**Supplementary Materials**

**S1 -Tasks assigned to cognitive domains**

In addition to the 7 categories derived from the National Institute of Mental Health–Measurement and Treatment Research to Improve Cognition in Schizophrenia Neurocognition Committee (MATRICS) (Kern et al., 2007; Nuechterlein et al., 2004) (i.e. speed of processing, attention and vigilance, working memory, verbal learning and memory, visual learning and memory, reasoning and problem solving, and social cognition), we also included “construction and visuospatial skills”, “motor skills” and “verbal and language skills”.(Heinrichs & Zakzanis, 1998; Lezak et al., 2004; Mesholam-Gately et al., 2009). There is evidence that verbal fluency should also be examined as an independent cognitive domain (Dickinson et al., 2007; Fett et al., 2011; Green et al., 2000) and was therefore categorised separately.The WAIS task ‘Block design’ is thought to measure construction and visuospatial skills (Heinrichs & Zakzanis, 1998; Lezak et al., 2004; Mesholam-Gately et al., 2009) and was therefore categorised under this heading, as opposed to ‘reasoning and problem-solving skills’.

**Table S1 – Tasks assigned to cognitive domains**

|  |  |
| --- | --- |
| Domain | Measures |
| Speed of processing | Digit symbol substitution test (DSST), Trail Making Task A (TMTa), Stroop Color, Stroop Words, Color Trails Test-1 (CTT-1) |
| Reasoning and Problem Solving | Wisconsin Card Sorting Task (WCST) categories completed, WCST-perseverative errors, WCST failure to maintain set, WCST trials to first, WCST conceptual level responses, WCST number of perseverative responses, WCST percentage of perseverative errors, WCST total errors, Stockings of Cambridge (SoC), Stroop interference (time/number correct), Stroop interference errors, Trail Making Task B (TMTb), Color Trails Test-2 (CTT-2), DKEFs Color Word Test, DKEFS-category switching , Neuropsychology Assessment Battery (NAB) executive function score., NAB-mazes |
| Attention and Vigilance | Continuous Performance Task (CPT), CPT-Identical Pairs (CPT-IP), Brief Test of Attention (BTA)-Letters, BTA-Numbers, CPT-Omission errors, CPT-Commission errors, Letter cancellation task, Span of Apprehension (SPAN), Backward Masking Test |
| Working Memory | Digit Span backwards, Digit span forwards, Digit span total, Letter Number Sequencing number (LNS) recalled, Paced Auditory Serial Addition Test (PASAT), CANTAB Spatial Working Memory, Wechsler Memory Scale (WMS) Spatial Span |
| Verbal Learning and Memory | California Verbal Learning Test (CVLT) immediate, CVLT Delayed, CVLT errors, CVLT recognition, Hopkins Verbal Learning Test-Revised (HVLT-R) immediate, Logical memory I (LMI(, Logical memory II (LMII), Logical memory II - delayed (LMII-D), Logical memory total, , Rey Auditory Verbal Learning (RAVLT) immediate, RAVLT delayed, Paired Associated Learning (PAL) Easy, PAL Hard, Spain Complutense Verbal Learning Test (TAVEC) |
| Visual Learning and Memory | Brief Visuospatial Memory Test-Revised (BVMT), REY Osterreith Complex Figure (RCF)- delayed, WMS visual reproduction (immediate, delayed), visual memory – faces/family pictures, WMS-R-HK visual reproduction, |
| Social Cognition | Mayer-Salvey-Caruso Emotional Intelligence test (MSCEIT) managing emotions, , MESCEIT total Facial Emotion Discrimination Test (FEDT), Facial Emotion Identification Test (FEIT), Hinting task, MATRICS Social cognition, relationships Across Domains (RAD), The Awareness of Social Inferences Test (TASIT) |
| Verbal Fluency | Controlled Oral Word Association test (COWAT), verbal fluency total, verbal fluency letters, verbal fluency categories, verbal fluency animals, Thurstone Word Fluency Test, |
| Verbal and Language Skills | Wechsler Adult Intelligence Scale (WAIS) - vocabulary |
| Construction and Visuospatial Skills | RCF-copy, WAIS Block Design |
| Motor Skills | Grooved pegboard (dominant, non-dominant), finger tapping test (FTT) -left/right |

**S2: Table of Outliers**

|  |  |
| --- | --- |
| Table S2: Table of Outliers | |
| Task/Domain | **Outlier** |
| Verbal learning | Torgalbasean et al., (2015),  Gonzalez-Ortega et al., (2013) |
| TMTa | Gonzalez et al., (2013) |
| WCST Perseverative | Gonzalez et al., (2013) |
| LNS Total | Pena et al., (2011) |
| LMI | Gonzalez et al., (2013) |
| VF Letters | Zhou et al., (2017) |
| Stroop Interference | Nopoulos et al., (1994) |

