on the WCST, only exhibiting a control-equivalent number of total correct. The final subgroup of 24 SSD participants was characterised by severe deficits on all seven WCST outcome indices assessed. The relatively intact participants significantly outperformed the remaining SSD participants on all comparisons except for one. A non-significant difference in the number of times the subgroup failed to maintain set was detected between the relatively intact and severely impaired subgroups; however, this was attributed to the severely impaired subgroups inability to establish set. The severely impaired subgroup performed significantly worse than the moderate impairment subgroup overall, with significant deficits in categories achieved, perseverative errors, trials to first category and total correct.

Gilbert et al. (2014) entered the performance of 112 SSD participants on a composite cognitive battery into a gaussian mixture model cluster analysis and identified three cognitive subgroups. Two clusters were labelled as having generally and selectively impaired cognition, with the third cluster said to represent near-normal cognitive functioning. Subgroups were not statistically compared on the variables entered into the cluster analysis, however the authors summarised the 18 generally impaired SSD participants as exhibiting diffuse cognitive dysfunction; the 46 selectively impaired SSD participants as having deficits in processing speed and visual episodic memory amongst near-average working memory and verbal episodic memory; and the 48 near-normal SSD participants as exhibiting average performances across the five cognitive domains assessed. Subgroups were also externally characterised using four measures of attention, executive function and motor control. The near-normal subgroup was reported to outperform the remaining two SSD subgroups on all adjunct cognitive measures. Whilst the selective impairment subgroup exhibited significantly better attention compared to the generally impaired subgroup.

Van Rheenen et al. (2017) and Sauvé et al. (2018) respectively entered participant performance on the MCCB and Cogstate battery into a HCA cluster analysis to identify two comparative three subgroup solutions. From a total of 564 SSD participants, Van Rheenen et al. (2017) identified 75 which were characterised by relatively intact cognitive function, exhibiting equivalent attention, premorbid IQ, reasoning, social cognition, processing speed, verbal and visual learning and working memory performance relative to 575 HCs. Two-hundred and sixty-two and 227 SSD participants emerged into subgroups defined by mild-to-moderate and relatively severe cognitive impairments respectively, with both groups performing significantly worse on all eight measures assessed relative to HCs. The relatively intact subgroup significantly outperformed the remaining SSD subgroups on all measures assessed, with the mild-to-moderate impairment subgroup also significantly outperforming the relatively severe subgroup on all measures. Sauvé et al. (2018) uniquely entered the seven MCCB domain scores of 121 SSD and 80 FEP participants into a HCA with *k*-means optimization to examine cognitive subgroups across first- and multiple-episode SSD. Whilst no statistical comparison was reported amongst the three subgroups that emerged, 23 participants labelled as having no cognitive impairment exhibited large performance effects on all seven MCCB domains compared to 113 intermediately impaired (d_{μ} = 1.94; $d_{\text{range:}}$ 1.03 to 3.18) and 65 generally impaired participants (d_{μ} = 1.08; $d_{\text{range:}}$ 0.66 to 1.59). The intermediately impaired subgroup also exhibited large performance effects over the generally impaired subgroup on six of the MCCB domains (d_{μ} = 1.15; $d_{\text{range:}}$ 0.60 to 1.73), with small performance effects in attention-vigilance detected (d = 0.29). The no impairment subgroup was comprised of significantly less FEP participants, with the opposite being reported for the generally impaired subgroup. A significant clinical group-by-cluster membership interaction was reported. In decomposing the significant interaction, the authors reported that generally

impaired FEP participants had significantly worse performance on the visual learning and memory tasks compared to their SSD counterparts.

Five studies presented four subgroup solutions, with a single study identifying five. Two studies administered a limited number of attention and/or executive function measures to subgroup participants, whilst the remaining three studies examined composite cognitive battery performance. The five studies presenting four subgroup solutions all reported the emergence of a high functioning or intact subgroup and a severely impaired or deficit subgroup; with two intervening subgroups characterised by intermediate or study-specific mixed impairments. Whilst limited statistical characterisation of emergent subgroups was conducted, the intact/high performing subgroups were consistently described to outperform the remaining subgroups on the majority of measures assessed. Study specific instances of equal performances on various neuropsychological measures were reported between the high performing and intermediate subgroups. A solitary study presented a five-cluster solution uniquely based on homogenous profiles of relative cognitive strengths and weaknesses. Details of the studies are provided below.

Gambini et al. (2003) entered the executive performance of 81 SSD participants into a HCA to reportedly evaluate the dorsolateral prefrontal cortex functions; with four clusters emerging. Thirty-nine participants were described as having good global frontal scores, exhibiting significantly better performance on the WCST, Wiegl Sorting Test (WST) and a test of verbal fluency compared to the remaining participants. Thirty participants were said to have poor scores on the WST and verbal fluency test, with three additional participants characterised by low scores on the WST and WCST. A cluster of nine participants was summarised as having poor WCST, WST and verbal fluency performance. Similarly, Liu et al. (2011) entered the performance of 549 SSD participants on the WCST and a continuous performance test of sustained attention into an exploratory HCA. A non-deficit subgroup of

106 SSD participants emerged which outperformed the 188 deficit participants on four indices of attention and executive function. Two intermediate subgroups emerged, one characterised by a significant executive deficit (n = 109) and the other a significant attention deficit (n = 146) relative to the non-deficit subgroup, outperforming the deficit subgroup in attention and executive function capacities respectively. The two intermediary subgroups were reported to significantly outperform each other on each subgroup's respective non-deficit domain.

Extending beyond the domains of executive function and sustained attention, Hill et al. (2002) entered the performance of 151 SSD participants on a composite cognitive battery into a HCA with minimum sum-of-squares iterative optimization. Subgroups were not statistically compared on the variables entered into the subgrouping analysis, however subgroup performance profiles were presented graphically and elaborated on. The 15 participants of cluster one were described as having severe impairments in language, motor and sensory skills, as well as spatial abilities (HC derived z-score range: -2.5 to -3.5), in addition to profound executive and memory deficits (HC derived z-score > -3.5). The 76 participants of cluster two were described as exhibiting average-to-mildly impaired overall performance, with verbal memory being the most impaired domain. Clusters three and four were described to have overlapping levels of moderate-to-severely impaired cognitive performance. The 41 participants of cluster three exhibited greater impairments in attention and abstraction relative to memory, whereas the remaining 19 participants of cluster four were characterised by more prominent memory impairments relative to their executive ability.

Seaton et al. (1999) used the 11 subtests of the WAIS-R to investigate cognitive subgroups amongst a sample of 102 SSD participants via a HCA. Four subgroups emerged labelled as reflecting high, average with motor deficit, low-average and impaired WAIS-R

performance profiles; subgroup participant numbers were not reported. For the arithmetic, comprehension, digit span, information, similarities and vocabulary subtests, all pairwise comparisons were reported as significant except for the low-average and motor deficit contrasts, which did not significantly differ on any of the 11 subtests. The high subgroup was also said to outperform the remaining SSD participants on the picture arrangement task, with the low average subgroup outperforming the impaired participants on the block design, digit symbol, object assembly and picture completion subtests.

Representing the culmination of several studies from the same lab, Goldstein et al. (1998) compared the clustering solutions that emerged using three overlapping cognitive batteries: an abstraction battery; the WAIS (III/R); and a combination of the abstraction battery the WAIS (III/R). The performance of 221 SSD participants on each of the three batteries was entered into three separate HCA, with the resulting cluster solutions compared. Four overlapping clusters emerged; near normal/above average functioning, moderately impaired, moderately impaired with a specific deficit, and a severely impaired subgroup. For the two analyses that included the abstraction battery, a moderately impaired plus psychomotor deficit subgroup emerged and with the WAIS (III/R), an average verbal IQ plus moderately poor performance IQ subgroup emerged. Whilst no statistical comparisons were performed across subgroups, the classification concordance amongst the three final solutions was, with poor levels of overall agreement reported for each of the three comparisons. However, it was stated that in all three comparisons, the near-normal and severely impaired subgroups exhibited high levels of concordance across clustering solutions, with the various intermediate deficit subgroups exhibiting a high level of classification variability.

Dawes et al. (2011) uniquely investigated cognitive subgroups based on the relative strength and weaknesses of 144 SSD participants on a composite cognitive battery. The authors created domain deviation scores based on each participant's performance on seven

cognitive domains relative to their overall cognitive performance and entered these scores into a HCA with *k*-means optimization. By classifying participants in this way, five subgroups emerged. A subgroup of 19 SSD participants exhibited a profile of visual learning and memory weakness, with a second subgroup of 40 participants characterised by a relative weakness in executive function. Thirty-eight participants were characterised by relative strengths on the verbal comprehension index of the WAIS-III, as well as processing speed, paired with impaired executive function, auditory and visual learning and memory. Profile four was comprised of 17 participants with a relative weakness in auditory learning and memory and a relative strength in visual learning and memory. The remaining 30 participants exhibiting a relative weakness in visual learning and memory. Participants of the latter two profiles also exhibited and overlapping relative weakness in executive function.

3.4. Putative cognitive symptom trajectories

In total, 13 studies investigated putative cognitive symptom trajectories by examining the relationship between a participant's estimated premorbid and current IQ. Ten studies classified participants into subgroups based on *a priori* defined criteria, with the remaining three studies employing various data-driven subgrouping techniques. Although the classification criteria employed varied across studies, they all centred around the relationship between a participant's estimated premorbid and current functioning. Participants were typically classified into one of three subgroups representing putatively preserved, deteriorated or compromised symptom trajectories. Overall, the preserved trajectory was characterised by an absence of a meaningful decline from near-normal premorbid estimates and subtle impairments in current cognitive function; deteriorated trajectory was characterised by a meaningful decline from near-normal premorbid estimates to severe and widespread impairments in current cognitive function; and finally compromised trajectory represents a

putative symptom course defined by pronounced, widespread and purportedly stable impairments that are said to manifest prior to the onset of psychosis. Across all studies employing this tripartite model, 33% of SSD participants were, on average classified as having preserved function, 41% as having deteriorated function and 21% as being compromised. The preserved SSD participants exhibited an overall profile of relatively intact or near-normal cognitive function, with deficits in processing speed and memory consistently reported. The preserved subgroups were reliably reported to outperform the deteriorated and compromised subgroups on the majority of measures assessed, with moderate-to-large effects present. Significant differences in cognitive performance between the deteriorated and compromised subgroups were uncommon when *a priori* classification criteria were employed. Conversely, when data-driven subgrouping techniques were utilised, deteriorated participants exhibited better cognitive performance than the compromised participants, albeit with a similar degree of processing speed and memory impairment. Details about the individual studies are reported below, with a results summary presented in Table 4.

Two early studies only examined preserved SSD participants. Evans et al. (1997) defined preserved function as having an estimated premorbid-current IQ difference of less than ten points, with Elliott et al. (1998) including an upper current IQ requirement of 90 points or higher to be classified as having preserved function. Evans et al. (1997) reported that 16 from 31 SSD participants met the criterion and exhibited equivalent cognitive flexibility and planning ability to that of 16 matched HCs. Significant impairments in practical problem solving, task monitoring temporal judgement were found, in addition to the overall BADS profile score. Twelve SSD participants met the updated criteria for having preserved function set by Elliott et al. (1998) and were reported to have intact executive function, pattern recognition and spatial working memory abilities relative to 12 HCs. The

SSD subgroup was reported as having significantly impaired cognitive flexibility, delayed pattern recognition, spatial recognition and span however.

Weickert et al. (2000) proposed a tripartite model of putative cognitive symptom trajectory that has since been regularly investigated amongst the SSD population; albeit with multiple alterations to their original classification criteria. They defined preserved functioning as having an premorbid IQ greater than 90 points and the absence of a 10-point premorbid-current IQ difference; Deteriorated functioning as having a premorbid-current IQ difference of 10-points or more; and compromised functioning as having an premorbid IQ less than 90 points in the absence of a meaningful premorbid-current IQ difference. Twenty-nine SSD participants were classified as having preserved functioning and exhibited intact arithmetic, current IQ, executive function, fluency, reasoning, premorbid IQ, processing speed, memory, verbal comprehension, verbal learning and verbal comprehension relative to 27 HCs. Significant impairments were also reported in attention and overall executive efficiency. Sixty deteriorated and 39 compromised SSD participants performed significantly worse than the HC group on 16 and 18 out of the 23 measures assessed respectively. Both subgroups exhibited control equivalent performance on a finger tapping test, fluency and measure of logical memory, with the deteriorated subgroup also exhibiting intact premorbid IQ and performance on the Boston Naming Test. The preserved subgroup outperformed the deteriorated subgroup on the majority of measures assessed; with equivalent performances in sustained attention and verbal retrieval detected. Likewise, the preserved subgroup exhibited significantly better performance compared to the compromised subgroup on all measures presented as having significant subgroup differences. Compromised SSD participants exhibited more pronounced deficits in attention, executive efficiency, verbal retrieval and visuospatial perception relative to their deteriorated counterparts. The three SSD subgroups did not significantly differ in motor dexterity or verbal fluency.

Using the same classificaiton criteria, Kremen et al. (2008) reported that 22 from 80 SSD participants with preserved function exhibited equivalent performance on nine indices of arithmetic, attention, fluency, premorbid IQ, verbal learning and visuospatial ability to that of 93 HCs, whilst also being significantly impaired on 12 alternate indices of attention, current IQ, executive function, processing speed, verbal learning and visuospatial ability. Forty deteriorated participants and 19 compromised participants performed significantly worse than the HCs on 18 and all 21 variables assessed respectively; with the deteriorated subgroup exhibiting intact premorbid IQ and graphic and visuospatial sequencing ability. The preserved and deteriorated participants exhibited equivalent performances on nine of the measures assessed, with the deteriorated subgroup displaying significantly impaired attention, current IQ, executive function, fluency, processing speed, verbal learning and visuospatial ability. In contrast, the compromised participants only performed as well as the preserved subgroup on five measures of attention, executive function and visuospatial ability. Compared to the deteriorated participants, the compromised subgroup exhibited significantly impaired arithmetic, current IQ, executive function, fluency, premorbid IQ, verbal learning and visuospatial ability.