**Table S3. Characteristics of included studies**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | |
| Author | | **Sample size Baseline** | **Sample size Follow-up** | **Sample Characteristics** | **Cognitive measures** | **Sensitivity Analyses Inclusion** | **Domain/Task Analyses** |
| Addington 2005 [1] | | 247 FEP  66 HCs | 154 FEP  66 HCs | Mean age: 24.5  Male: 67% at baseline (no sig difference at FU)  Follow-up period: 12 months  98% taking SGA at baseline,  82% at follow-up  Schizophrenia 62.8%  Schizophreniform 21.9%  Delusional disorder 1.6%  Brief psychotic disorder 2.8%  Psychotic disorder NOS 9.3%  Schizoaffective disorder 1.6%  Patients performed worse than HCs on all tasks at baseline | RAVLT, COWAT, WMS, RAVLT, REY copy, REY figure recall LNS, WCST, CPT, SPAN, TMTa and TMTb, Grooved Pegboard, Stroop Color | HCs: Yes  Non-affective: Yes | Global  Speed of processing  Reasoning and problem solving  Attention and vigilance  Verbal learning and memory  Verbal fluency  Construction and visuospatial skills  Motor skill |
| Addington 2006, [2]a | | 50 FEP  55 HCs | 50 FEP  55 HCs | Mean age: 25.1  Male: 60%  Follow-up period: 12 months  82% taking SGA at baseline, 87.5% at FU  One year diagnoses were as follows:  schizophrenia 62.0%, schizophreniform 24.0%,  delusional disorder 2.0%, brief psychotic disorder  2.0%, psychotic disorder NOS 6.0%, and schizoaffective  disorder 2%.  Patients performed worse than HCs on all tasks at baseline | FEIT, FEDT | HCs: Yes  Non-affective: Yes | Social cognition |
| Albus 2002 [3] | | 50 FES  50 HCs | 50 FES  50 HCs | Mean age: 29  Male: 46%  Follow-up period: 24 months  23/50 FGA, 18/50 Clozapine  Diagnoses Non-affective (first-episode schizophrenia)  Patients performed worse than HCs on all tasks at baseline | WMS (LMI, LMII, Visual memory immediate, visual memory delayed, PAL), CVLT, TMTb, DSST, WCST categories, WCST perseverative responses | HCs: No  Non-affective: Yes | Global  Speed of processing  Reasoning and problem solving  Verbal learning and memory  Visual learning and memory |
| Ayres 2010 [4] | | 30 recent-onset schizophrenia  67 HCs | 30 recent-onset schizophrenia  67 HCs (no longitudinal data) | Mean age: 28.82 (for 34)  Male: 76.5% (for 34)  Follow-up period: 22 months  Diagnoses: Schizophrenia FEP for my data  No HC data in paper  Patients performed worse than HCs on all tasks at baseline – not significant for attention, visual memory or executive function | Digit forward, digit backwards, Verbal fluency | HCs: No  Non-affective: Yes | Global  Working memory  Verbal fluency |
| Chang 2014 [5] | | 93 FES | 93 FES | Mean age: 31.2  Male: 45.2%  Follow-up period: 12 months  48 were first assessed in a medication-naive state, and the rest were evaluated within 7 days of starting first-generation antipsychotic treatment  Diagnoses: first-episode schizophrenia, schizophreniform disorder or schizoaffective disorder | WMS-R, Digit span forward , WCST, verbal fluency, WMS-R-HK visual reproduction | HCs: No  Non-affective: Yes | Global  Reasoning and problem solving Working memory  Verbal learning and memory  Visual learning and memory  Verbal fluency |
| Gold 1999 [6] | | 59 recent-onset schizophrenia spectrum | 59 recent-onset schizophrenia spectrum | Mean age: 24  Male: 76%  Follow-up period: 60 months  40/54 APs 29/54 FGA  Diagnoses: First-episode and recent-onset schizophrenia spectrum | IQ, TMTb, Letter cancellation task, FTT, COWAT, LMI, LMII, WCST, RCF-delay | HCs: No  Non-affective: Yes | Visual learning and memory |
| Gonzalez-Ortega 2013 [7] | | 26 FEP | 26 FEP | Mean age: 25.77  Male: 57.7%  Follow-up period: 54 months  16 on atypical APs  9 not on APs  Diagnoses: schizophreniform disorder; schizoaffective disorder; schizophrenia; delusional disorder; brief psychotic disorder; atypical psychosis; bipolar I/II disorder; or major depressive disorder | WAIS-Vocabulary, WCST errors, total, WCST perseverative errors, WCST conceptual level responses, Stroop interference, TMTb, Verbal fluency -categories, LMI, LMII, PALI, PALII, Digit span backwards, LNS, TMTa, DSST | HCs: No  Non-affective: No | Global  Speed of processing  Reasoning and problem solving  Working memory  Verbal learning and memory  Verbal fluency  Verbal and language skills |
| Haatveit 2015 [8] | | 82 FEP  107 HC | 82 FEP  107 HC | Mean age: 26.7  Male: 60.98%  Follow-up period: 12 months  78.1% on Antipsychotics  Diagnosis of schizophrenia (42.7%), schizophreniform disorder (11.0%), schizoaffective disorder (4.9%), major depression with mood incongruent psychotic symptoms (11.0%), or other psychosis (30.5%).  Patients performed worse than HCs on all tasks at baseline | Digit span, LNS, Letter-Fluency, Category Fluency, WCST, DKEFS Color-word interference, DKEFS category switching | HCs: Yes  Non-affective: Yes | Global  Reasoning and problem solving  Working memory  Verbal fluency |
| Higgins 2021 [9] | | 54 early psychosis  42 HCs (baseline only) | 30 early psychosis (slight variations by tasks) | Mean age: 23.07  Male: 61.11%  Follow-up period: 12 months  CPZ equivalents by group (schizophrenia = 205.94; BD = 153.38)  Diagnoses: schizophrenia-spectrum diagnoses (44%) and BD with psychotic features (66%)  No HC follow-up  No baseline differences in MSCEIT or TASIT | MCCB composite, TASIT, MCCB Social Sub score | HCs: No  Non-affective: No | Global  Social cognition |
| Horan 2012 [10] | | 55 FES | 55 FES | Mean age: 22.3  Male: 76.79%  Follow-up period: 12 months  All participants were clinically stabilized on oral risperidone and received clinically determined antipsychotic medications and dosages during the follow-up period.  Diagnoses: First-episode schizophrenia | MSCEIT, TASIT, RAD, | HCs: No  Non-affective: Yes | Social Cognition |
| Kenney 2015 [11] | | 37 FEP  59 HCs | 23 FEP  21 HCs | Mean age: 28.3  Male: 65%  Follow-up period: 48 months  11 on AP at follow-up  Diagnoses: Schizophrenia (n = 6), schizophreniform disorder (n = 1), schizoaffective disorder (n=3), psychotic disorder not otherwise specified (4) and delusional disorder (n=1) were defined as non-affective types of psychoses whereas bipolar I disorder (n = 6) and major depressive  disorder, recurrent (n = 2) defined as affective types of psychoses  Patients performed worse than HCs on all tasks at baseline | TMTa, BACS, Verbal fluency – categories, Verbal fluency – letters, CPT-IP, WMS-Spatial span, LNS, HVLT-R, BVMT, NAB-mazes, MSCEIT-managing emotions | HCs: Yes  Non-affective: No | Global  Speed of processing  Reasoning and problem solving  Attention and vigilance  Visual learning and memory  Verbal learning and memory  Social cognition |
| Leeson 2009 [12] | | 54 FEP | 54 FEP | Diagnoses: First-episode schizophrenia or schizoaffective disorder | IQ, CANTAB SWM | HCs: No  Non-affective: Yes | Global  Working memory |
| Leeson 2010 [13]b | | 29 FEP  29 HCs | 29 FEP  29 HCs | Mean age: 26.77  Male: 64% (for entire sample, not just those with follow-up data)  Follow-up period: 12 months  “Most patients” being prescribed antipsychotic medication  Diagnoses: First-episode schizophrenia or schizoaffective disorder  IQ-matched - so no difference between groups on IQ | Information, Arithmetic, Block Design, Digit Symbol | HCs: Yes  Non-affective: Yes | Visuospatial  Speed of processing  Construction and visuospatial skills |