Ammari et al. (2014), Badcock et al. (2005) and Leeson et al. (2011) altered the original criteria set for both preserved and compromised functioning. In the absence of a meaningful premorbid-current IQ difference, preserved functioning was defined by a premorbid and current IQ greater than or equal to 90-points, with compromised function defined by a premorbid and current IQ less than 90-points; with Ammari et al. (2014) deviating from previous descriptive labels and denoting the compromised subgroup as being premorbidly impaired. Ammari et al. (2014) classified both SSD and HC participants using the above criteria and compared across diagnostic subgroups. Thirty-two SSD and 20 HC particpants met the criteria for preserved function and did not significantly differ in attention-

vigilance or current IQ. The preserved SSD participants did perform significantly worse than the matched-HC subgroup in premorbid IQ and on seven of the eight MCCB cognitive domains; with the largest effect being calculated for the overall composite score. The preserved SSD subgroup was reported not to significantly differ on nine of the 10 variables assessed from a subgroup of nine HCs meeting the criteria for deteriorated function. The preserved SSD subgroup was found to have significantly higher current IQ than the deteriorated HCs. Twenty-six SSD participants were classified as having deteriorated function and were found to have significantly impaired current IQ and MCCB domain performance relative to the preserved HC subgroup. Conversely, the SSD and HC subgroups matched as having deteriorated function did not significantly differ in premorbid or current IQ and on six MCCB domains, with a significant difference in attention and MCCB overall composite score detected in favour of the HC subgroup. The remaining 13 SSD participants were classified as premorbidly impaired and were found to be significantly impaired on all 10 variables assessed relative to the preserved HC subgroup and on four variables compared to the deteriorated HC subgroup. The premorbidly impaired SSD participants exhibited the same level of performance in their current IQ, reasoning, processing speed, social cognition and verbal and visual learning ability as per the deteriorated HC subgroup. In comparing cognitive performance amongst the SSD subgroups, preserved participants outperformed the deteriorated subgroup in attention, processing speed, visual learning and on the overall composite score of the MCCB, whilst outperforming the premorbidly impaired SSD subgroup on six of the MCCB domains, premorbid and current IQ; with non-significant differences in reasoning and social cognition reported. The deteriorated and premorbidly impaired subgroups were reported to significantly differ in their premorbid IQ only.

Badcock et al. (2005) reported that 45 from 109 SSD participants were classified as having preserved function and exhibited equivalent performance on a measure of phonemic

fluency relative to 149 HCs. After covarying for a significant difference in age and education, it was reported that the preserved SSD participants exhibited a significant impairment in attention, processing speed and verbal learning. The preserved subgroup was found to outperform 47 deteriorated SSD participants on four measures of attention and verbal learning, in addition to no fluency or processing speed performance differences. Compared to 17 compromised participants, the preserved subgroup exhibited significantly better performance on three measures of attention, fluency and verbal learning; with the two subgroups not differing in their performance on an alternate measure of attention, processing speed and verbal learning. The deteriorated and compromised subgroups did not significantly differ on any of the six measures assessed.

Applying the same criteria to a sample of 129 FEP participants, Leeson et al. (2011) classified 40 as exhibiting preserved function and reported no significant differences on measures of premorbid and current IQ, executive function, verbal learning and working memory relative to 120 HCs. The preserved subgroup was reported as having significantly impaired working memory manipulation ability however. Fifty-seven deteriorated SSD participants and 32 compromised participants, labelled in this instance as having low IQ, performed significantly worse than controls on all seven measures assessed. The preserved subgroup performed significantly better than both the deteriorated and compromised/low IQ subgroups on all seven measures, with the deteriorated and compromised/low IQ subgroups only differing in their premorbid IQ.

Kravariti et al. (2009) provided an alteration to the criteria outlined above by Ammari et al. (2014), Badcock et al. (2005) and Leeson et al. (2011), by relabelling the preserved and compromised subgroups to stable good and stable poor respectively and substituting the deteriorated subgroup with a deteriorated-poor IQ subgroup (dpIQ); dpIQ was defined by a premorbid IQ of less than 90-points and a premorbid-current IQ difference of 10 points or

more. The authors classified 101 FEP and 317 HC participants into three subgroups and compared cognitive performance between each subgroup pair. Twenty-two FEP and 137 HC participants were classified as stable good and exhibited equal performances on 14 measures of current IQ, executive function, fluency, premorbid IQ, processing speed, verbal comprehension, verbal and visual learning and visuospatial perception. Significant impairments in executive function, processing speed and working memory were detected for the FEP subgroup. Thirty-seven FEP dpIQ participants were reported as performing significantly worse on three measures of processing speed and working memory relative to 26 dpIQ HCs. Nineteen FEP participants classified as stable poor performed significantly worse on four measures of fluency and processing speed relative to 14 stable poor HCs.

Joyce et al. (2005) defined preserved function as a premorbid IQ greater than or equal to 90 points, in the absence of a 10-point premorbid-current IQ difference; deteriorated function as having an premorbid IQ of less than 90 points with an premorbid-current IQ difference of 10 points or more; and low IQ as having both a premorbid and current IQ of less than 90 points. Using this criteria, 47 FEP participants were classified as having preserved function and were reported to have no significant impairments on eight measures of current IQ, executive function, premorbid IQ, visual learning and working memory relative to 60 HCs. Thirty-seven FEP participants were classified as having deteriorated function and were reported to exhibit significant deficits on all measures except for intact premorbid IQ relative to HCs. Nine participants were labelled as having low IQ and were reported to exhibit significant deficits in premorbid and current IQ, as well as executive function compared to the HC group. Control equivalent performance in planning, visual learning and working memory were detected for the low IQ subgroup, however a lack of statistical power was a likely contributing factor. In comparing cognitive performance amongst the FEP subgroups, it was reported that the preserved participants outperformed the

deteriorated subgroup in current and premorbid IQ, executive function and working memory and the low IQ subgroup in both premorbid and current IQ. The deteriorated subgroup exhibited significantly better premorbid IQ compared to the low IQ subgroup, with no other significant difference detected.

MacCabe et al. (2012) provided yet another iteration of the classification criteria by investigating SSD participant with superior intellect or a premorbid IQ greater than 115. In comparing a subgroup of 10 high IQ-intact (HI) SSD participants to 19 HCs with above-average current IQ (unspecified), MacCabe et al. (2012) reported no significant performance differences across 20 measures assessing multiple core cognitive domains. Twenty-four high IQ-decline (HD) SSD participants performed as well as the HCs on 11 measures of current IQ executive function, fluency, premorbid IQ and working memory, whilst exhibiting significantly impaired performance on alternate measures of executive function, fluency and working memory, as well as processing speed and verbal learning. The HI subgroup performed significantly better than the HD subgroup on only five measures of current IQ, executive function, verbal learning and working memory. In contrast, the HI subgroup exhibited significantly better performance on 12 of the measures assessed than a subgroup of 16 SSD participants classified as having typical IQ (premorbid IQ < 110). Whereas the HD subgroup only outperformed the typical IQ subgroup on three measures of fluency, premorbid and current IQ.

Three of the included studies examined putative cognitive symptom trajectories using varying data-driven subgrouping techniques (see section 3.8 below for an overview on data-driven methodologies). Based on the works of Weickert et al. (2000) and Kremen et al. (2008), Potter and Nestor (2010) entered the premorbid and current IQ of 73 SSD participants into a *k*-means iterative partitioning analysis to subgroup participants into one of three *a priori* defined subgroups. Twenty-one SSD participants emerged as being characterised by

preserved functioning, 21 as deteriorated and 31 as compromised. Compared to 74 HCs, the preserved subgroup exhibited equivalent performance on four of the five WAIS-III derived intellectual abilities, with a significant deficit on the processing speed index reported. The preserved participants were also found to exhibit control-equivalent premorbid and current IQ, WCST performance, as well as intact working, auditory immediate and delayed memory as measured on by the WMS-III. Significant deficits in memory, processing speed, verbal and visual learning were detected. Both of the deteriorated and compromised subgroups were found to exhibit significant impairments on the majority of the 20 measures assessed. The deteriorated subgroup performed as well as HCs in premorbid IQ and executive function, with the compromised subgroup not significantly differing from controls in the number of categories achieved on the WCST. Preserved participants outperformed both remaining subgroups on all five WAIS-III derived intellectual abilities. The deteriorated participants exhibited significantly impaired current IQ, memory, processing speed, reasoning, verbal comprehension and verbal and visual learning compared to the preserved subgroup, with the compromised participants registering intact executive functioning and processing speed relative to their preserved counterparts. The preserved and deteriorated subgroups did not significantly differ in their premorbid or current IQ, executive function, verbal and visual leaning, verbal comprehension and working memory capacity.

Both Wells et al. (2015) and Weinberg et al. (2016) utilised exploratory hierarchical cluster analyses with *k*-means optimization to identify putative cognitive symptom trajectories amongst their respective SSD samples. Rather than using a single index of current IQ as per previous research, Wells et al. (2015) used the attention and immediate memory index scores of the RBANS and the letter number sequencing subtest of the WASI-III as a composite index of current cognitive ability, in addition to premorbid IQ to explore putative cognitive symptom trajectories in 543 SSD participants. One-hundred and fifty-seven SSD

participants emerged to be characterised by preserved function after exhibiting intact attention, premorbid IQ, verbal learning, visuospatial perception and working memory relative to 635 HCs. Significant fluency, language and verbal learning deficits were detected amongst the preserved subgroup however. Two-hundred and thirty-nine SSD participants emerged as having deteriorated function and were reported to be significantly impaired on all eight measures assessed relative to HCs. Similarly, 138 SSD participants characterised by compromised function were also significantly impaired on all eight measures assessed relative to the HCs. Preserved SSD participants outperforming both remaining SSD subgroups on all eight measures, with the deteriorated subgroup exhibiting significantly better performance than the compromised subgroup on all measures.

More in-line with the traditional subgrouping method, Weinberg et al. (2016) entered only the premorbid and current IQ of 96 SSD participants into an exploratory hierarchical cluster analysis to identify four putative cognitive symptom trajectories. Twenty-five SSD participants were described as following a putatively preserved symptom trajectory and were reported to exhibit intact premorbid and current IQ, fluency, memory and working memory relative to 87 HCs, with only a significant processing speed deficit detected. Unlike previous research however, Weinberg et al. (2016) identified two deteriorated subgroups, one characterised by a moderate and the other a severe degree of decline in cognitive function. The moderately deteriorated subgroup was comprised of 33 SSD participants and performed significantly worse than HCs on five measures of current IQ, processing speed and working memory, whilst exhibiting intact premorbid IQ and fluency ability. Twenty-seven participants were identified as having severely deteriorated cognitive profiles and were found to exhibit more pronounced deficits on the above five measures, in addition to premorbid IQ and fluency. The putatively preserved subgroup was reported to significantly outperform both the moderate and severely deteriorated subgroups in premorbid and current IQ and memory,

with the severely deteriorated subgroup exhibiting an additional fluency deficit. The moderate and severely deteriorated subgroups did not differ in their premorbid IQ, memory or processing speed capacities, however were matched for current IQ and fluency. A putatively compromised subgroup of 11 SSD participants emerged, however no statistical comparisons were conducted due to the small subgroup size.

4. Discussion

The aim of the current review was to provide a comprehensive and systematic account of the structure of cognitive heterogeneity in SSD. Examination of cognitive function in SSD typically involves documenting the extent of cognitive deficit from HCs, however it is apparent that considerable cognitive variability exists within the population. Despite the methodological variability amongst the included studies, overlapping subgroups of SSD participants characterised by homogenous cognitive profiles were consistently found. Although the number of cognitive subgroups varied from two to five, a consistent pattern generally emerged. That is, two opposing cognitive subgroups were clearly evident; one characterised by relatively intact cognitive function and the other severe and widespread cognitive deficits, with a third, intermediate subgroup also likely to be present. Below we discuss each of these in turn.

4.1. Relatively intact cognitive function

This synthesis indicates that a proportion of individuals with SSD exhibit relatively intact cognitive abilities. Of the SSD participants empirically classified into three or more subgroups that were characterised via appropriate *post-hoc* comparisons, we identified 25% of the overall sample (eight studies, N = 2220) as exhibiting relatively intact levels of current cognitive function. This subgroup was frequently reported to exhibit control-equivalent

performance across a range of cognitive measures, was spared from any identifiable decline in cognitive function from premorbid estimates and consistently outperformed the remaining SSD participants with moderate-to-large effects. This led some authors to label this subgroup as reflecting a cognitively normal subgroup (e.g. Ammari et al., 2010; González-Blanch et al., 2010; Holthausen et al., 2002; McDermid Vaz and Heinrichs, 2006). However, whilst there was some variability amongst the included studies, a degree of impairment was apparent relative to HCs overall. Impairments to executive function, processing speed and verbal learning were some of the most consistently reported amongst the high performing subgroups (Cobia et al., 2011; Kravariti et al., 2009; Ruiz et al., 2007; Weickert et al., 2000). It is therefore evident that a proportion of individuals with SSD have better cognitive functions relative to others with the disorder, however can still experience mild impairments when compared to the HC population. This review therefore supports the validity of naming conventions already applied by some authors (e.g. Carruthers et al., 2018; Van Rheenen et al., 2017), in which labelling of this subgroup reflects its composition of individuals with SSD as manifesting a relatively intact cognitive subgroup. Here, 'relatively' is a crucial descriptor, in contrast to the work that has suggested complete cognitive normality in such individuals.

4.2. Globally impaired cognitive function

A second subgroup was consistently characterised by pronounced cognitive impairments not restricted to any specific domain. Of the SSD participants empirically classified into three or more subgroups that were characterised via appropriate *post-hoc* comparisons, we identified 44% of the overall sample (eight studies, N = 2220) as exhibiting severe cognitive impairments. Globally impaired participants would perform up to three standard deviations below the standardized HC mean and up to one-and-a-half standard

deviations below the average SSD performance across the core cognitive domains. Several studies did report indifferent performance between the severely impaired participants and the relatively intact participants on a select number of measures (e.g. Ammari et al., 2010; Guimond et al., 2016; Ruiz et al., 2007), however the overall evidence depicts a profile of global and pronounced impairment.