| Lindgren 2020 [14] | | 52 FEP  62 HCs | 32 FEP  44 HCs | Mean age: 26.7  Male: 59.6% Follow-up period: 12 months  94.2% on antipsychotic at baseline  71.9% on antipsychotic at FU  Diagnoses: Non-affective  Patients performed worse than HCs on all tasks at baseline | WAIS-III (Block design, Vocabulary, DSST) WMS-III (LMI, LMII, LMII-delayed), TMTa, TMTb, FTT, CPT-IP d-prime, Hinting Task, Spatial span, | HCs: Yes  Non-affective: Yes | Global  Speed of processing  Reasoning and problem solving  Attention and vigilance  Working memory  Verbal learning and memory  Visual learning and memory  Verbal fluency  Social cognition  Construction and visuospatial skills  Verbal and language skills |
| Liu 2015 [15] | | 49 FEP  137 HCs | 27 FEP  117 HCs | Mean age: 22.34  Male: 43.86%  Follow-up period: 12 months  Almost all subjects with FEP were under treatment with second generation antipsychotics  Diagnoses: Participants with schizophrenia, schizophreniform disorder, brief psychotic disorder, or schizoaffective disorder meeting the DSM-IV criteria in the preceding one year.  Patients performed worse than HCs on all tasks at baseline | CPT, WCST. WAIS-III, WMS-III, Verbal fluency. | HCs: Yes  Non-affective: Yes | Global  Reasoning and problem solving  Working memory  Visual learning and memory  Verbal fluency |
| Lho 2020 [16] | | 25 FEP  25 HCs | 25 FEP  25 HCs (no longitudinal data) | Mean age: 23.1  Male: 40%  Follow-up period: 12 months  Diagnoses: Schizophrenia, schizoaffective disorder, or schizophreniform disorder,  Patients performed worse on TMTa, but not IQ or TMTb at baseline | TMTa, TMTb | HCs: No  Non-affective: Yes | Global  Speed of processing Reasoning and Problem Solving |
| Noupolous 1994c [17] | | 35 recent-onset or FEP  68 HCs | 35 Recent-onset or FEP  68 HCs (no data) | Mean age: 23.71  Male: 83%  Follow-up period: 12 or 24 months  29/35 on AP, 24/35 FGA, baseline all on FGA  Diagnoses: DSM-III criteria for schizophrenia or schizophreniform disorder | FTT, LMI, LMII, RAVLY total, RAVLY delayed, PAL-easy, PAL-hard, Stroop words, Stroop colors, Stroop interference, CPT – omission errors, CPT commission errors,  COWAT total | HCs: No  Non-affective: Yes | Global  Reasoning and problem solving  Attention and vigilance  Verbal learning and memory  Verbal fluency  Motor skill |
| Pena 2011 [18] | | 71 FEP  34 HCs | 71 FEP  34 HCs | Mean age: 28.5  Male: 68.8%  Follow-up period: 24 months  Diagnoses: Includes schizophreniform and affective psychosis  Treated with SGA  Diagnoses: At baseline: schizophrenia (5.8%), schizophreniform (30.2%), brief psychotic disorder (32.6%), bipolar disorder (22.1%), delusional disorders (4.7%), drug-induced psychosis (2.3%), major depressive disorder with psychotic features (2.3%)  . | BTA-N, BTA-L Verbal fluency categories, VF letters, Verbal fluency -categories. digit span backwards, LNS, DSST, Stroop colour, TMTa, WCST categories completed, WCST perseverative errors, LMII | HCs: Yes  Non-affective: No | Global  Speed of processing  Reasoning and problem solving  Attention and vigilance  Working memory  Verbal learning and memory |
| Rund 2007 [19] | | 111 FEP | 111 FEP | Mean age: 28.2  Male: 53%  Follow-up period: 12 months  78 % AP 73 % SGA  Diagnoses: At the 2-year follow-up, schizophrenia (52.25%), schizophreniform disorder (5.55%), schizoaffective disorder (10.81%), affective psychosis with mood incongruent delusions (18.92%), delusional disorder (5.41%), brief psychotic episode (2.7%), psychotic disorder not otherwise specified (5.41). | IQ (BD, Similarities, Digit span), FTT, WCST, CVLT, Digit span, CPT, Backward masking test | HCs: No  Non-affective: No | Global  Speed of processing  Reasoning and problem solving  Attention and vigilance  Working memory  Verbal learning and memory  Verbal fluency  Motor skill |
| Sanchez-Torres 2018 [20] | | 159 FEP  151 HCs | 159 FEP  151 HCs | Mean age: 25.83  Male: 65.4%  Follow-up period: 23.64 months  Cpz equivalent dose = 451 at baseline, 188.71 at follow-up  Diagnoses: Patients were grouped into three diagnostic categories: (1) schizophrenia spectrum disorders, which included schizophrenia, schizophreniform and schizoaffective disorders; (2) affective psychosis spectrum, including bipolar I and II disorders, and manic and depressive episodes with psychotic symptoms; and (3) other psychoses, including brief psychotic disorders, psychoses not otherwise specified  Patients performed worse than HCs on all tasks at baseline | WAIS III Vocab, CPT, TMT-a, TMTb, WCST, Stroop interference, COWAT, VF -animals, Digit span, LNS, CVLT, MSCEIT, TAVEC | HCs: Yes  Non-affective: No | Global  Speed of processing  Reasoning and problem solving  Social Cognition  Attention and vigilance  Working memory  Verbal learning and memory  Verbal fluency  Social cognition |
| Torgalsbøen 2015 [21] | | 28 FES  28 HCs | 28 FES  28 HCs | Mean age: 21  Male: 60.7%  Follow-up period: 24 months  80% on SGA  20% on FGA  Diagnoses: Non-affective only  Patients performed worse than HCs on all tasks at baseline  except for WM | TMTa, symbol coding, HVLT-R, WMS Spatial Span, LNS, BMVT-R, NAB, Category fluency, MSCEIT, CPT | HCs: Yes  Non-affective: Yes | Global  Reasoning and problem solving  Attention and vigilance  Working memory  Verbal learning and memory  Visual learning and memory  Social cognition |
| Townsend 2002 [22] | | 83 FEP | 83 FEP (but slight differences by task) | Mean age: 24.9  Male: 75%  Follow-up period: 12 months  40 Risperidone, 17 Olanzapine, 11 Quetiapine, 8 Clozapine, 7 typical antipsychotics  Diagnoses: schizophrenia, schizoaffective disorder, or schizophreniform psychosis | IQ, DSST, WAIS-Vocabulary, similarities, information, picture completion, block design) WMS-III (LMI, LMII, PAL, immediate memory – faces, family pictures, LNS, spatial span), WCST categories completes/perseverative errors, Stroop colour, Stroop Word, Stroop Interference, CPT d-prime, TMTa, TMTb, PASAT, Thurstone word fluency | HCs: No  Non-affective: Yes | Global  Speed of processing  Reasoning and problem solving  Working memory  Verbal fluency |
| Wright 2019 [23] | | 26 FEP | 26 FEP | Mean age: 25.9  Male: 73.08%  Follow-up period: 36 months  22 on medication at baseline  18 at follow-up  Diagnoses: First Episode Psychosis | WAIS - Vocabulary | HCs: No  Non-affective: Yes | Verbal and language skills |
| Zabala 2008 [24] | | 57 FES | 57 FES | Mean age: 24.46  Male: 70%  Follow-up period: 12 months  All patients on AP treatment  Diagnoses: First-episode schizophrenia | Stroop interference, WCST categories completed, WCST perseverative errors | HCs: No  Non-affective: Yes | Reasoning and Problem Solving |
| Zhou 2017 [25] | | 32 FES  17 HCs | 32 FEP  17 HCs | Mean age: 26.2  Male: 59.4%  Follow-up period: 12 months  At baseline:  10 AP naïve  3 FGA  19 SGA  At FU: All on SGA  Diagnoses: First episode of schizophrenia  Patients performed worse than HCs on all tasks at baseline | HVLT-R, CTT-1, CTT-2, Verbal Fluency categories, verbal fluency letters, Stroop Interference | HCs: Yes  Non-affective: Yes | Global  Reasoning and problem solving  Speed of processing  Attention and vigilance  Verbal learning and memory  Verbal fluency |
| HC = healthy controls, FES = first-episode schizophrenia, FEP = first-episode psychosis, CPZ = chlorpromazine, APs = antipsychotics, FGA = first-generation antipsychotics, SGA = second-generation antipsychotics  a overlaps with Addington et al., 2005  b overlaps with Leeson et al., 2009  c overlaps with Gold et al., 1999  See Supplementary Table S1 for cognitive task abbreviations | | | | | | | |

**References of studies included in meta-analyses**

1. Addington J, Saeedi H, Addington D. The course of cognitive functioning in first episode psychosis: Changes over time and impact on outcome. Schizophrenia Research. 2005;78(1):35–43.

2. Addington J, Saeedi H, Addington D. Facial affect recognition: A mediator between cognitive and social functioning in psychosis? Schizophrenia Research. 2006;85(1–3):142–50.

3. Albus M, Hubmann W, Mohr F, Tiedemann T v, Pechler S, Drießlein D, et al. Neurocognitive functioning in patients with first-episode schizophrenia: results of a prospective 15-year follow-up study. European Archives of Psychiatry and Clinical Neuroscience. 2020 Sep;270(6):689–98.

4. de Mello Ayres A, Scazufca M, Menezes PR, Nakano EY, Regina ACB, Schaufelberger MS, et al. Cognitive functioning in subjects with recent-onset psychosis from a low-middle-income environment: Multiple-domain deficits and longitudinal evaluation. Psychiatry Research. 2010;179(2):157–64.

5. Chang WC, Hui CLM, Chan SKW, Lee EHM, Wong GHY, Chen EYH. Relationship between diminished expression and cognitive impairment in first-episode schizophrenia: A prospective three-year follow-up study. Schizophrenia Research. 2014;152(1):146–51.

6. Gold S, Arndt S, Nopoulos P, O’Leary DS, Andreasen NC. Longitudinal study of cognitive function in first-episode and recent-onset schizophrenia. Am J Psychiatry. 1999 Sep;156(9):1342–8.

7. González-Ortega I, de los Mozos V, Echeburúa E, Mezo M, Besga A, Ruiz de Azúa S, et al. Working memory as a predictor of negative symptoms and functional outcome in first episode psychosis. Psychiatry Research. 2013;206(1):8–16.

8. Haatveit B, Vaskinn A, Sundet KS, Jensen J, Andreassen OA, Melle I, et al. Stability of executive functions in first episode psychosis: One year follow up study. Psychiatry Research. 2015;228(3):475–81.

9. Higgins A, Lewandowski KE, Liukasemsarn S, Hall MH. Longitudinal relationships between mismatch negativity, cognitive performance, and real-world functioning in early psychosis. Schizophrenia Research. 2021;228:385–93.

10. Horan WP, Green MF, Degroot M, Fiske A, Hellemann G, Kee K, et al. Social cognition in schizophrenia, part 2: 12-month stability and prediction of functional outcome in first-episode patients. Schizophrenia Bulletin. 2012;38(4):865–72.

11. Kenney J, Anderson-Schmidt H, Scanlon C, Arndt S, Scherz E, McInerney S, et al. Cognitive course in first-episode psychosis and clinical correlates: A 4 year longitudinal study using the MATRICS Consensus Cognitive Battery. Schizophr Res. 2015 Dec;169(1–3):101–8.

12. Leeson VC, Barnes TRE, Hutton SB, Ron MA, Joyce EM. IQ as a predictor of functional outcome in schizophrenia: A longitudinal, four-year study of first-episode psychosis. Schizophrenia Research. 2009;107(1):55–60.

13. Leeson VC, Barnes TRE, Harrison M, Matheson E, Harrison I, Mutsatsa SH, et al. The relationship between IQ, memory, executive function, and processing speed in recent-onset psychosis: 1-year stability and clinical outcome. Schizophrenia Bulletin. 2010;36(2):400–9.

14. Lindgren M, Birling H, Kieseppä T, Tuulio-Henriksson A. Is cognitive performance associated with anxiety and depression in first-episode psychosis? Journal of Affective Disorders. 2020;263:221–7.

15. Liu CC, Hua MS, Hwang TJ, Chiu CY, Liu CM, Hsieh MH, et al. Neurocognitive functioning of subjects with putative pre-psychotic states and early psychosis. Schizophrenia Research. 2015;164(1–3):40–6.

16. Lho SK, Kim M, Park J, Hwang WJ, Moon SY, Oh S, et al. Progressive Impairment of Mismatch Negativity Is Reflective of Underlying Pathophysiological Changes in Patients With First-Episode Psychosis. Frontiers in Psychiatry [Internet]. 2020;11. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087290818&doi=10.3389%2ffpsyt.2020.00587&partnerID=40&md5=420c963ac8951590b1c958449040fa4e

17. Nopoulos P, Flashman L, Flaum M, Arndt S, Andreasen N. Stability of cognitive functioning early in the course of schizophrenia. Schizophrenia Research. 1994;14(1):29–37.

18. Peña J, Ojeda N, Segarra R, Eguiluz JI, García J, Gutiérrez M. Executive functioning correctly classified diagnoses in patients with first-episode psychosis: Evidence from a 2-year longitudinal study. Schizophrenia Research. 2011;126(1–3):77–80.

19. Rund BR, Melle I, Friis S, Johannessen JO, Larsen TK, Midbøe LJ, et al. The course of neurocognitive functioning in first-episode psychosis and its relation to premorbid adjustment, duration of untreated psychosis, and relapse. Schizophrenia Research. 2007;91(1–3):132–40.

20. Sánchez-Torres AM, Moreno-Izco L, Lorente-Omeñaca R, Cabrera B, Lobo A, González-Pinto AM, et al. Individual trajectories of cognitive performance in first episode psychosis: a 2-year follow-up study. European Archives of Psychiatry and Clinical Neuroscience. 2018 Oct;268(7):699–711.

21. Torgalsbøen AK, Mohn C, Czajkowski N, Rund BR. Relationship between neurocognition and functional recovery in first-episode schizophrenia: Results from the second year of the Oslo multi-follow-up study. Psychiatry Research. 2015;227(2–3):185–91.

22. Townsend LA, Norman RMG, Malla AK, Rychlo AD, Ahmed RR. Changes in cognitive functioning following comprehensive treatment for first episode patients with schizophrenia spectrum disorders. Psychiatry Research. 2002;113(1–2):69–81.

23. Wright AC, Davies G, Fowler D, Greenwood K. Three-year follow-up study exploring metacognition and function in individuals with first episode psychosis. Frontiers in Psychiatry [Internet]. 2019;10(APR). Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067939293&doi=10.3389%2ffpsyt.2019.00182&partnerID=40&md5=e7513c1af204546353ba15b3ddc55de7

24. Zabala A, Eguiluz JI, Segarra R, Enjuto S, Ezcurra J, Elizagarate E, et al. Course of cognitive functioning in first-episode psychosis: A comparison between schizophrenia and non schizophrenia psychosis. Revista de Psiquiatria y Salud Mental. 2008;1(1):18–25.