The literature that succeeded the work of Weickert et al. (2000) has increasingly noted that multiple cognitive trajectories might exist for such individuals. Some individuals with SSD who experience severe cognitive impairments appear to have once had near-normal functioning prior to the onset of the psychotic illness; experiencing a decline in function to their current severely impaired levels. Conversely, others appear to have had severe cognitive dysfunction before developing SSD. Whilst the classification criteria most commonly applied characterises such individuals as following a stable trajectory of severely impaired cognitive function, it is likely that a proportion of such individuals also experience an undetectable decline from premorbid levels. At present though, further characterisation of the course of cognitive symptom development is needed; as the evidence reviewed here is cross-sectional and this tripartite model of cognitive symptom trajectory has yet to be comprehensively supported by longitudinal evidence.

4.3. Intermediate cognitive function

The above dichotomy begins to meaningfully characterise the cognitive heterogeneity that is present in SSD, providing moderate discrimination between two distinct subgroups with opposing cognitive profiles. Eighteen criteria-based and two data-driven studies classified SSD participants into relatively intact and impaired performers, with significant differences reported for the majority of subgroup comparisons. However, our synthesis

suggests that a third cognitive subgroup exists within the SSD population, characterised by an intermediate degree of cognitive impairment.

Ten of the 12 studies employing non-trajectory based exploratory data-driven techniques presented final cluster solutions comprised of three or more cognitive subgroups, with Weinberg et al. (2016) also empirically identifying a subgroup characterised by only a moderate level of deterioration from premorbid estimates to an intermediate level of impairment in current cognitive function. Two additional data-driven studies also defined three-cluster solutions *a priori* (McDermid Vaz and Heinrichs, 2006; Turetsky et al., 2002). Whilst several studies presented multiple cognitive subgroups between the two anchoring subgroups (e.g. Goldstein et al., 1998; Hill et al., 2002; Liu et al., 2011; Seaton et al., 1999), it is contended that these represent a differentiation of the same intermediate impairment subgroup. In such instances, two subgroups are typically characterised by a study-specific cognitive strength or weakness amongst an overall profile of moderate cognitive impairments. Considerable overlap between the two intermediate profiles is commonly reported and when both profiles are viewed together, a single intermediate cognitive subgroup can be conceptualized. Furthermore, the largest exploratory data-driven study to-date supports the existence of three cognitive subgroups adhering to profiles of relatively intact, intermediate and severely impaired functioning. In drawing the multidimensional cognitive performance of 564 SSD participants from one Australian and three American independent research sites, Van Rheenen et al. (2017) reported that three subgroups emerged; characterised by profiles of relatively intact, moderately and severely impaired cognitive function. The relatively intact subgroup was reported to outperform the moderately impaired subgroup, who went on to outperform the severely impaired subgroup on all cognitive domains with moderate-to-large effects present; supporting the proposed tripartite model of cognitive subgroups in SSD.

Evidence pertaining to the course of cognitive impairment for the intermediate subgroup is limited. Weickert et al. (2000) sought to validate their classification methodology using data-driven techniques and reported two subgroups as exhibiting varying degrees of meaningful decline. Whilst the two clusters were taken as validation of their deteriorated subgroup, one cluster was reported to exhibit only a small decline from estimated premorbid levels to an intermediately impaired level of current functioning. Similarly, Weinberg et al. (2016) presented a moderately and severely deteriorated subgroup of SSD participants. The moderately deteriorated subgroup was characterised by a meaningful decline from estimated premorbid levels to an intermediately impaired level of current functioning. Whilst restricted in scope, it would appear that the course of cognitive impairment for the intermediate subgroup may be of mild deterioration. However, as the trajectory-based studies included in the review focus on characterising SSD participants with relatively intact or severely impaired current cognitive capabilities and restricted to proxy measures of premorbid function, further characterisation of the intermediate cognitive subgroup is required.

4.4. Methodological considerations

Overall, the results of our systematic review of the literature indicate that whilst considerable cognitive heterogeneity exits within the SSD population, three distinct subgroups of relatively intact, intermediate and globally impaired cognitive functions appear to exist. Research using classification criteria has made substantial contributions to our understanding of cognitive subgroups in SSD, however it appears to be limited to classifying participants into either relatively intact or impaired cognitive subgroups. Furthermore, it was common throughout this body of research for inconsistent and somewhat arbitrary classification criteria to be utilised, without considering the intrinsic performance distribution characteristics of the measures being employed. For example, Ayesa-Arriola et al. (2018) and

Vaskinn et al. (2014) both employed an upper boundary of 90 IQ points or less as reflective of low IQ, in contrast to 85 points or less, which has previously been accepted as the upper boundary for borderline intellectual functioning or rather representing 1 s.d. below normal intellectual functioning (Wieland and Zitman, 2016). An exclusion to this is the work which succeeded the seminal paper of Weickert et al. (2000) in characteristing the putative cognitive symptoms trajectories of SSD. However, as mentioned above, this body of litearture is limited by its cross-sectional nature and use of proxy measures of premorbid funcitoning. Validation of these putative symptom trajectories using comprehensive longitudinal research designs is therefore needed to better characterise the multiple courses of cognitive impairment that appear to be prevelant in SSD.

Empirical subgrouping techniques enable a relatively unbiased data-driven classification method to be employed and have provided valuable insight into the multiple cognitive subgroups that exists within the SSD population. However, considerable methodological variability exists in the empirical characterisation of the cognitive heterogeneity of SSD. Included studies varied in the use and reporting of pre-process standardization, subgrouping methods, basis for determining final subgroup solutions, data transparency, validation and *post-hoc* statistical characterisation. Access to large and comprehensive data-sets is becoming more readily available and the use of data-driven techniques, particularly cluster analysis more frequent. It is therefore important that methodological consistency is upheld, and adequate information is reported to ensure transparency and the validity of final subgroup solutions. As the use of cluster analysis is becoming increasingly popular in the study of cognitive heterogeneity in SSD, we have created a brief set of guidelines based on our review, to overcome the inconsistencies identified in the reviewed literature. These can be found in Table 5. It is hoped that these guidelines will promote consistent data-driven methodologies and reporting and will

therefore enable more valid and comprehensive characterisation of the cognitive heterogeneity prevalent in SSD.

Finally, the majority of studies reviewed here did not account for the meaningful cognitive heterogeneity that is likely to exist within the HC population. Whilst the focus of this body of work has been directed towards the structure of cognitive heterogeneity in SSD, the potential variability within the HC population should be considered. As SSD subgroups under investigation are typically compared against a single HC sample, high performing SSD participants are being compared against the HC average; which is likely to contain both poorly and very strongly performing participants. When HC samples are classified using the same criteria as per the SSD participants, mixed results have been found between cognitively-matched SSD and HC subgroups. Some studies report parallel performances (e.g. Ayesa-Arriola et al., 2018; Heinrichs et al., 2008; Heinrichs et al., 2015), whilst others report significant deficits amongst the SSD subgroups (e.g. Kremen et al., 2001; Vaskin et al., 2014). To-date this methodological avenue has yet to be comprehensively explored using data-driven subgrouping techniques (although see Van Rheenen et al., 2017). Future research should therefore consider the cognitive heterogeneity within the HC population to enable better characterisation of the SSD subgroups which emerge.

4.5. Future directions and conclusions.

Considerable cognitive heterogeneity is known to exist within the SSD population. This review provided a systematic and comprehensive account of the empirical characterisation of cognitive subgroups in SSD. Despite the substantial methodological variability amongst the included studies, it is apparent that at least three distinct cognitive subgroups can be reliably identified: a relatively intact subgroup defined by a level of cognitive performance that goes against the existing consensus regarding cognition in SSD

by falling within ~0.5 s.d. of the HC mean performance; an intermediate cognitive subgroup defined by mixed or moderate levels of cognitive function/dysfunction that lies approximately between 0.5 to 1.5 s.d. below HCs; and a globally impaired subgroup characterised by severe cognitive deficits (> 1.2 s.d. below HCs) not specific to any single domain. Moving forward, the challenge will be to establish whether these subgroups are simply representative of subdivisions on a linear continuum, as the current synthesis tentatively suggests, or whether they reflect variations of the disorder that map onto differences in underlying aetiology, pathophysiology or disease trajectory. The functional and treatment relevance of these subgroups is also not clear, but we can begin to address these questions in three key ways:

Firstly, we must confirm the cognitive validity of subgroups by identifying whether they differ on independent measures of the same cognitive domains used to form them. As previously mentioned, subgroups that result from data-driven classification techniques are heavily influenced by the measures used to subgroup participants. Post-subgrouping characterisation using independent cognitive measures is therefore important for determining whether the resulting typology is related to the specific methodology employed. This review indicates the existence of similar subgroups across studies despite the use of different batteries to create them. Whilst this suggests that the subgroups are likely to be cognitively consistent across measures, it also raises the question about whether they are mapping onto a single latent cognitive (*g*) dimension that differs in severity. Further, research examining the generalisation of cognitive deficits in subgroups classified by a single domain (e.g. executive function or memory) will contribute to our understanding of this issue, by clarifying the breadth and selectivity of cognitive variation. It should be noted that the inclusion of social cognitive measures here will be important, given that preliminary, psychosis spectrum evidence shows that social cognitive deficits are evident in only those individuals considered

to be cognitively impaired on more traditional measures relative to those that are not (Bora et al., 2016; Van Rheenen and Rossell, 2016).

Secondly, we must ascertain whether there are meaningful pathophysiological differences between the subgroups. To date, a small number of studies have investigated cognitive subgroup-brain relationships in psychosis (e.g. Guimond et al., 2016; Shepherd et al., 2015; Van Rheenen et al., 2018; Weinberg et al., 2016; Wexler et al., 2009; Woodward and Heckers, 2015). Indicating that there may be some variation between subgroups, although this may not be extensive. More comprehensive research is needed to confirm or deny the biological validity of the cognitive subgrouping approach to SSD.

Thirdly, we must establish the importance of these subgroups in terms of functioning and treatment response. Preliminary evidence indicates greater overall functional impairment in cognitively compromised subgroups (e.g. Holthausen et al., 2002; Wells et al., 2015), but further work is required to determine whether specific aspects of functional impairment map to the different subgroups. Further, one study found that cognitive subgroups differ in their profiles of disease outcome and lifetime response to treatment (Gilbert et al., 2014). Taken in the context of evidence that neurobiology and behaviour do not necessarily have a one-to-one relationship (e.g. Van Rheenen et al., 2019), it therefore seems possible that the cognitive subgrouping approach may have utility in terms of psychological treatment assignment and outcome even in the absence of differences between subgroups that would render them neurobiologically meaningful. Indeed, an understanding of what the cognitive profile needs to look like, or where the cognitive deficit level needs to be for cognitive remediation to be effective might be informed by more homogenous subgrouping of patients based on cognitive patterns irrespective of underlying biology. For this reason, our second and third perspectives on future research need not be considered sequentially.

In sum, this review suggests that cognitive subgroups in SSD can be reliably identified despite considerable methodological variability. However, future research needs to establish the validity of the cognitive subgroups and determine the extent to which they can be independently targeted to provide more beneficial treatments.

5. Acknowledgements

This work was supported by Australian Postgraduate Awards (S.P.C, P.J.S) and by the Australian National Health and Medical Research Council (NHMRC; fellowships to S.L.R (ID 1154651), C.G (ID: 5467262), T.V.R (1088785) and a project grant to S.L.R (ID: 1060664)).

6. References

Ammari, N., Heinrichs, R.W., Miles, A.A., 2010. An investigation of 3 neurocognitive subtypes in schizophrenia. Schizophr. Res. 121, 32-38.

Ammari, N., Heinrichs, W.R., Pinnock, F., Miles, A.A., Muharib, E., Vaz, S.M., 2014. Preserved, deteriorated, and premorbidly impaired patterns of intellectual ability in schizophrenia. Neuropsychol. 28, 353-358.

Ayesa-Arriola, R., Setien-Suero, E., Neergaard, K.D., Belzunces, A.A., Contreras, F., van Haren, N.E.M., Crespo-Facorro, B., 2018. Premorbid IQ subgroups in first episode non affective psychosis patients: Long-term sex differences in function and neurocognition. Schizophr. Res. 197, 370-377.

Badcock, J.C., Dragovic, M., Waters, F.A.V., Jablensky, A., 2005. Dimensions of intelligence in schizophrenia: evidence from patients with preserved, deteriorated and compromised intellect. J. Psychiatr. Res. 39, 11-19.

Bell, M.D., Zito, W., Greig, T., Wexler, B.E., 2008. Neurocognitive enhancement therapy with vocational services: Work outcomes at two-year follow-up. Schizophr. Res. 105, 18-29.

Bora, E., Veznedaroğlu, B., Vahip, S., 2016. Theory of mind and executive functions in schizophrenia and bipolar disorder: A cross-diagnostic latent class analysis for identification of neuropsychological subtypes. Schizophr. Res. 176, 500-505.

Bowie, C. R., McGurk, S. R., Mausbach, B., Patterson, T. L., & Harvey, P. D. (2012). Combined cognitive remediation and functional skills training for schizophrenia: Effects on cognition, functional competence, and real-world behavior. *Am. J. Psychiatr. 169*, 710-718. doi:10.1176/appi.ajp.2012.11091337

Bowie, C.R., Reichenberg, A., Rieckmann, N., Parrella, M., White, L., Harvey, P.D., 2004. Stability and functional correlates of memory-based classification in older schizophrenia patients. Am. J. Geriatr. Psychiatr. 12, 376-386.

Butler, R.W., Jenkins, M.A., Sprock, J., Braff, D.L., 1992. Wisconsin Card Sorting Test deficits in chronic paranoid schizophrenia – Evidence for a relatively discrete subgroup. Schizophr. Res. 7, 169-176.

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., 2019. Exploring heterogeneity on the Wisconsin Card Sorting Test in schizophrenia spectrum disorders: A cluster analytical investigation. J. Int. Neuropsychol. Soc. 1-11.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., Horne, R., 2005. The use and reporting of cluster analysis in health psychology: A review. Br. J. Health Psychol. 10, 329-358.

Cobia, D.J., Csernansky, J.G., Wang, L., 2011. Cortical thickness in neuropsychologically near-normal schizophrenia. Schizophr. Res. 133, 68-76.

Dawes, S.E., Jeste, D.V., Palmer, B.W., 2011. Cognitive profiles in persons with chronic schizophrenia. J. Clin. Exp. Neuropsychol. 33, 929-936.

Dickerson, F.B., Stallings, C., Origoni, A., Boronow, J.J., Sullens, A., Yolken, R., 2007. The association between cognitive functioning and occupational status in persons with a recent onset of psychosis. J. Nerv. Ment. Dis. 195, 566-571.