25. Fu-Chun Z, Chuan-Yue W, Ungvari GS, Ng CH, Zhou Y, Zhang L, et al. Longitudinal changes in prospective memory and their clinical correlates at 1-year follow-up in first-episode schizophrenia. PLoS One [Internet]. 2017 Feb;12(2). Available from: https://search.proquest.com/scholarly-journals/longitudinal-changes-prospective-memory-their/docview/1872789940/se-2?accountid=11862

|  |  |  |
| --- | --- | --- |
| Table S4: Average follow-up period for comparisons | | |
| Domain | **Patients (months)** | **Controls (months)** |
| Global | 18.9 | 20.4 |
| Processing Speed | 19.2 | 20.1 |
| Attention | 19.8 | 22.2 |
| Working Memory | 19.3 | 21.3 |
| Problem solving | 18.9 | 21.1 |
| Verbal Learning | 21.8 | 18.6 |
| Verbal Fluency | 17.9 | 15.4 |
| Visual learning | 25.4 | 23.6 |
| Social Cognition | 20.3 | 26.5 |

***S4 - Patient-Control Differences at baseline***

Most studies found evidence of significantly poorer cognition across all domains in the patient group (Addington et al., 2005, 2005; Albus et al., 2006; Haatveit et al., 2015; Kenney et al., 2015; Lindgren et al., 2020; Liu et al., 2011; Sánchez-Torres et al., 2018; Zhou et al., 2017). Other studies found impairment across most domains, with exceptions of attention, visual memory and executive function (de Mello Ayres et al., 2010), working memory (Torgalsbøen et al., 2014), IQ (Lho et al., 2020), and social cognition (Higgins et al., 2021). Our baseline analysis showed large impairments in the patient group on global cognition (g = 0.85) and large or medium impairments across all sub-domains; attention (g = 0.66), reasoning and problem solving (g = 0.83), processing speed (g = 1.24), verbal learning (g = 1.14), visual learning and memory (0.80), working memory (g = 0.83), social cognition (0.59), and verbal fluency (g = 0.97).

***S5 – Change in cognition***

***Patient change in cognition – Domains***

Patient samples showed significant improvement in 7 out of the 11 domains (Table 1), including global cognition (Figure 2), with no significant improvement found in visual learning and memory, verbal fluency, verbal and language, construction and visuospatial, or motor skills. Significant heterogeneity was observed for Verbal Learning and Memory, with two clear outliers (González-Ortega et al., 2013; Torgalsbøen et al., 2014). Removal of these outliers increased the effect size [g = 0.21, p = <0.001] and homogeneity of the sample [Q(p) = 0.61, I2 = 0%], but did not change the direction or significance of the findings. Forest plots for each domain are included in the supplementary materials.

***Healthy control change in cognition – Domains***

Healthy controls significantly improved on 7 of 10 domains for which data were available, with no significant improvement in verbal learning and memory, visual learning and memory, or social cognition. There was significant heterogeneity for Verbal Learning and Memory [Q(p) = 0.04, I 2 = 55%], and Construction and Visuospatial Skills [Q(p) = 0.02, I 2 = 74%], with no clear outliers detected.

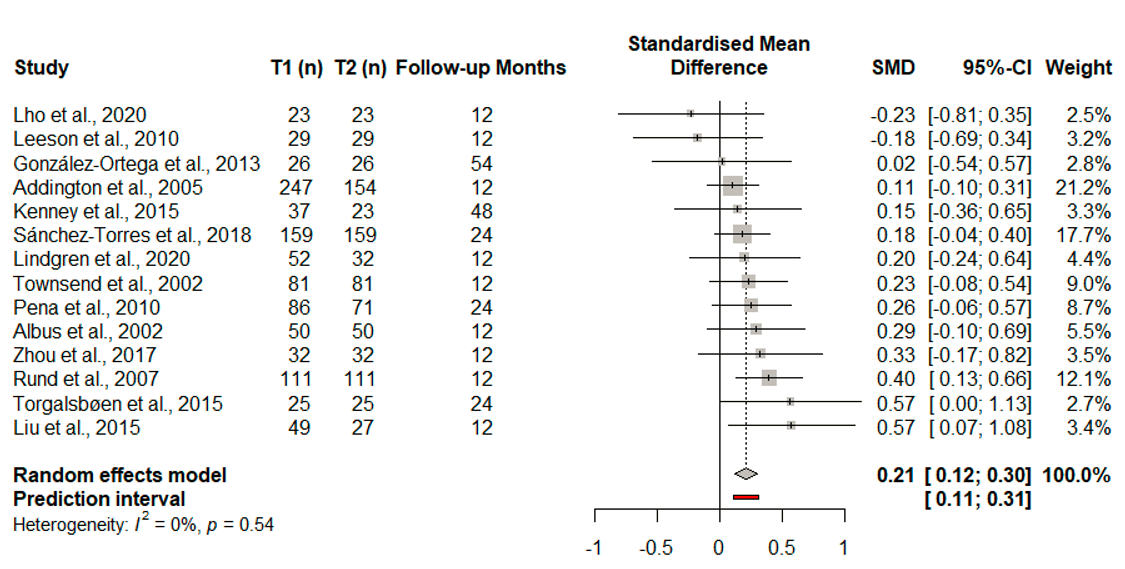
**Patient vs Control difference in change – Domains**

In order to directly compare patient vs control change for only studies which included both patient and healthy control groups, we performed a sensitivity meta-analysis (i.e., calculating an effect size of difference in group-change for each study, omitting those without a control group) (Becker., 1998; Morris., 2007). Hedges' g was calculated for each study using a conservative value of 0.5 for pre-post correlations, and random-effects meta-analyses were performed using the inverse variance method. Effect sizes were adjusted in each case so that positive effect sizes indicated greater improvement in cognitive performance. Results of this analysis did not differ from the primary analysis, with an effect size of 0.01 for global cognition and no significant differences across any domain (see supplementary table S5)

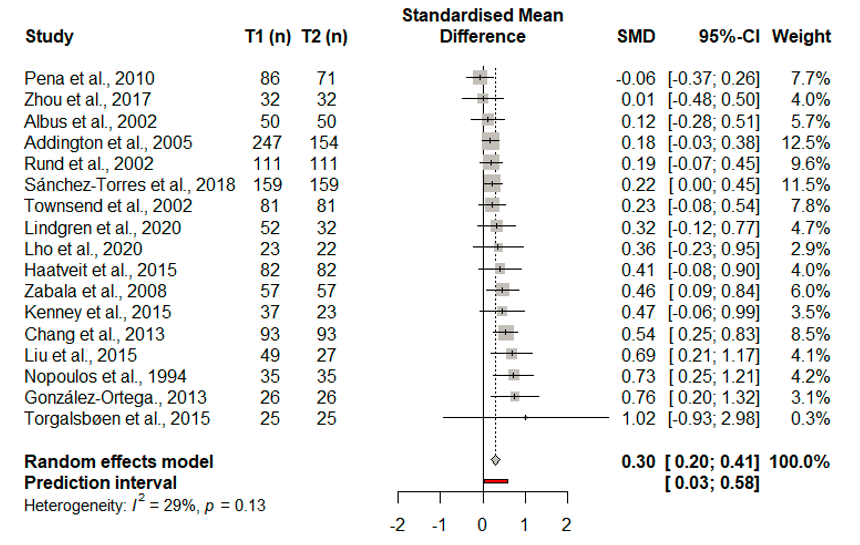
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S5– Meta-analyses of pooled effect sizes for patient-control differences in change scores within studies | | | | | | | | | | |
| Domain | **k** | **Estimated effect (g)** | **95% CI** | **Z** | **p** | **Q** | **Q (p)** | **R2** | ***I2%*** | **Bias (*p*)** |
| Global Cognition | 11 | 0.01 | -0.10 to 0.12 | 0.18 | 0.96 | 9.63 | 0.47 | 0 | 0 | 0.75 |
| Speed of Processing | 10 | -0.15 | -0.35 to 0.04 | -1.52 | 0.13 | 23.87 | <0.01 | 0.06 | 62 | 0.76 |
| Reasoning and Problem Solving | 9 | 0.15 | -0.03 to 0.33 | 1.66 | 0.10 | 18.01 | 0.02 | 0.04 | 56 | 0.61 |
| Attention and Vigilance | 5 | 0.13 | -0.17 to 0.43 | 0.83 | 0.40 | 15.36 | <0.01 | 0.08 | 74 | 0.64 |
| Working Memory | 6 | 0.00 | -0.14 to 0.14 | 0.00 | 0.99 | 4.06 | 0.54 | 0 | 0 | 0.84 |
| Verbal Learning and Memory | 6 | 0.08 | -0.06 to 0.23 | 1.15 | 0.25 | 2.96 | 0.71 | 0 | 0 | 0.40 |
| Visual Learning and Memory | 5 | 0.09 | -0.11 to 0.29 | 0.91 | 0.37 | 8.17 | 0.09 | 0.02 | 51 | 0.80 |
| Social Cognition | 4 | 0.17 | -0.19 to 0.54 | 0.92 | 0.36 | 11.08 | 0.01 | 0.09 | 73 | 0.37 |
| Verbal Fluency | 6 | -0.08 | 0.23 to 0.06 | -1.09 | 0.27 | 6.28 | 0.28 | 0.01 | 20 | 0.86 |
| Estimated effect (g) = pooled effect of patient-control difference in change scores for studies with patient and control groups  *Q =* measure of the heterogeneity of the distribution of effect size  I2 = quantifies of the percentage of total variation across studies due to heterogeneity  Bias = p value of Egger’s test  Results reported before any exclusion of outliers | | | | | | | | | | |

## **S6 - Forest Plots for Patient Domain Meta-analyses (Scores to the left of 0 represent worse performance at follow-up, scores to the right represent improvement)**

## Speed of Processing

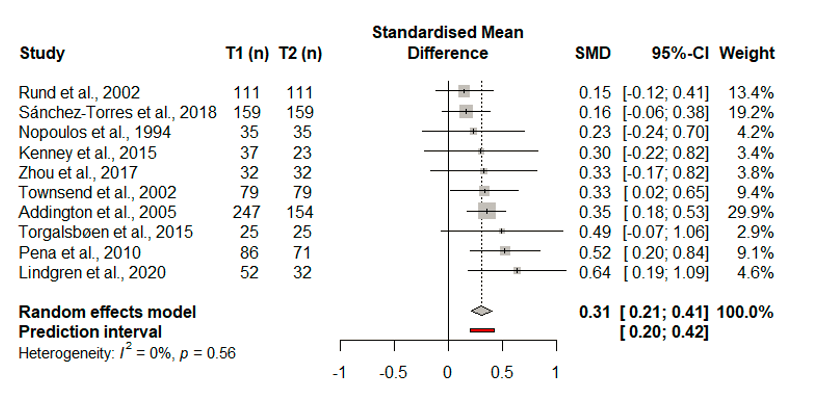
Forest plot showing change in speed of processing in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Reasoning and Problem Solving



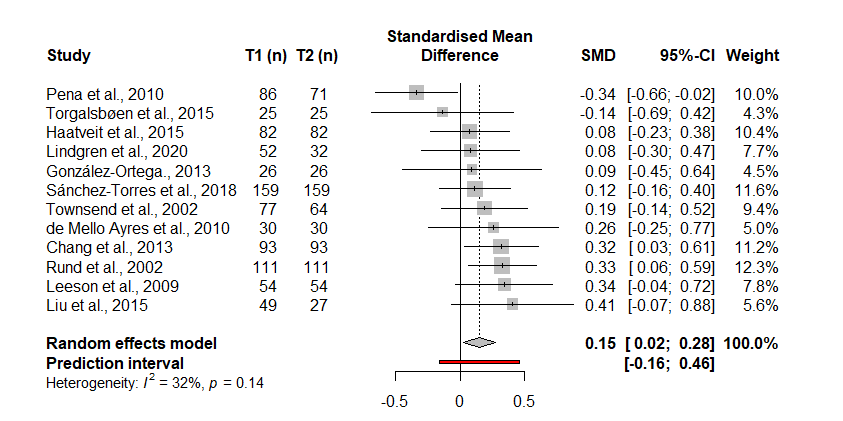
Forest plot showing change in Reasoning and Problem Solving in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

### Attention and Vigilance



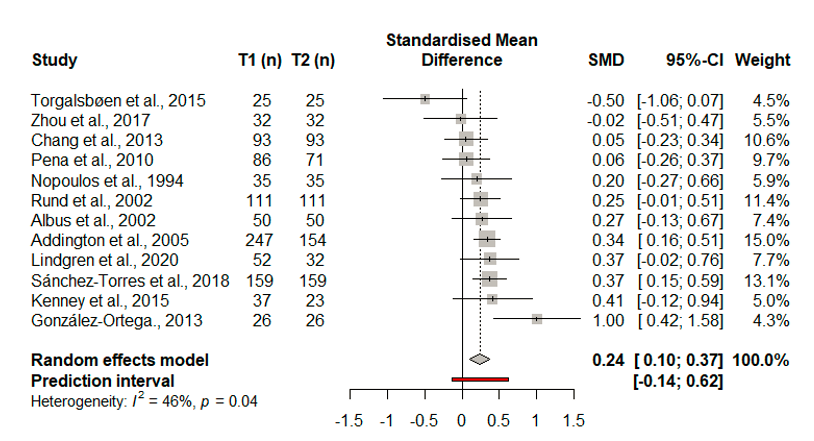
Forest plot showing change in Attention and Vigilance in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Working Memory



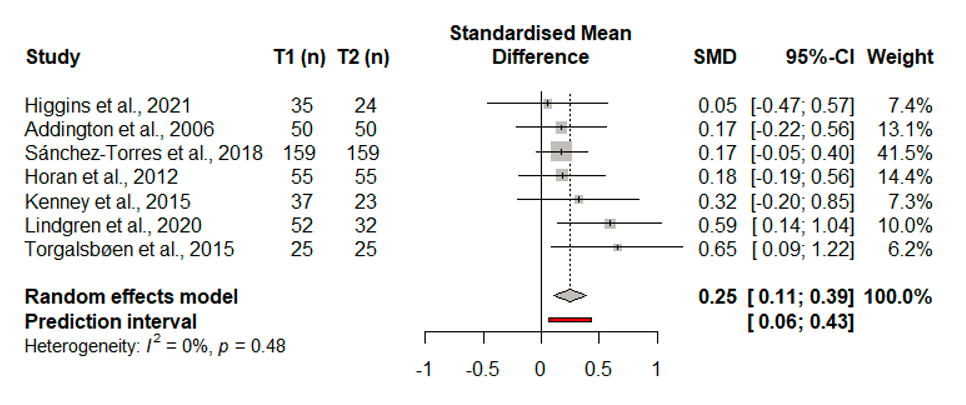
Forest plot showing change in Working Memory in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Verbal Learning and Memory