Dickerson, F.B., Stallings, C., Origoni, A., Boronow, J.J., Sullens, A., Yolken, R., 2008. Predictors of occupational status six months after hospitalization in persons with a recent onset of psychosis. Psychiatr. Res. 160, 278-284.

Dickson, H., Laurens, K.R., Cullen, A.E., Hodgins, S., 2012. Meta-analyses of cognitive and motor function in youth aged 16 years and younger who subsequently develop schizophrenia. Psychol. Med. 42, 743-755.

Elliott, R., McKenna, P.J., Robbins, T.W., Sahakian, B.I., 1998. Specific neuropsychological deficits in schizophrenic patients with preserved intellectual function. Cognit. Neuropsychiatry 3, 45-70.

Evans, J.J., Chua, S.E., McKenna, P.J., Wilson, B.A., 1997. Assessment of the dysexecutive syndrome in schizophrenia. Psychol. Med. 27, 635-646.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. John Wiley & Sons. Ltd., New York. doi:10.1002/9780470977811

Fett, A.K.J., Viechtbauer, W., Dominguez, M.D.G., Penn, D.L., van Os, J., Krabbendam, L., 2011. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. Neurosci. Biobehav. Rev. 35, 573-588.

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M.E., Clare, L., 2005. A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. Neuropsychol. Rev. 15, 73-95.

Gambini, O., Campana, A., Garghentini, G., Scarone, S., 2003. No evidence of a homogeneous frontal neuropsychological profile in a sample of schizophrenic subjects. J. Neuropsychiatry Clin. Neurosci. 15, 53-57.

Geisler, D., Walton, E., Naylor, M., Roessner, V., Lim, K.O., Schulz, S.C., Gollub, R.L., Calhoun, V.D., Sponheim, S.R., Ehrlich, S., 2015. Brain structure and function correlates of cognitive subtypes in schizophrenia. Psychiatr. Res. Neuroimaging 234, 74-83.

Gilbert, E., Mérette, C., Jomphe, V., Émond, C., Rouleau, N., Bouchard, R.H., Roy, M.A., Paccalet, T., Maziade, M., 2014. Cluster analysis of cognitive deficits may mark heterogeneity in schizophrenia in terms of outcome and response to treatment. Eur. Arch. Psychiatr. Clin. Neurosci. 264, 333-343.

Goldstein, G., 1990. Neuropsychological heterogeneity in schizophrenia: A consideration of abstraction and problem-solving abilities. Arch. Clin. Neuropsychol. 5, 251-264.

Goldstein, G., Allen, D.N., Seaton, B.E., 1998. A comparison of clustering solutions for cognitive heterogeneity in schizophrenia. J. Int. Neuropsychol. Soc. 4, 353-362.

González-Blanch, C., Rodríguez-Sánchez, J.M., Pérez-Iglesias, R., Pardo-García, G., Martínez-García, O., Vázquez-Barquero, J.L., Crespo-Facorro, B., 2010. First-episode schizophrenia patients neuropsychologically within the normal limits: Evidence of deterioration in speed of processing. Schizophr. Res. 119, 18-26.

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

Guimond, S., Chakravarty, M.M., Bergeron-Gagnon, L., Patel, R., Lepage, M., 2016. Verbal memory impairments in schizophrenia associated with cortical thinning. NeuroImage: Clin. 11, 20-29.

Harvey, P.D., Heaton, R.K., Carpenter, W.T., Green, M.F., Gold, J.M., Schoenbaum, M., 2012. Functional impairment in people with schizophrenia: Focus on employability and eligibility for disability compensation. Schizophr. Res. 140, 1-8.

Harvey, P.D., Reichenberg, A., Bowie, C.R., Patterson, T.L., Heaton, R.K., 2010. The course of neuropsychological performance and functional capacity in older patients with schizophrenia: Influences of previous history of long-term institutional stay. Bio. Psychiatr. 67, 933-939.

Harvey, P.D., Strassnig, M., 2012. Predicting the severity of everyday functional disability in people with schizophrenia: Cognitive deficits, functional capacity, symptoms, and health status. World Psychiatr. 11, 73-79.

Heinrichs, R.W., 2005. The primacy of cognition in schizophrenia. Am. Psychol. 60, 229.

Heinrichs, R.W., 2007. Cognitive improvement in response to antipsychotic drugs: Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE trial. Arch. Gen. Psychiatr. 64, 631-632.

Heinrichs, R.W., Awad, A.G., 1993. Neurocognitive subtypes of chronic schizophrenia. Schizophr. Res. 9, 49-58.

Heinrichs, R.W., Miles, A.A., Smith, D., Zargarian, T., Vaz, S.M., Goldberg, J.O., Ammari, N., 2008. Cognitive, clinical, and functional characteristics of verbally superior schizophrenia patients. Neuropsychol. 22, 321-328.

Heinrichs, R.W., Parlar, M., Pinnock, F., 2017. Normal-range verbal-declarative memory in schizophrenia. Neuropsychol. 31, 778-786.

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

Heinrichs, R.W., Ruttan, L., Zakzanis, K.K., Case, D., 1997. Parsing schizophrenia with neurocognitive tests: Evidence of stability and validity. Brain Cognit. 35, 207-224.

Heinrichs, R.W., Zakzanis, K.K., 1998. Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. Neuropsychol. 12, 426-445.

Hill, S.K., Ragland, J.D., Gur, R.C., Gur, R.E., 2002. Neuropsychological profiles delineate distinct profiles of schizophrenia, an interaction between memory and executive function, and uneven distribution of clinical subtypes. J. Clin. Exp. Neuropsychol. 24, 765-780.

Holthausen, E.A., Wiersma, D., Sitskoorn, M.M., Hijman, R., Dingemans, P.M., Schene, A.H., van den Bosch, R.J., 2002. Schizophrenic patients without neuropsychological deficits: subgroup, disease severity or cognitive compensation? Psychiatr. Res. 112, 1-11.

Insel, T.R., 2010. Rethinking schizophrenia. Nature 468, 187.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31, 651-666.

Joyce, E.M., Hutton, S.B., Mutsatsa, S.H., Barnes, T.R.E., 2005. Cognitive heterogeneity in first-episode schizophrenia. Br. J. Psychiatr. 187, 516-522.

Kahn, R.S., Keefe, R.S.E., 2013. Schizophrenia is a cognitive illness: Time for a change in focus. JAMA Psychiatr. 70, 1107-1112.

Keefe, R.S.E., Bilder, R.M., Davis, S.M., Harvey, P.D., Palmer, B.W., Gold, J.M., Meltzer, H.Y., Green, M.F., Capuano, G., Stroup, T.S., McEvoy, J.P., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., Davis, C.E., Hsiao, J.K., Lieberman, J.A., 2007. Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE trial. Am. J. Geriatr. Psychiatr. 64, 633-647.

Keefe, R.S.E., Eesley, C.E., Poe, M.P., 2005. Defining a cognitive function decrement in schizophrenia. Bio. Psychiatr. 57, 688-691.

Klingberg, S., Wittorf, A., Sickinger, S., Buchkremer, G., Wiedemann, G., 2008. Course of cognitive functioning during the stabilization phase of schizophrenia. J. Psychiatr. Res. 42, 259-267.

Kravariti, E., Morgan, K., Fearon, P., Zanelli, J.W., Lappin, J.M., Dazzan, P., Morgan, C., Doody, G.A., Harrison, G., Jones, P.B., Murray, R.M., Reichenberg, A., 2009. Neuropsychological functioning in first-episode schizophrenia. Br. J. Psychiatr. 195, 336-345.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Toomey, R., Tsuang, M.I., 2004. Heterogeneity of schizophrenia: A study of individual neuropsychological profiles. Schizophr. Res. 71, 307-321.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Toomey, R., Tsuang, M.T., 2000. The paradox of normal neuropsychological function in schizophrenia. J. Ab. Psychol. 109, 743.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Tsuang, M.T., 2001. Intelligence quotient and neuropsychological profiles in patients with schizophrenia and in normal volunteers. Bio. Psychiatr. 50, 453-462.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Tsuang, M.T., 2008. IQ decline in cross-sectional studies of schizophrenia: Methodology and interpretation. Psychiatr. Res. 158, 181-194.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics, 159-174.

Leeson, V.C., Sharma, P., Harrison, M., Ron, M.A., Barnes, T.R.E., Joyce, E.M., 2011. IQ trajectory, cognitive reserve, and clinical outcome following a first episode of psychosis: A 3-year longitudinal study. Schizophr. Bul. 37, 768-777.

Liu, C.M., Fann, C.S., Chen, C.Y., Liu, Y.L., Oyang, Y.J., Yang, W.C., Chang, C.C., Wen, C.C., Chen, W.J., Hwang, T.J., Hsieh, M.H., Liu, C.C., Faraone, S.V., Tsuang, M.T., Hwu, H.G., 2011. ANXA7, PPP3CB, DNAJC9, and ZMYND17 genes at chromosome 10q22 associated with the subgroup of schizophrenia with deficits in attention and executive function. Bio. Psychiatr. 70, 51-58.

MacCabe, J.H., Brébion, G., Reichenberg, A., Ganguly, T., McKenna, P., Murray, R.M., David, A.S., 2012. Superior intellectual ability in schizophrenia: Neuropsychological characteristics. Neuropsychol. 26, 181-190.

MacCabe, J.H., Lambe, M.P., Cnattingius, S., Torrång, A., Björk, C., Sham, P.C., David, A.S., Murray, R.M., Hultman, C.M., 2008. Scholastic achievement at age 16 and risk of schizophrenia and other psychoses: A national cohort study. Psychol. Med. 38, 1133-1140.

McDermid Vaz, S., Heinrichs, R.W., 2006. Stability and validity of memory-based subtypes of schizophrenia. J. Int. Neuropsychol. Soc. 12, 782-791.

Mesholam-Gately, R.I., Giuliano, A.J., Goff, K.P., Faraone, S.V., Seidman, L.J., 2009. Neurocognition in first-episode schizophrenia: A meta-analytic review. Neuropsychol. 23, 315-336.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J.A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J.J., Devereaux, P.J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P.C., Grimshaw, J., Guyatt, G., Higgins, J., Ioannidis, J.P.A., Kleijnen, J., Lang, T., Magrini, N., McNamee, D., Moja, L., Mulrow, C., Napoli, M., Oxman, A., Pham, B., Rennie, D., Sampson, M., Schulz, K.F., Shekelle, P.G., Tovey, D., Tugwell, P., 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 6.

Morissette, L., Chartier, S., 2013. The k-means clustering technique: General considerations and implementation in Mathematica. Tutorials in Quant. Methods Psychol. 9, 15-24.

Moses, J.A., 1983. Schizophrenic subgroups with normal and abnormal cognitive functioning on the luria-nebraska neuropsychological battery. Int. J. Neurosci. 21, 129-135.

Muharib, E., Heinrichs, R.W., Miles, A., Pinnock, F., Vaz, S.M., Ammari, N., 2014. Community outcome in cognitively normal schizophrenia patients. J. Int. Neuropsychol. Soc. 20, 805-811.

Nuechterlein, K.H., Green, M.F., Kern, R.S., Baade, L.E., Barch, D.M., Cohen, J.D., Essock, S., Fenton, W.S., Frese Iii, 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. Am. J. Psychiatr. 165, 203-213.

Ortiz-Gil, J., Pomarol-Clotet, E., Salvador, R., Canales-Rodríguez, E.J., Sarró, S., Gomar, J.J., Guerrero, A., Sans-Sansa, B., Capdevila, A., Junqué, C., McKenna, P.J., 2011. Neural correlates of cognitive impairment in schizophrenia. Br. J. Psychiatr. 199, 202-210.

Palmer, B.W., Heaton, R.K., Paulsen, J.S., Kuck, J., Braff, D., Harris, M.J., Zisook, S., Jeste, D.V., 1997. Is it possible to be schizophrenic yet neuropsychologically normal? Neuropsychol. 11, 437-446.

Paulsen, J.S., Heaton, R.K., Sadek, J.R., Perry, W., Delis, D.C., Braff, D., Kuck, J., Zisook, S., Jeste, D.V., 1995. The nature of learning and memory impairments in schizophrenia. J. Int. Neuropsychol. Soc. 1, 88-99.

Potter, A.I., Nestor, P.G., 2010. IQ subtypes in schizophrenia: Distinct symptom and neuropsychological profiles. J. Nerv. Ment. Dis. 198, 580-585.

Ruiz, J.C., Soler, M.J., Fuentes, I., Tomás, P., 2007. Intellectual functioning and memory deficits in schizophrenia. Compr. Psychiatr. 48, 276-282.

Rüsch, N., Spoletini, I., Wilke, M., Bria, P., Di Paola, M., Di Iulio, F., Martinotti, G., Caltagirone, C., Spalletta, G., 2007. Prefrontal-thalamic-cerebellar gray matter networks and executive functioning in schizophrenia. Schizophr. Res. 93, 79-89.

Sauvé, G., Malla, A., Joober, R., Brodeur, M.B., Lepage, M., 2018. Comparing cognitive clusters across first- and multiple-episode of psychosis. Psychiatr. Res. 269, 707-718.

Schaefer, J., Giangrande, E., Weinberger, D.R., Dickinson, D., 2013. The global cognitive impairment in schizophrenia: Consistent over decades and around the world. Schizophr. Res. 150, 42-50.

Seaton, B.E., Allen, D.N., Goldstein, G., Kelley, M.E., Van Kammen, D.P., 1999. Relations between cognitive and symptom profile heterogeneity in schizophrenia. J. Nerv. Ment. Dis. 187, 414-419.

Shepherd, A.M., Quide, Y., Laurens, K.R., O'Reilly, N., Rowland, J.E., Mitchell, P.B., Carr, V.J., Green, M.J., 2015. Shared intermediate phenotypes for schizophrenia and bipolar disorder: neuroanatomical features of subtypes distinguished by executive dysfunction. J. Psychiatr. & Neurosci. 40, 58-68.

Tandon, R., Nasrallah, H.A., Keshavan, M.S., 2009. Schizophrenia, "just the facts" 4. Clinical features and conceptualization. Schizophr. Res. 110, 1-23.

Torres, I.J., Flashman, L.A., O'Leary, D.S., Swayze Ii, V., Andreasen, N.C., 1997. Lack of an association between delayed memory and hippocampal and temporal lobe size in patients with schizophrenia and healthy controls. Bio. Psychiatr. 42, 1087-1096.

Turetsky, B.I., Moberg, P.J., Mozley, L.H., Moelter, S.T., Agrin, R.N., Gur, R.C., Gur, R.E., 2002. Memory-delineated subtypes of schizophrenia: relationship to clinical, neuroanatomical, and neurophysiological measures. Neuropsychol. 16, 481-490.

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. Schizophr. Bull. 44, 560-574.

Van Rheenen, T., Cropley, V., Fagerlund, B., Wannan, C.M.J., Bruggemann, J., Lenroot, R.K., Sundram, S., Weickert, C.S., Weickert, T.W., Zalesky, A., 2019. Cognitive reserve attenuates age-related cognitive decline in the context of putatively accelerated brain ageing in schizophrenia-spectrum disorders. Psychol. Med. 1-15.

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. Psychol. Med. 47, 1848-1864.

Van Rheenen, T., Rossell, S., 2016. Facial emotion recognition impairments in bipolar disorder. A cognitive problem? J. Int. Neuropsychol. Soc. 22, 583-585.

Vaskinn, A., Sundet, K., Friis, S., Ueland, T., Simonsen, C., Birkenaes, A.B., Engh, J.A., Opjordsmoen, S., Andreassen, O.A., 2008. Can learning potential in schizophrenia be assessed with the standard CVLT-II? An exploratory study. Scand. J. Psychol. 49, 179-186.

Vaskinn, A., Ueland, T., Melle, I., Agartz, I., Andreassen, O.A., Sundet, K., 2014. Neurocognitive decrements are present in intellectually superior schizophrenia. Front. Psychiatr. 5.

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. Arch. Gen. Psychiatr. 57, 907-913.

Wieland, J., Zitman, F.G., 2016. It is time to bring borderline intellectual functioning back into the main fold of classification systems. Psychiatrist 40, 204-206.

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 Psychiatr. 73, 1251-1259.

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 Schizophr. 1, 15043.

Wexler, B.E., Zhu, H.T., Bell, M.D., Nicholls, S.S., Fulbright, R.K., Gore, J.C., Colibazzi, T., Amat, J., Bansal, R., Peterson, B.S., 2009. Neuropsychological near normality and brain structure abnormality in schizophrenia. Am. J. Psychiatr. 166, 189-195.

Wilk, C.M., Gold, J.M., McMahon, R.P., Humber, K., Iannone, V.N., Buchanan, R.W., 2005. No, it is not possible to be schizophrenic yet neuropsychologically normal. Neuropsychol. 19, 778-786.

Woodbury, M.A., Clive, J., Garson Jr, A., 1978. Mathematical typology: A grade of membership technique for obtaining disease definition. Comput. Biomed. Res. 11, 277-298.

Woodward, N.D., Heckers, S., 2015. Brain structure in neuropsychologically defined subgroups of schizophrenia and psychotic bipolar disorder. Schizophr. Bull. 41, 1349-1359.

Wykes, T., Huddy, V., Cellard, C., McGurk, S.R., Czobor, P., 2011. A meta-analysis of cognitive remediation for schizophrenia: Methodology and effect sizes. Am. J. Psychiatr. 168, 472-485.

Yim, O., Ramdeen, K.T., 2015. Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. The Quant. Methods Psychol. 11, 8-21.

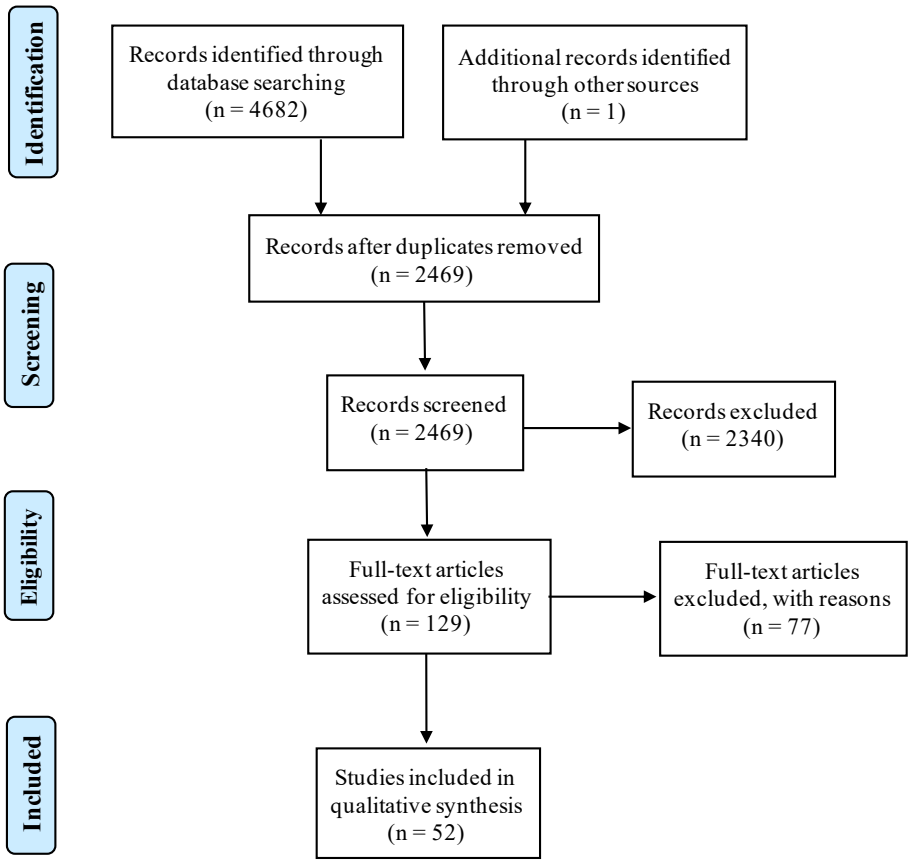

Figure 1. PRISMA flow diagram

Table 1. Summary of group comparisons for studies employing criterion-related subgrouping based on current cognitive function.

Study	Subgroups: n (% of N)	Pairwise Comparison	d_μ	d_{range}	Pairwise Comparison	d_μ	d_{range}	Pairwise Comparison	d_μ	d_{range}	Pairwise Comparison	d_μ	d_{range}	Pairwise Comparison	d_μ	d_{range}	Pairwise Comparison	d_μ	d_{range}
Intellectual Functioning		1 = 2			1 > 3			1 = 3			2 > 3			2 = 3			1 = A		
Ayesa-Arriola et al. (2018)	1. High epIQ: 30 (10) 2. Normal epIQ: 178 (61) 3. Low epIQ: 84 (29)	AV, EF, GF, MC, SoP, VerL, VisL *n* variables: 8	0.42	0.09 to 0.79	EF, GF, SoP, VerL, VisL *n* variables: 6	1.16	0.64 to 1.86	AV, MC *n* variables: 2	0.49	0.39 to 0.59	EF, GF, SoP, VerL, VisL *n* variables: 6	0.60	0.42 to 0.93	AV, MC *n* variables: 2	0.25	0.16 to 0.34	AV, EF, GF MC, SoP, VerL, VisL, WM *n* variables: 8	0.61	0.19 to 0.97
Note: FEP	A. HC High epIQ: 24 (12) B. HC Normal epIQ: 146 (73) C. HC Low epIQ: 29 (15)																		
	N variables: 8	2 = B			3 < C			3 = C											
		AV, EF, GF MC, SoP, VerL, VisL, WM *n* variables:8	0.77	0.42 to 1.32	AV, EF, GF, MC, SoP, VerL *n* variables: 6	1.03	0.75 to 1.82	VisL, WM *n* variables: 2	0.38	0.35 to 0.41									
		1 < A			1 = A			2 < B			2 = B								
Kremen et al. (2001)	1. Average IQ: 20 (27) 2. Low Average IQ: 16 (21) A. HC Average IQ: 21 (23) B. HC Low Average IQ: 15 (17)	AV, EF, VerL *n* variables: 3	NA	NA	MC, RaPS, SoP, VS *n* variables: 4	NA	NA	AV, EF, SoP, VerL *n* variables: 4	NA	NA	MC, RaPS, VS *n* variables: 3	NA	NA						
	N variables: 7																		
		1 > 2			1 = 2			1 < A			1 = A			2 < A			2 = A		
Ruiz et al. (2007)	1. High IQ: 21 (48) 2. Low IQ: 23 (52) A. HC: 34 (100)	EF, RaPS, SoP, WM, VC, Mem *n* variables: 9	1.27	0.89 to 2.00	EF, Fl, Mem, VerL, VisL *n* variables: 11	0.38	0.02 to 0.74	EF, Fl SoP *n* variables: 5	0.99	0.60 to 1.79	EF, Fl, Mem, RaPS, VC, VerL, VisL, WM *n* variables: 15	0.31	0.02 to 0.78	EF, Fl, Mem, RaPS, SoP, VC, WM, *n* variables: 15	1.92	0.79 to 4.42	Mem, VerL, VisL *n* variables: 5	0.38	0.12 to 0.53
	N variables: 20																		
		1 < A			1 > A			1 = A											
Wilk et al. (2005)	1. High Average IQ: 13 (20) A. HC High Average IQ: 13 (22)	Mem *n* variables: 1	NA	NA	RaPS *n* variables: 1	NA	NA	Mem, SoP, VC, WM *n* variables: 4	NA	NA									
	N variables: 6																		
Neuropsychological Functioning		1 > 2			1 = 2			1 > 3			1 = 3			2 > 3			2 = 3		
Ammari et al. (2010)[i]	1. Cognitively Normal: 24 (16) 2. Verbal Memory Impaired: 26 (17) 3. Generalized Impairment: 23 (15) A. HC: 18 (100)	VerL *n* variables: 1	2.93		AV, epIQ, FL, SoP, RaPS, VC, WM *n* variables: 8	0.26	0.05 to 0.77	epIQ, FL, RaPS, VC, WM, VerL *n* variables: 6	2.53	0.94 to 4.10	AV, FL, SoP *n* variables: 3	0.73	0.84 to 0.98	epIQ, Fl RaPS, VC, WM *n* variables: 5	1.94	0.73 to 3.03	AV, Fl, SoP, VerL *n* variables: 4	0.55	0.44 to 0.69
	N variables: 9																		
		1 = A			2 < A			2 = A			3 < A			3 = A					
		AV, epIQ, Fl, RaPS, SoP, VC, VerL, WM *n* variables: 9	0.22	0.02 to 0.48	VerL, WM *n* variables: 2	1.71	1.18 to 2.24	AV, epIQ, FL, RaPS, SoP, VC *n* variables: 7	0.26	0.05 to 0.47	epIQ, FL, RaPS, SoP, VC, VerL WM *n* variables: 8	2.20	1.03 to 3.76	AV *n* variables: 1	0.50				
		1 > 2			1 = 2														
Ayesa-Arriola et al. (2013)	1. Non-Deficit: 60 (49) 2. Deficit: 86 (51)	AV, EF, epIQ, SoP, MC, VerL *n* variables: 7	0.93	0.73 to 1.37	VisL *n* variables: 1	0.47													
Note: FEP	*N* variables: 8																		
		1 > 2			1 < A			1 = A			2 < A								
González-Blanch et al. (2010)	1. Cognitive Normal: 25 (23) 2. Cognitive Impaired: 86 (77) A. HC: 28 (100)	AV, EF, GF, MC, VC, VerL, VisL *n* variables: 8	1.25	0.85 to 1.88	MC *n* variables: 1	0.87		AV, EF, GF, VC, VerL, VisL *n* variables: 7	0.34	0.21 to 0.47	AV, EF, GF, MC, VC, VerL, VisL *n* variables: 8	1.30	1.01 to 1.95						
Note: FEP	*N* variables: 8																		
		1 > 2			1 = 2			1 = A			2 < A			2 = A			1 > B		
Heinrichs et al. (2008)	1. Verbally Superior: 25 (17) 2. Comparison Patients: 126 (83) A. HC Verbally Superior: 22 (31)	AV, epIQ, FL, RaPS, SoP, VC, VerL	1.17	0.78 to 1.33	AV *n* variables: 1	0.15		AV, epIQ, FL, RaPS, SoP,	0.34	0.05 to 0.56	AV, epIQ, FL, RaPS, SoP,	0.34	0.05 to 0.56	AV, epIQ, FL, RaPS, SoP,	0.34	0.05 to 0.56	epIQ, VC *n* variables: 2	1.98	1.03 to 2.93