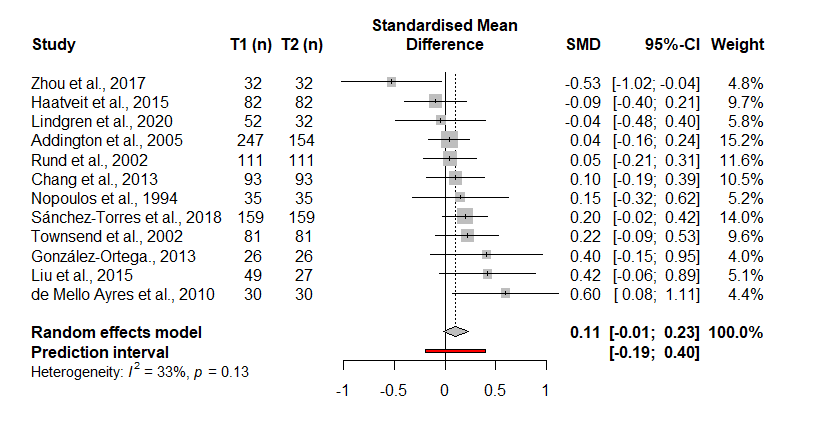
Forest plot showing change in Working Memory in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Social Cognition



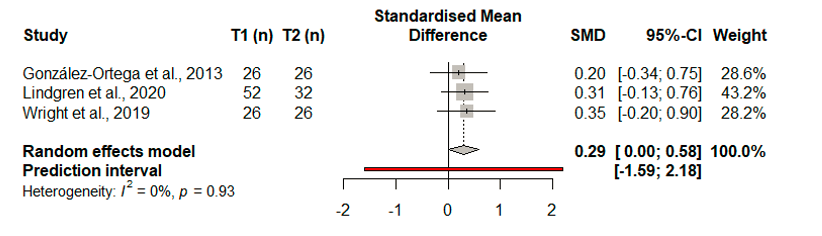
Forest plot showing change in Social Cognition in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Verbal Fluency



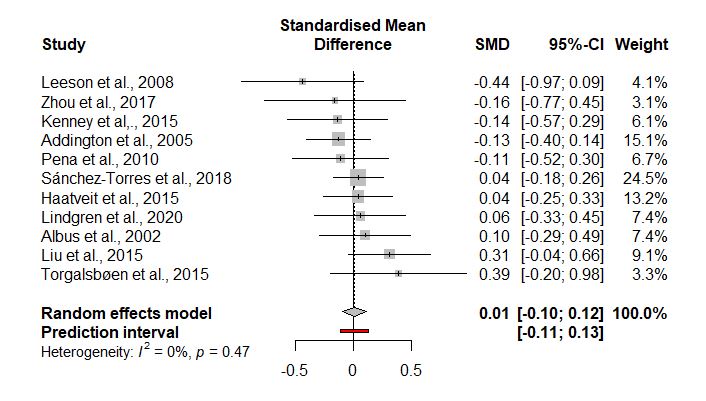
Forest plot showing change in Verbal Fluency in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Verbal and Language Skills



Forest plot showing change in Verbal and Language Skills in the patient samples (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate.

## Global Cognition – Patients vs Controls



Forest plot showing comparison of change in Global Cognition in patients vs controls for studies including a healthy control group (SMD = Hedges’ g); p value is for Q test; diamond = overall estimate. Left of centre line = greater improvement in controls, right of centre line = greater improvement in patients

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table S7: Results of Meta-regressions for the patient samples | | | | | | |  |
| Variable | **k** | **g** | **s.e.** | **z** | **95% CI** | **ToM (p)** | **ToRH** |
| Global |  |  | . |  |  |  |  |
| Age | 19 | -0.02 | 0.02 | -1.37 | -0.06 to 0.01 | 0.17 | 0.99 |
| Gender (% male) | 19 | 0.003 | 0.003 | 0.63 | 09.91 to 0.01 | 0.53 | 0.97 |
| Education | 12 | 0 | 0.04 | 0.98 | -0.04 to 0.12 | 0.33 | 0.99 |
| Positive change | 11 | 0 | 0.00 | 0.19 | -0.01 to 0.01 | 0.85 | 0.86 |
| Negative Change | 10 | 0 | 0.001 | 0.82 | -0,01 to 0.01 | 0.41 | 0.85 |
| PANSS Positive Baseline | 10 | -0.01 | 0.01 | -0.61 | -0.03 to 0.01 | 0.54 | 0.84 |
| PANSS Negative Baseline | 10 | 0 | 0.01 | -0.37 | 0.03 to -0.02 | 0.71 | 0.82 |
| % Antipsychotic | 13 | 0.002 | 0.00 | 0.48 | -/0.00 to 0.01 | 0.63 | 0.95 |
| FU Months | 19 | 0 | 0.00 | 0.78 | -0.01 to 0.01 | 0.43 | 0.98 |
| Speed of Processing |  |  |  |  |  |  |  |
| Gender (% male) | 12 | 0 | 0.01 | -0.29 | -0.01 to 0.01 | 0.78 | 0.67 |
| FU Months | 13 | -0.001 | 0.00 | -0.27 | -0.01 to 0.01 | 0.79 | 0.55 |
| Reasoning and Problem Solving |  |  |  |  |  |  |  |
| Age | 17 | -0.02 | 0.03 | -0.75 | -0.07 to 0.03 | 0.45 | 0.10 |
| Gender | 17 | 0 | 0.01 | -0.54 | -0.01 to 0.01 | 0.58 | 0.12 |
| Education | 10 | 0 | 0.06 | 0.01 | -0.12 to 0.12 | 0.99 | 0.10 |
| % Antipsychotic | 12 | -0.01 | 0.01 | -1.02 | -0.02 to 0.001 | 0.31 | 0.19 |
| FU Months | 17 | 0 | 0.01 | 0.89 | -001 to 0.02 | 0.11 | 0.38 |
| Attention and Vigilance |  |  |  |  |  |  |  |
| Age | 10 | -0.01 | 0.03 | -0.32 | -0.06 to 0.05 | 0.74 | 0.47 |
| Gender | 10 | 0 | 0.01 | 0.60 | -0.01 to 0.02 | 0.55 | 0.50 |
| FU Months | 10 | -0.00 | 0.01 | -0.21 | -0.01 to 0.01 | 0.83 | 0.46 |
| Working Memory |  |  |  |  |  |  |  |
| Age | 11 | 0.01 | 0.03 | 0.42 | -0.04 to 0.07 | 0.67 | 0.09 |
| Gender | 11 | -0.01 | 0.01 | -2.14 | -0.02 to -0.00 | 0.03 | 0.35 |
| FU Months | 12 | -0.01 | 0.01 | -1.48 | -0.02 to -0.00 | 0.13 | 0.23 |
| Verbal Learning and Memory |  |  |  |  |  |  |  |
| Age | 12 | 0.01 | 0.03 | 0.27 | -0.05 to 0.07 | 0.79 | 0.03 |
| Gender | 12 | 0 | 0.01 | 0.10 | -0.01 to 0.02 | 0.92 | 0.03 |
| FU Months | 12 | 0.01 | 0.01 | 1.63 | -0.00 to 0.02 | 0.10 | 0.06 |
| Verbal Fluency |  |  |  |  |  |  |  |
| Age | 12 | -0.01 | 0.03 | -0.34 | -0.07 to 0.05 | 0.73 | 0.09 |
| Gender | 12 | 0 | 0.01 | -0.11 | -0.01 to 0.01 | 0.91 | 0.09 |
| FU Months | 12 | 0.01 | 0.01 | 1.72 | -0.00 to 0.02 | 0.08 | 0.24 |
| ToM = Test of moderator  ToRH = test of residual heterogeneity | | | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table S8: Results of sensitivity analyses for the patient sample | | | | | |
| Variable | **k** | **g** | **95% CI** | ***I*2%** | *Q*bet |
| Global |  |  |  |  |  |
| Affective Yes | 8 | 0.24 | 0.11 to 0.36 | 0 | Qbet 0.07  p = 0.79 |
| Affective No | 11 | 0.26 | 0.15 to 0.37 | 0 |
| Study Quality |  |  |  |  |  |
| Speed of Processing |  |  |  |  |  |
| Affective Yes | 5 | 0.24 | 0.10 to 0.38 | 0 | Qbet 0.72  P=0.40 |
| Affective No | 8 | 0.16 | 0.03 o 0.29 | 0 |
| Study Quality |  |  |  |  |  |
| Reasoning and Problem Solving |  |  |  |  |  |
| Affective Yes | 6 | 0.24 | 0.07 to 0.43 | 25 | Qbet 0.64  p = 0.43 |
| Affective No | 11 | 0.34 | 0.21 to 0.48 | 25 |
| Study Quality |  |  |  |  |  |
| Attention and Vigilance |  |  |  |  |  |
| Affective Yes | 6 | 0.40 | 0.02 to 0.33 | 0 | Qbet 5.14  p = 0.02 |
| Affective No | 4 | 0.18 | 0.28 to 0.53 | 0 |
| Study Quality |  |  |  |  |  |
| Working Memory |  |  |  |  |  |
| Affective Yes | 5 | 0.06 | -0.17 to 0.29 | 61 | Qbet 1.57  p = 0.21 |
| Affective No | 7 | 0.24 | 0.09 0.38 | 0 |
| Study Quality |  |  |  |  |  |
| Verbal Learning and Memory |  |  |  |  |  |
| Affective Yes | 5 | 0.34 | 0.12 to 0.56 | 8.65 | Qbet 1.51  p = 0.22 |
| Affective No | 7 | 0.16 | -0.02 to 0.34 | 10.63 |  |
| Study Quality |  |  |  |  |  |
| Verbal Fluency |  |  |  |  |  |
| Affective Yes | 4 | 0.10 | -0.06 to 0.26 | 17.7 | Qbet 0.51  p = 0.78 |
| Affective No | 8 | 0.11 | -0.06 to 0.29 | 45.4 |
| Study Quality |  |  |  |  |  |
| I2 quantifies of the percentage of total variation across studies due to heterogeneity | | | | | |