	B. HC Comparison: 50 (69) *N* variables: 9	*n* variables: 8						VC, VerL, WM *n* variables: 9			VC, VerL, WM *n* variables: 9			VC, VerL, WM *n* variables: 9				
		1 = B AV, FL, RaPS, SoP, VerL, WM *n* variables: 7	0.28	0.05 to 0.54	2 < B SoP, VerL *n* variables: 2	0.80	0.69 to 0.90											
Heinrichs et al. (2015)[i]	1. Cognitively Normal Range: 18 (18) 2. Cognitively Below Normal Range: 19 (20) A. HC Cognitively Normal Range: 24 (17) B. HC Cognitively Below Normal Range: 20 (14) 1 *v.* 2: *N* variables: 4 1, 2 *v.*A, B: *N* variables:12	1 > 2 epIQ, RaPS, SC *n* variables: 3	2.35	0.72 to 5.43	1 = 2 SC *n* variables: 1	0.73		1 < A SoP *n* variables: 1	0.92		1 = A AV, epIQ GF, RaPS, SC, VerL, VisL, WM *n* variables: 11	0.28	0.07 to 0.82	2 = B AV, epIQ GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 12	0.23	0.00 to 0.73		
Holthausen et al. (2002) Note: FEP	1. Cognitively Normal: 23 (20) 2. Cognitively Impaired: 95 (80) A. HC: 45 (100) *N* variables: 9	1 > 2 AV, Fl, SoP VerL, VisL, VS *n* variables: 9	0.81	0.48 to 1.13	1 < A SoP VerL *n* variables: 2	0.67	0.62 to 0.71	1 = A AV, Fl, VisL, VS *n* variables: 7	0.17	0.00 to 0.35	2 < A AV, Fl, SoP, VerL, VisL, VS *n* variables: 9	0.94	0.38 to 1.63						
Kremen et al. (2000)	1. Neuropsychologically Within Normal Limits: 17 (23) 2. Neuropsychologically Abnormal: 58 (77) A. HC *N* variables: 8	1 > 2 AV, EF, MC, SoP, VC, VerL, VS, WM *n* variables: 8	NA	NA	1 < A EF, SoP *n* variables: 2	NA	NA	1 = A AV, MC, VC, VerL, VS, WM *n* variables: 6	NA	NA	2 < A AV, EF, MC, SoP, VC, VerL, VS, WM *n* variables: 8	NA	NA						
Moses (1983)	1. Normal: 69 (69) 2. Borderline: 16 (16) 3. Abnormal: 15 (15) *N* variables: 14	1 > 2 Intel, LHem, MC, Mem, Patho, RHem Rhy, Tac, VS *n* variables: 9	1.10	0.76 to 1.52	1 = 2 Arith, Read, Speech, Write *n* variables: 5	0.46	0.07 to 0.72	1 > 3 Arith, Intel, LHem, MC, Mem, Patho, Read, RHem Rhy, Speech, Tac, VS, Write *n* variables: 14	1.90	0.98 to 3.69	2 > 3 Arith, Read, RHem, Speech, Write *n* variables: 6	0.96	0.37 to 1.34	2 = 3 Intel, LHem, MC, Mem, Patho, Rhy, Tac, VS *n* variables: 8	0.55	0.17 to 1.02		
Muharib et al. (2014)	1. Normal Range Cognition: 14 (14) 2. Cognitively Impaired: 21 (21) A. HC Normal Range Cognition: 21 (26) *N* variables: 10	1 > 2 AV, cIQ, epIQ, GF, RaPS, SoP, VerL, VisL, WM *n* variables: 9	1.78	1.05 to 3.30	1 = 2 SC *n* variables: 1	0.71		1 = A AV, cIQ, epIQ, GF, RaPS, SC SoP, VerL, VisL, WM *n* variables: 10	0.26	0.01 to 0.68	2 < A AV, cIQ, epIQ, GF, RaPS, SC SoP, VerL, VisL, WM *n* variables: 10	1.80	1.03 to 3.52						
Ortiz-Gil et al. (2011)	1. Cognitively Preserved: 23 (47) 2. Cognitively Impaired: 26 (53) A. HC: 39 (100) 1 *v.* 2: *N* variables: 6 SSD *v.* HC: *N* variables: 4	1 > 2 EF, Mem *n* variables: 2	1.54	1.29 to 1.78	1 = 2 cIQ, epIQ, RaPS, VC *n* variables: 4	0.53	0.42 to 0.59	1 < A cIQ, RaPS *n* variables: 2	0.33	0.30 to 0.36	1 = A epIQ, VC *n* variables: 2	0.11	0.08 to 0.14	2 < A cIQ, RaPS *n* variables: 2	0.88	0.83 to 0.92	2 = A epIQ, VC *n* variables: 2	0.44	0.37 to 0.50
Palmer et al. (1997)	1. Neuropsychologically Normal: 44 (28) 2. Neuropsychologically Impaired: 124 (72) A. HC: 63 (100) *N* variables: 9	1 > 2 AV, EF, GF, MC, Mem, SA, SoP, VC *n* variables: 9	0.92	0.46 to 1.33	1 < A Mem *n* variables: 1	0.77		1 = A AV, EF, GF, MC, Mem, SA, SoP, VC *n* variables: 8	0.35	0.12 to 0.62	2 < A AV, EF, GF, MC, Mem, SA, SoP, VC *n* variables: 9	1.12	0.37 to 1.84						
Memory																			
Bowie et al. (2004)	1. Unimpaired: 34 (4) 2. Subcortical: 156 (26) 3. Cortical: 238 (40) *N* variables: 7	1 > 2 Fl, Mem, VC, VerL *n* variables: 7	1.25	0.83 to 4.25	1 > 3 Fl, Mem, VC, VerL *n* variables: 7	4.30	3.29 to 18.40	2 > 3 Fl, Mem, VC, VerL *n* variables: 7	2.05	0.38 to 5.91									
		1 > 2			1 = 2			1 < A			1 = A			2 < A			2 = A		

Guimond et al. (2016)	1. Low to Mild Verbal Memory Impairment: 23 (46) 2. Moderate to Severe Verbal Memory Impairment: 27 (54) A. HC: 23 (100) 1 *v.* 2; *N* variables 9 1, 2 *v.* A; *N* variables: 2	cIQ, EF, VerL *n* variables: 4	1.73	0.74 to 3.19	AV, SC, SoP, VisL, WM *n* variables: 5	0.42	0.27 to 0.84	cIQ *n* variables: 1	1.00		VerL *n* variables: 1	0.57		cIQ *n* variables: 1	1.69		VerL *n* variables: 1	3.02	
		1 > 2			1 = 2			1 = A			2 < A			2 = A					
Heinrichs et al. (2017)[1]	1. Normal Range: 23 (17) 2. Below Normal Range: 129 (83) A. HC Normal Range: 51 (69) 1 *v.* 2; *N* variables 5 1, 2 *v.* A; *N* variables: 13	epIQ, VC, VerL *n* variables: 4	0.77	0.55 to 0.91	WM *n* variables: 1	0.40		epIQ, VC, VerL, WM *n* variables: 13	0.15	0.01 to 0.49	epIQ, VC, VerL *n* variables: 4	0.77	0.55 to 0.91	VerL, WM *n* variables: 9	1.22	0.40 to 1.98			
		1 > 2			1 = 2														
McDermid Vaz and Heinrichs (2006)	1. Memory Unimpaired: 12 (22) 2. Memory Impaired: 43 (78) *N* variables: 5	VerL *n* variables: 1	1.22		cIQ, EF, MC, VerL *n* variables: 4	0.23	0.08 to 0.65												
		1 > 2			1 = 2														
Torres et al. (1997)	1. High Memory: 10 (11) 2. Low Memory: 10 (11) *N* variables: 8	cIQ, Mem, VerL, VisL *n* variables: 8	2.98	0.34 to 8.15	Fl *n* variables: 1	0.84													
		1 > 2			1 = 2			1 > 3			1 = 3			2 > 3			2 = 3		
Vaskinn et al. (2008)	1. High Achievers: 24 (21) 2. Learners: 61 (56) 3. Non-learners: 25 (23) *N* variables: 28	VerL *n* variables: 8	1.34	0.95 to 2.68	cIQ, EF, Fl, MC, SoP, VerL, WM *n* variables: 20	0.36	0.07 to 1.09	VerL *n* variables: 12	2.01	0.89 to 2.74	cIQ, EF, Fl, MC, SoP, VerL, WM *n* variables: 16	0.65	0.01 to 0.78	VerL *n* variables: 9	1.48	0.88 to 2.78	cIQ, EF, Fl, MC, SoP, VerL, WM *n* variables: 19	0.43	0.09 to 0.85
		1 > 2			1 = 2			1 = A			2 < A								
Wexler et al. (2009)	1. Neuropsychologically Near Normal: 14 (17) 2. Neuropsychologically Impaired: 32 (40) A. HC: 22 (100) *N* variables: 2	VerL *n* variables: 1	0.86		AV *n* variables: 1	0.61		AV, VerL *n* variables: 2	0.39	0.10 to 0.65	AV, VerL *n* variables: 2	1.01	0.51 to 1.52						
Executive Function		1 > 2			1 = 2														
Butler et al. (1992)	1. WCST Non-Impaired: 20 (46) 2. WCST Impaired: 24 (54) *N* variables: 6	EF, Fl, RT, VC *n* variables: 4	1.31	0.84 to 2.36	Fl, WM *n* variables: 2	0.70	0.70 to 0.70												
		1 > 2			1 = 2														
Rüsch et al. (2007)	1. High WCST Performers: 30 (57) 2. Poor WCST Performers: 21 (37) *N* variables: 6	EF, GF *n* variables: 2	1.37	1.18 to 1.56	WM *n* variables: 4	0.43	0.25 to 0.69												

Note: Only statistical pairwise comparisons with one or more variables included are presented; [1]Overlapping samples; AV, attention-vigilance; Arith, arithmetic; cIQ, current IQ; EF, executive function; epIQ, estimated premorbid IQ; FEP, first episode psychosis; GF, global functioning; Fl, fluency; HC, healthy control; Intel, intellect; IQ, intelligence quotient; LHem; left hemisphere; Mem, memory; MC, motor-control; NA, not available; Patho, pathognomonic; RaPS, reasoning and problem solving; RHem, right hemisphere; Rhy, rhythm; RT, reaction time; SA, sensory ability; SC, social cognition; SoP, speed of processing; SSD, schizophrenia spectrum disorder; Tac, tactile; VC, verbal comprehension; VerL, verbal learning; VisL, visual learning; VS, visuospatial; WM, working memory

Table 2. Summary of data-driven methodologies employed.

Study	Sample	Domains entered into subgrouping analysis	Preprocessing	Subgrouping analysis	Distance	Basis for the number of clusters	Presented:		Validation Procedure	*Post hoc* comparison of variables:			Subgroup: n (% of N)
							Dendrogram	Agglomeration Schedule Plot		Included in subgrouping analysis	Not included in subgrouping analysis	Against HC	
Carruthers et al. (2019)	210 SSD 192 HC	EF	Z-standardized: HC	Ward's+ *k*-means	sEd	Visual inspection of dendrogram + meaningful jump in agglomeration schedule coefficient	Yes	Yes	Classification concordance over multiple methods	Yes	No	Yes	Relatively Intact: 72 (34) Moderate Impairment: 114 (54) Severe Impairment: 24 (12)
Cobia et al. (2011)	79 SSD 65 HC	cIQ, EF, Fl, SoP, VerL, VisL, WM	Z-standardized: HC	Ward's+ *k*-means	sEd	Visual inspection of dendrogram	Yes	No	Classification concordance over multiple methods	Yes	No	Yes	Neuropsychologically Near Normal: 45 (57) Neuropsychologically Impaired: 34 (43)
Dawes et al. (2011)	144 SSD	AV, EF, PerOrg, SoP, VC, VerL, VisL	T-score standardized: Normative data	Pearson's Correlation Coefficient + *k*-means	sEd	Visual inspection of dendrogram + meaningful jump in agglomeration schedule coefficient	No	No	No	No	No	NA	K1: 19 (13) K2: 38 (26) K3: 40 (28) K4: 17 (12) K5: 30 (21)
Gambini et al. (2003)	81 SSD	EF, Fl	Z-standardized: Unspecified	Ward's	sEd	Visual inspection of dendrogram	No	No	No	Yes	No	NA	[#]Cluster 1: 30 (37) Cluster 2: 39 (48) [$]Cluster 3: 9 (11) Cluster 4: 3 (4)
Geisler et al. (2015)	129 SSD 165 HC	Fl, Intel, MC, SD, SoP, VerL, VisL, WM	Z-standardized: SSD + PCA scores	*k*-means	NA	Consistent with previous research	NA	NA	No	No	No	No	Diminished Fl: 38 (30) Diminished VerL + MC: 26 (20) Diminished FM + SoP: 21 (16) Diminished IQ: 44 (34)
Gilbert et al. (2014)	112 SSD	SoP, VerL, VisL, WM	Z-standardized: Normative data	GMA	NA	Comparing the adjustment of different clustering solutions using the Bayesian Information Criterion	NA	NA	Overlap with neuropsychologist clinical classification	No	Yes	NA	Near-normal Functioning: 48 (43) Selectively Impaired: 46 (41) Generally Impaired: 18 (16)
		Ab	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	Near Normal: 85 (38) Moderately Impaired: 45 (20) Moderately Impaired + Psychomotor Deficit: 54 (24) Severely Impaired: 40 (18)
Goldstein et al.(1998)	221 SSD	cIQ	NS	Ward's	sEd	Visual inspection of dendrogram + discriminant function plot	No	No	Classification concordance across the three solutions	No	No	NA	Average or Above: 19 (10) Moderately Poor: 50 (25) Average Verbal IQ + Moderately Poor Performance IQ: 76 (38) Severely Poor: 54 (27)
		Ab, PerIQ, VerIQ	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	Near Normal: 59 (26) Moderately Impaired: 104 (46) Moderately Impaired + Psychomotor Deficit: 26 (12) Severely Impaired: 35 (16)
Green et al. (2013)[1]	617 SSD 764 HC	AV, cIQ, epIQ, Fl, Lang, Mem, VS	Categorical standardization	Grade of Membership Analysis	NA	"Considered a combination of the goodness of fit criterion + the utility of the solution when evaluating possible models"	NA	NA	No	Yes	No	Yes	Cognitively Spared: 323 (52) Cognitive Deficit: 294 (48)

Hill et al. (2002)	151 SSD	AV, EF, Lang, MC, SP, VerL, VisL	Z- standardized: HC	Ward's + MSSC	Ed	Visual inspection of dendrogram + meaningful jump in agglomeration schedule coefficient	No	No	No	No	Yes	NA	[$]Cluster 1: 15 (10) [#]Cluster 2: 76 (50) Cluster 3: 41 (27) Cluster 4: 19 (13)
Liu et al. (2011)	549 SSD 345 HC	AV, EF	Z- standardized: HC	Average Linkage	Ed	Not stated	No	No	No	Yes	No	No	Non-deficit: 106 (19) EF Deficit: 109 (20) AV Deficit: 146 (27) EF + AV Deficit: 188 (34)
McDermid Vaz & Heinrichs (2006)	102 SSD	VerL	Z-standardized: Normative data	*k*-means	NA	Consistent with previous research	NA	NA	No	Yes	Yes	NA	Unimpaired: 40 (39) Subcortical: 42 (41) Cortical: 20 (20)
Potter and Nestor (2010)	73 SSD 74 HC	epIQ, FSIQ	NS	*k*-means	NA	Consistent with previous research	NA	NA	No	Yes	Yes	Yes	Preserved: 21 (29) Deteriorated: 21 (29) Compromised: 31 (42)
Sauvé et al. (2018)	121 SSD 80 FEP 55 HC	AV, EF, RaPS SC, SoP, VerL, VisL, WM	Z-standardized: HC	Complete Linkage + *k*-means	sEd	Visual inspection of dendrogram + meaningful jump in agglomeration schedule coefficient	Yes	Yes	Adjusted Rand Index + DFA evaluation of multiple method cluster solutions	Yes	No	No	No Impairment: 23 (12) Intermediately Impaired: 113 (56) Generally Impaired: 65 (32)
Seaton et al. (1999)	102 SSD	cIQ	Z-standardized: HC	Ward's	sEd	Visual inspection of dendrogram	No	No	Classification concordance with *k*-means + discriminant function plot	Yes	No	NA	High: NS Motor Deficit: NS Low: NS Impaired: NS
Turetsky et al. (2002)	116 SSD 129 HC	VerL	Z-standardized: HC	*k*-means	NA	Consistent with previous research	No	No	No	Yes	No	No	Unimpaired: 59 (51) Subcortical: 36 (31) Cortical: 21 (18)
Van Rheenen et al. (2017)	564 SSD 575 HC	AV, EF, RaPS SC, SoP, VerL, VisL, WM	Z-standardized: HC	Ward's	sEd	Visual inspection of dendrogram + DFA	No	Yes	Classification concordance with random split-half	Yes	No	Yes	Relatively Intact: 75 (13) Mild-Moderate: 262 (47) Relatively Severe: 227 (40)
Weinberg et al. (2016)	96 SSD	epIQ, cIQ	NS	Complete Linkage + *k*-means	sEd	Visual inspection of dendrogram	No	No	No	Yes	Yes	Yes	Putatively Preserved: 25 (26) Moderately Deteriorated: 33 (34) Severely Deteriorated: 27 (28) Compromised: 11 (12)
Wells et al. (2015)[1]	534 SSD 635 HC	AV, epIQ, Mem, VS, WM	Z-standardized: HC	Complete Linkage + *k*-means	sEd	Visual inspection of dendrogram	No	No	Consistent with previous research	Yes	Yes	Yes	Preserved: 157 (29) Deteriorated: 239 (45) Compromised: 138 (26)

Notes: [1]Sample overlap; [#]Denotes author defined best performing cluster; [$] Denotes author defined worst performing cluster; Ab, abstraction; AV, attention-vigilance; cIQ, current IQ; DFA, discriminant function analysis; Ed, Euclidean distance; EF, executive function; epIQ, estimated premorbid IQ; FEP, first episode psychosis; Fl, fluency; FM, face memory; GMA, gaussian mixture model; GoM, Grade of Membership; HC, healthy control; Intel, intellect; IQ, intelligence quotient; Lang, language; Mem, memory; MC, motor control; NA, not applicable; NS, not stated; PCA, principle component analysis; PerIQ, performance IQ; PerOrg, perceptual organization; MSSC, minimum sum-of-squares clustering; SC, social cognition; sEd, squared Euclidean distance; SD, signal detection; SoP, speed of processing; SSD, schizophrenia spectrum disorders; VerComp, verbal comprehension; VerL, verbal learning; VerIQ, verbal IQ; VisL, visual learning; VS, visuospatial; WM, working memory

Table 3. Summary of group comparisons for research employing data-driven subgrouping based on current cognitive function.

Study	Subgroups: n (% of N)	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}
Two Subgroups		1 > 2			1 = 2			1 < A			1 = A			2 < A					
Cobia et al. (2011)	1. Neuropsychologically Near Normal: 45 (57) 2. Neuropsychologically Impaired: 34 (43) A. HC: 65 (100)	EF, Fl, Mem, SoP, VC, WM *n* variables: 9	1.34	0.79 to 2.19	WM variables: 1	1.49		EF, Fl, Mem, SoP, VC, WM *n* variables: 9	0.86	0.48 to 1.56	RaPS *n* variables: 1	0.25		EF, Fl, Mem, RaPS, SoP, VC, WM *n* variables: 10	2.00	1.34 to 2.76			
	N variables: 10	1 > 2			1 < A			1 = A			2 > A								
Green et al. (2013)	1. Cognitively Spared: 323 (52) 2. Cognitive Deficit: 294 (48) A. HC: 764 (100)	AV, cIQ, epIQ, Fl, GF, Mem, VC, VS, WM *n* variables: 10	1.42	0.86 to 2.14	AV, cIQ, Fl, GF, Mem, VC, WM *n* variables: 8	0.47	0.25 to 0.71	epIQ, VS *n* variables: 2	0.12	0.09 to 0.15	AV, cIQ, epIQ, Fl, GF, Mem, VC, VS, WM *n* variables: 10	1.65	0.99 to 2.30						
	N variables: 10																		
Three Subgroups		1 > 2			1 > 3			1 = 3			2 > 3			2 = 3			1 < A		
Carruthers et al. (2019)	1. Relatively Intact: 72 (34) 2. Moderate Impairment: 114 (54) 3. Severe Impairment: 24 (12) A. HC: 192 (100)	EF *n* variables: 7	1.83	0.60 to 4.50	EF *n* variables: 6	6.60	1.5 to 14.8	EF *n* variables: 1	0.50		EF *n* variables: 4	2.45	0.90 to 5.20	EF *n* variables: 3	0.17	0.00 to 0.40	EF *n* variables: 4	0.70	0.40 to 0.90
	N variables: 7	1 = A			2 > A			2 = A			3 > A								
		EF *n* variables: 3	0.13	0.10 to 0.20	EF *n* variables: 6	2.83	1.00 to 4.40	EF *n* variables: 1	0.30		EF *n* variables: 7	4.87	1.40 to 12.90						
Gilbert et al. (2014)	1. Near-normal Functioning: 48 (43) 2. Selectively Impaired: 46 (41) 3. Generally Impaired: 18 (16)	1 > 2 AV, EF, MC *n* variables: 4	0.81	0.57 to 1.22	1 > 3 AV, EF, MC *n* variables: 4	1.10	0.87 to 1.21	2 > 3 AV *n* variables: 1	0.86		2 = 3 EF, MC *n* variables: 3	0.32	0.08 to 0.60						
	N variables: 4																		
McDermid Vaz and Heinrichs (2006)	1. Unimpaired: 40 (39) 2. Subcortical: 42 (41) 3. Cortical: 20 (20)	1 > 2 Mem *n* variables: 1	0.89		1 = 2 cIQ, EF, MC, Mem *n* variables: 4	0.32	0.02 to 0.51	1 > 3 cIQ, Mem *n* variables: 2	1.51	1.22 to 1.79	1 = 3 EF, MC, Mem *n* variables: 3	0.45	0.18 to 0.61	2 = 3 cIQ, EF, Mem, MC	0.33	0.10 to 0.59			
	N variables: 5																		
Van Rheenen et al. (2017)	1. Relatively Intact: 75 (13) 2. Mild-Moderate: 262 (47) 3. Relatively Severe: 227 (40) A. HC: 575 (100)	1 > 2 AV, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	1.01	0.58 to 1.35	1 > 3 AV, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	2.08	1.22 to 3.09	2 > 3 AV, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	1.07	0.49 to 1.98	1 < A SC, SoP, WM *n* variables: 3	0.36	0.31 to 0.43	1 = A AV, epIQ, RaPS, VerL, VisL *n* variables: 5	0.09	0.01 to 0.15	2 < A AV, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	0.96	0.62 to 1.48
	N variables: 8	3 < A AV, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	1.93	1.41 to 2.11															
Four Subgroups		1 > 2			1 = 2			1 > 3			1 = 3			1 > 4			2 > 3		
Liu et al. (2011)	1. ANEN: 106 (19) 2. ANED: 109 (20) 3. ADEN: 146 (27) 4. ADED: 188 (34)	EF *n* variables: 2	1.25	0.82 to 1.68	AV *n* variables: 2	0.40	0.36 to 0.44	AV *n* variables: 2	1.85	1.41 to 2.29	EF *n* variables: 2	0.17	0.11 to 0.22	AV, EF *n* variables: 4	1.55	1.03 to 2.03	AV *n* variables: 2	1.37	0.98 to 1.76
	N variables: 4	2 < 3 EF *n* variables: 2	1.12	0.89 to 1.34	2 > 4 AV, EF *n* variables: 3	1.12	0.75 to 1.56	2 < 4 EF *n* variables: 1	0.63		3 > 4 EF *n* variables: 2	1.59	1.18 to 2.00	3 < 4 AV *n* variables: 4	0.19	0.15 to 0.23			
Seaton et al. (1999)	1. High: NS 2. Motor Deficit: NS 3. Low: NS 4. Impaired: NS	1 > 2 Arth, Intel, RaPS VC, WM *n* variables: 7	NA	NA	1 = 2 RaPS, WM *n* variables: 4	NA	NA	1 < 3 Arth, Intel, RaPS VC, WM *n* variables: 7	NA	NA	1 = 3 RaPS, WM *n* variables: 4	NA	NA	1 > 4 Arth, Intel, RaPS VC, WM *n* variables: 11	NA	NA	2 = 3 Arth, Intel, RaPS VC, WM *n* variables: 11	NA	NA
	N variables: 11	2 > 4 Arth, Intel, RaPS VC *n* variables: 6	NA	NA	2 = 4 RaPS, WM *n* variables: 5	NA	NA	3 > 4 Arth, Intel, RaPS VC, WM *n* variables: 10	NA	NA	3 = 4 RaPS *n* variables: 1	NA	NA						

Note: Only statistical pairwise comparisons with one or more variables included are presented; Arith, arithmetic; ADED, attention deficit, executive deficit; ADEN, attention deficit, executive non-deficit; ANED, attention non-deficit, executive deficit; ANEN, attention non-deficit, executive non-deficit; AV, attention-vigilance; cIQ, current IQ; EF, executive function; epIQ, estimated premorbid IQ; GF, global functioning; Fl, fluency; HC, healthy control; Intel, intellect; IQ, intelligence quotient; MC, motor control; Mem, memory; NS, not stated; RaPS, reasoning and problem solving; SC, social cognition; SoP, speed of processing; VC, verbal comprehension; VerL, verbal learning; VisL, visual learning; VS, visuospatial; WM, working memory

Table 4. Summary of group comparisons for research investigating putative cognitive symptom trajectories.

Study	Subgroups: n (% of N)	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}	Pairwise Comparison	d_{μ}	d_{range}
Ammari et al. (2014)	1. Preserved Patients: 32 (32) 2. Deteriorated Patients: 26 (26) 3. Premorbidly Impaired Patients: 13 (14) A. Preserved Controls: 20 (25) B. Deteriorated Controls: 9 (11) *N* variables: 10	1 > 2 AV, cIQ, GF, SoP, VisL *n* variables: 5	1.22	0.77 to 2.03	1 = 2 epIQ, RaPS, SC, VerL, WM *n* variables: 5	0.47	0.18 to 0.68	1 > 3 AV, cIQ, epIQ, GF, SoP, VerL, VisL, WM *n* variables: 8	1.96	1.23 to 3.55	1 = 3 RaPS, SC *n* variables: 10	0.96	0.87 to 1.04	2 > 3 epIQ *n* variables: 1	2.47		2 = 3 AV, cIQ, GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 9	0.65	0.04 to 0.92
		1 < A epIQ, GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	0.94	0.61 to 1.17	1 = A cIQ, AV *n* variables: 2	0.86	0.54 to 0.82	1 > B cIQ *n* variables: 1	2.30		1 = B AV, epIQ, GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 9	0.20	0.05 to 0.72	2 < A AV, cIQ, GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 9	1.98	1.28 to 2.83	2 = A epIQ *n* variables: 1	0.79	
		2 < B AV, GF *n* variables: 2	1.52	1.43 to 1.60	2 = B cIQ, epIQ, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 8	0.59	0.28 to 1.07	3 < A AV, cIQ, epIQ, GF, RaPS, SC, SoP, VerL, VisL, WM *n* variables: 10	2.67	1.86 to 4.00	3 < B AV, epIQ, GF, WM *n* variables: 4	1.46	0.85 to 2.02	3 = B cIQ, RaPS, SC, SoP, VerL, VisL *n* variables: 6	1.03	0.07 to 1.66			
Badcock et al. (2005)	1. Preserved: 45 (41) 2. Deteriorated: 47 (43) 3. Compromised: 17(16) A. HC: 149 (100) *N* variables: 6 A. only compared with 1.	1 > 2 AV, VerL *n* variables: 4	0.74	0.55 to 0.85	1 = 2 Fl, SoP *n* variables: 2	0.34	0.19 to 0.49	1 > 3 AV, Fl, VerL *n* variables: 3	0.77	0.73 to 0.80	1 = 3 AV, SoP, VerL *n* variables: 3	0.43	0.08 to 0.48	2 = 3 AV, Fl, SoP, VerL *n* variables: 6	0.17	0.01 to 0.34	1 < A AV, SoP, VerL *n* variables: 5	0.68	0.37 to 1.08
		1 = A Fl *n* variables: 1	0.33																
Elliott et al. (1998)	1. PIQ: 12 (100) A. HC: 12 (100) *N* variables: 10	1 < A EF, VisL, WM *n* variables: 4	NA	NA	1 = A EF, VisL, WM *n* variables: 6	NA	NA												
Evans et al. (1997)	1. PIQ: 16 (52) A. HC: 16 (100) *N* variables: 7	1 < A EF *n* variables: 4	0.96	0.91 to 1.02	1 = A EF *n* variables: 3	0.23	0.00 to 0.67												
Joyce et al. (2005) Note: FEP	1. PIQ: 47 (50) 2. DIQ: 37 (40) 3. Low IQ: 9 (10) A. HC: 50 (100) *N* variables: 8	1 > 2 cIQ, EF, WM *n* variables: 4	0.75	0.52 to 1.19	1 = 2 epIQ, EF, VisL *n* variables: 4	0.29	0.08 to 0.37	1 > 3 cIQ, epIQ *n* variables: 2	1.65	1.60 to 1.68	1 = 3 EF, VisL, WM *n* variables: 6	0.31	0.06 to 0.61	2 > 3 epIQ *n* variables: 1	2.30		2 = 3 cIQ, EF, VisL, WM *n* variables: 7	0.33	0.19 to 0.66
		1 = A cIQ, EF, epIQ, VisL, WM *n* variables: 8	0.26	0.10 to 0.35	2 < A cIQ, EF, VisL, WM *n* variables: 7	0.83	0.60 to 1.10	2 = A epIQ *n* variables: 1	0.17		3 < A cIQ, EF, epIQ *n* variables: 3	1.33	1.04 to 1.48	3 = A EF, VisL, WM *n* variables: 5	0.55	0.39 to 0.73			
Kravariti et al. (2009) Note: FEP	1. Stable Good: 22 (22) 2. Deteriorated Poor: 37 (37) 3. Stable Poor: 19 (19) A. HC Stable Good: 137 (43) B. HC Deteriorated Poor: 26 (8) C. HC Stable Poor: 14 (4)	1 < A EF, SoP, WM *n* variables: 3	0.42	0.35 to 0.53	1 = A cIQ, EF, epIQ, Fl, SoP, VC, VerL, VisL, VS *n* variables: 14	0.17	0.03 to 0.34	2 < B Fl, SoP, WM *n* variables: 3	0.79	0.70 to 0.87	2 = B cIQ, EF, epIQ, Fl, SoP, VC, VerL, VisL VS *n* variables: 14	0.33	0.00 to 0.61	3 < C Fl, SoP *n* variables: 4	1.22	0.78 to 1.53	3 = C cIQ, EF, epIQ, Fl, SoP, VC, VerL, VisL, VS, WM *n* variables: 13	0.45	0.06 to 1.07

	N variables: 17																			
Kremen et al. (2008)	1. Preserved: 22 (28) 2. Deteriorated: 40 (50) 3. Compromised: 18 (22) A. HC: 93 (100)	1 > 2 AV, cIQ, Fl, EF, SoP, VerL, VS *n* variables: 11	0.65	0.47 to 1.16	1 = 2 Arth, AV, EF, VerL, VS *n* variables: 10	0.33	0.20 to 0.48	1 > 3 AV, cIQ, EF, epIQ, Fl, SoP, VerL, VS *n* variables: 16	1.10	0.49 to 3.22	1 = 3 AV, EF, VS *n* variables: 5	0.45	0.27 to 0.78	2 > 3 Arth, cIQ, EF, epIQ, Fl, VerL, VS *n* variables: 7	1.17	0.51 to 3.58	2 = 3 AV, EF, Fl, SoP, VerL, VS *n* variables: 14	0.18	0.00 to 0.48
	N variables: 21	1 < A AV, EF, cIQ, SoP, VerL, VS *n* variables: 12	0.98	0.56 to 1.27	1 = A Arth, AV, epIQ, Fl, VerL, VS *n* variables: 9	0.24	0.07 to 0.41	2 < A Arth, AV, cIQ, EF, Fl, SoP, VerL, VS *n* variables: 18	1.21	0.40 to 1.74	2 = A EF, VS *n* variables: 2	0.40	0.30 to 0.49	3 < A Arth, AV, EF, cIQ, EF, epIQ, Fl, SoP, VerL, VS *n* variables: 21	1.63	0.75 to 2.33			
Leeson et al. (2011) Note: FEP	1. PIQ: 40 (32) 2. DIQ: 57 (44) 3. Low IQ: 32 (24) A. HC: 120 (100)	1 > 2 cIQ, EF, epIQ, VerL, WM *n* variables: 7	1.29	0.82 to 2.73	1 > 3 cIQ, EF, epIQ, VerL, WM *n* variables: 7	1.50	0.67 to 2.62	2 > 3 epIQ *n* variables: 1	1.22		2 = 3 cIQ, EF, VerL, WM *n* variables: 6	0.20	0.01 to 0.51	1 < A WM *n* variables: 1	0.61		1 = A cIQ, EF, epIQ, VerL, WM *n* variables: 6	0.29	0.06 to 0.56
	N variables: 7	2 < A cIQ, EF, epIQ, VerL, WM *n* variables: 7	1.49	0.93 to 1.83	3 < A cIQ, EF, epIQ, VerL, WM *n* variables: 7	1.56	0.88 to 2.38													
MacCabe et al. (2012)	1. High IQ-intact: 10 (20) 2. High IQ-decline: 24 (48) 3. Typical IQ: 16 (32) A. Above-average HC: 19 (100)	1 > 2 cIQ, EF, VerL, WM *n* variables: 5	1.55	1.06 to 2.13	1 = 2 EF, epIQ, Fl, SoP, VerL, WM *n* variables: 15	0.60	0.00 to 1.16	1 > 3 cIQ, EF, epIQ, Fl, SoP, VerL, WM *n* variables: 12	1.62	0.97 to 2.62	1 = 3 EF, Fl, SoP, WM *n* variables: 8	0.71	0.08 to 1.18	2 > 3 cIQ, epIQ, Fl *n* variables: 3	1.73	1.05 to 2.89	2 = 3 EF, Fl, SoP, VerL, WM *n* variables: 12	0.35	0.06 to 0.84
	N variables: 20	1 = A cIQ, EF, epIQ, Fl, SoP, VerL, WM *n* variables: 20	0.38	0.03 to 1.25	2 < A EF, Fl, SoP, VerL, WM *n* variables: 9	1.25	0.99 to 1.71	2 = A cIQ, EF, epIQ, Fl, WM *n* variables: 11	0.61	0.06 to 1.29	3 < A cIQ, EF, epIQ, Fl, SoP, VerL, WM *n* variables: 12	1.60	1.27 to 2.13	3 = A EF, Fl, WM *n* variables: 8	0.76	0.59 to 0.96			
Weickert et al. (2000)	1. PIQ: 29 (25) 2. DIQ: 60 (51) 3. CIQ: 28 (24) A. HC: 27 (100)	1 > 2 EF, SoP, VerL, VS *n* variables: 10	0.70	0.46 to 1.01	1 = 2 AV, Fl, Mem, SoP, VC *n* variables: 7	0.44	0.18 to 0.57	1 > 3 AV, EF, SoP, VerL, VC, VS *n* variables: 12	0.99	0.40 to 1.46	1 = 3 Fl, Mem, SoP *n* variables: 5	0.47	0.39 to 0.77	2 > 3 AV, EF, VC, VS *n* variables: 4	0.48	0.23 to 0.71	2 = 3 EF, Fl, Mem, SoP, VerL *n* variables: 13	0.11	0.01 to 0.29
	N variables: 23	1 < A AV, EF *n* variables: 2	0.76	0.67 to 0.88	1 = A Arth, cIQ, EF, epIQ, Fl, RaPS, SoP, Mem, VC, VerL, VS *n* variables: 21	0.37	0.05 to 1.15	2 < A Arth, AV, cIQ, EF, Fl, RaPS, SoP, Mem, VC, VerL, VS *n* variables: 16	0.98	0.13 to 1.42	2 = A epIQ, Fl, SoP, Mem, VC *n* variables: 7	0.60	0.13 to 1.07	3 < A Arth, AV, cIQ, EF, epIQ, Fl, RaPS, SoP, Mem, VC, VerL, VS *n* variables: 18	1.37	0.13 to 1.42	3 = A Fl, Mem, SoP *n* variables: 5	0.82	0.24 to 1.31
Potter and Nestor (2010)[&]	1. Preserved: 21 (29) 2. Deteriorated: 21 (29) 3. Compromised: 31 (42) A. HC 74 (100)	1 > 2 cIQ, Intel, Mem, RaPS, SoP, VC, VerL, VisL *n* variables: 12	1.33	0.68 to 2.21	1 = 2 EF, epIQ, VerL, VisL, WM *n* variables: 8	0.39	0.17 to 0.65	1 > 3 cIQ, EF, Intel, Mem, RaPS, SoP, VC, VerL, VisL, WM *n* variables: 18	1.41	0.55 to 2.94	1 = 3 EF, SoP *n* variables: 2	0.39	0.37 to 0.40	2 > 3 cIQ, EF, epIQ, Intel, VerL, VisL, VC, WM *n* variables: 9	1.08	0.48 to 3.55	2 = 3 EF, Int Mem, RaPS, SoP, VerL, VisL *n* variables: 11	0.25	0.08 to 0.51
	N variables: 20	1 < A Mem, SoP, VerL, VisL *n* variables: 6	0.75	0.51 to 0.91	1 = A cIQ, EF, epIQ, pIQ, RaPS, VC, VerL vIQ, VisL, WM *n* variables: 14	0.31	0.03 to 0.86	2 < A cIQ, EF, Mem, pIQ, RaPS, SoP, VC, VerL vIQ, VisL, WM *n* variables: 17	1.44	0.71 to 1.99	2 = A EF, epIQ *n* variables: 3	0.66	0.45 to 1.04	3 < A cIQ, EF, epIQ, Mem, pIQ, RaPS, SoP, VC, VerL vIQ, VisL, WM *n* variables: 19	1.66	0.58 to 2.36	3 = A EF *n* variables: 1	1.17	
		1 > 2			1 = 2			1 > 3			1 = 3			2 > 3			2 = 3		

Weinberg et al. (2016)[&]	1. Putatively Preserved: 25 (26) 2. Moderately Deteriorated: 33 (34) 3. Severely Deteriorated: 27 (28) A. HC 87 (100) *N* variables: 7	cIQ, epIQ, Mem, WM *n* variables: 5	1.61	0.81 to 3.18	Fl, SoP *n* variables: 2	0.29	0.11 to 0.47	cIQ, epIQ, Fl, Mem, WM *n* variables: 6	2.28	1.16 to 5.51	SoP *n* variables: 1	0.56		cIQ, Fl *n* variables: 2	2.18	1.01 to 3.35	epIQ, Mem, SoP, WM *n* variables: 5	0.42	0.11 to 0.79
		1 < A SoP *n* variables: 1	1.19		1 = A cIQ, epIQ, Fl, Mem, WM *n* variables: 6	0.29	0.06 to 0.53	2 < A cIQ Mem, SoP, WM *n* variables: 5	1.29	1.10 to 1.60	2 = A epIQ, Fl *n* variables: 2	0.53	0.51 to 0.54	3 < A cIQ, epIQ, Fl, Mem, SoP WM *n* variables: 7	1.61	1.05 to 2.31			
		1 > 2 AV, epIQ, Fl, Lang, VerL, VS, WM *n* variables: 8	0.76	0.30 to 1.40	1 > 3 AV, epIQ, Fl, Lang, VerL, VS, WM *n* variables: 8	1.43	0.60 to 2.20	2 > 3 AV, epIQ, Fl, Lang, VerL, VS, WM *n* variables: 8	0.70	0.30 to 2.20	1 < A Fl, Lang, VerL *n* variables: 3	0.43	0.20 to 0.60	1 = A AV, epIQ, VerL, VS, WM *n* variables: 5	0.26	0.08 to 0.40	2 < A AV, epIQ, Fl, Lang, VerL, VS, WM *n* variables: 8	1.01	0.20 to 1.60
Wells et al. (2015)[&]	1. Preserved: 157 (29) 2. Deteriorated: 239 (45) 3. Compromised: 138 (26) A. HC: 635 (100) *N* variables: 8																	
		3 < A AV, epIQ, Fl, Lang, VerL, VS, WM *n* variables: 8	1.63	1.00 to 2.2														

Note: Only statistical pairwise comparisons with one or more variables included are presented; [&] Data-driven methodology employed; Arith, arithmetic; AV, attention-vigilance; cIQ, current IQ; CIQ, compromised IQ; DIQ, deteriorated IQ; EF, executive function; epIQ, estimated premorbid IQ; FEP, first episode psychosis; GF, global functioning; Fl, fluency; HC, healthy control; Intel, intellect; IQ, intelligence quotient; Lang, language; Mem, memory; PIQ, preserved IQ; RaPS, reasoning and problem solving; SC, social cognition; SoP, speed of processing; VC, verbal comprehension; VerL, verbal learning; VisL, visual learning; VS, visuospatial; WM, working memory

Table 5. Recommended guidelines for reporting on cluster analysis

Guideline:	Explanation:
1. Statistical Package	Report the statistical software package used (e.g. "Cluster analysis was performed using SPSS v.25.0").
2. Data standardization	Describe the type of data standardization that was performed on the variables being entered into the clustering analysis (e.g. z-scores based on healthy control performance); if none was performed, state that the raw test data was entered and why.
3. Similarity/Distance measure and clustering algorithm	Detail the combination of distance measure (e.g. squared Euclidian distance) and clustering algorithm (e.g. Ward's Method) employed to attain the final cluster solution reported on. If multiple methodological combinations were explored, detail the process. Providing the results of the unsuccessful methodological combinations as supplementary material is also recommended. If a *k*-means clustering was employed, clearly detail this.
4a. Basis for determining the number of clusters	Detail the process through which the number of clusters to retain was determined (e.g. visual inspection of dendrogram, meaningful jump in agglomeration schedule coefficient). If a *k*-means clustering was employed, clearly describe the basis from which the number of clusters was defined *a priori* (e.g. "based on the classifications of Weickert et al. (2000), a three-cluster solution was defined").
4b. Presentation of pertinent plots	If visually based measures were used to determine the number of clusters to retain (e.g. dendrogram, inverse scree plot of agglomeration schedule coefficients) or to compare potential cluster solutions (e.g. Canonical Discriminant Function Plots) present them in the supplementary material.
5a. Validation	Describe the steps taken to validate the final cluster solution. Provide evidence, either in-text or in the supplementary material, of final cluster solution validation.
5b. Validation Recommendations	The gold-standard is external replication with an independent dataset and research team, however there are ways to provide a degree of validation of the final cluster solution using the same dataset. If independent replication is not possible, it is recommended that the following internal validation procedure be followed: Assess the degree of classification concordance between the final clustering solution and the solutions identified using varying random split-samples and alternate methodologies. For example; 1) Random 75% of original sample, same analysis pipeline; 2) Random 50% of original sample, same analysis pipeline; 3) Original sample, alternate analysis pipeline (e.g. average-linkage clustering in lieu of Ward's method) With a high degree of classification concordance required to validate the final cluster solution obtained (e.g., using Cohen's Kappa or adjusted rand index)
6. Subgroup Characterisation	Provide statistical characterisation of the emergent subgroups using appropriate *post-hoc* comparisons (e.g. Games-Howell corrected *post-hoc* pairwise contrasts) to establish cognitive profiles. Refrain from detailing subjective interpretations of cognitive profiles without providing empirical support. Statistically compare subgroups to each other (using both variables entered and not entered into the subgrouping analysis) and to a healthy control comparison group. If possible, investigate the healthy control comparison group for meaningful cognitive heterogeneity and compare cognitive performance across matched subgroups.
7. Presentation of Results	Provide a summary of subgroup neuropsychological performance (e.g. mean, s.d., confidence intervals) and any subsequent *post-hoc* results in a table. Refrain from only including a figure or not including either. If it is not possible to provide a detailed summary of subgroup test performance in-text, include in the supplementary material.

Note: Based on the recommendations found in Everitt, Landau, Leese, and Stahl (2011) and Balijepally et al. (2011).