**S9 - Task-specific analyses**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S8 – Meta-analyses of change in cognition across tasks, for patient and healthy controls of includes samples | | | | | | | | | | | |
| Task | **k** | ***n*** | **Estimated effect (g)** | **95% CI** | **Z** | **p** | **Q** | **Q (*p*)** | **R2** | ***I*2%** | *Q*bet |
| TMTa |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 8 | 787 | 0.15 | -0.02 to 0.32 | 1.176 | 0.08 | 16.43 | 0.02 | 0.03 | 57 | Qbet 0.59  p = 0.44 |
| HCs | 4 | 313 | 0.25 | 0.07 to 0.43 | 2.66 | 0.01 | 3.77 | 0.28 | 0.01 | 20 |
| Digit Symbol Substitution | | |  |  |  |  |  |  |  |  |  |
| Patients | 5 | 243 | 0.32 | 0.08 to 0.56 | 2.59 | 0.009 | 5.91 | 0.21 | 0.02 | 32,3 | Not enough data |
| HCs | 4 | 175 | 0.32 | 0.11 to 0.53 | 2.99 | 0.003 | 1.33 | 0.72 | 0 | 0 |
| TMTb |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 9 | 786 | 0.11 | 0.00 to 0.21 | 2.05 | 0.04 | 7.26 | 0.51 | 0 | 0 | Qbet 0.30  p = 0.59 |
| HCs | 4 | 329 | 0.17 | -0.01 to 0.35 | 1.80 | 0.07 | 3.93 | 0.27 | 0.01 | 23 |
| WCST Categories Correct | | |  |  |  |  |  |  |  |  |  |
| Patients | 9 | 938 | 0.29 | 0.19 to 0.40 | 1.72 | <0.001 | 9.78 | 0.28 | 0 | 18 | Qbet 3.02  p = 0.08 |
| HCs | 4 | 301 | 0.12 | -0.04 to 0.28 | 1.51 | 0.13 | 1.61 | 0.66 | 0 | 0 |
| WCST Perseverative Errors | | |  |  |  |  |  |  |  |  |  |
| Patients | 8 | 860 | 0.20 | -0.03 to 0.43 | 1.72 | 0.84 | 34.73 | <0.001 | 0.08 | 79.8 | Qbet 0.06  p = 0.81 |
| HCs | 4 | 301 | 0.25 | -0.09 to 0.59 | 1.45 | 0.15 | 11.43 | 0.01 | 0.09 | 73.7 |
| Stroop Interference | | |  |  |  |  |  |  |  |  |  |
| Patients | 8 | 687 | 0.26 | 0.10 to 0.42 | 3.20 | 0.001 | 12.32 | 0.09 | 0.02 | 43 | Qbet 0.01  p = 0.94 |
| HCs | 4 | 284 | 0.25 | 0.09 to 0.42 | 2.99 | 0.002 | 0.99 | 0.80 | 0 | 0 |
| LNS Total | |  |  |  |  |  |  |  |  |  |  |
| Patients | 5 | 405 | -0.09 | 0.55 to 0.37 | -0.38 | 0.70 | 39.16 | <0.001 | 0.24 | 90 | Not enough data |
| HCs | 3 | 247 | 0.20 | -0.10 to 0.32 | 2,22 | 0.03 | 1.51 | 0.47 | 0 | 0 |
| LMI |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 6 | 503 | 0.44 | 0.08 to 0.80 | 2.41 | 0.02 | 28.59 | 0.15 | 28.59 | 83 | Not enough data |
| HCs | 3 | 178 | 0.11 | -0.10 to 0.32 | 1.00 | 0.32 | 0.27 | 0.87 | 0 | 0 |
| LMII |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 5 | 410 | 0.27 | 0.08 to 0.80 | 2.9 | 0.004 | 4.82 | 0.31 | 0.01 | 17 | Not enough data |
| HCs | 4 | 212 | 0.20 | 0.01 to 0.39 | 2.03 | 0.04 | 1.10 | 0.78 | 0 | 0 |
| Patients | 3 |  | 0.21 | -0.14 to 0.57 | 1.17 | 0.25 | 3.43 | 0.18 | 0.04 | 42 |  |
| VF Categories | |  |  |  |  |  |  |  |  |  |  |
| Patients | 6 | 639 | 0.07 | -0.05 to 0.15 | 1.16 | 0.24 | 3.17 | 0.67 | 0 | 0 | Qbet 0.65  p = 0.42 |
| HCs | 5 | 375 | 0.15 | 0.00 to 0.29 | 1.97 | 0.05 | 3.41 | 0.49 | 0 | 0 |
| VF Letters | |  |  |  |  |  |  |  |  |  |  |
| Patients | 6 | 666 | -0.04 | -0.28 to 0.21 | -0.31 | 0.76 | 20.66 | 0.001 | 0.07 | 76 | Qbet 3.65  p = 0.06 |
| HCs | 5 | 375 | 0.25 | 0.08 to 0.42 | 2.94 | 0.003 | 4.86 | 0.30 | 0.01 | 18 |
| CPT-IP |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 242 | 0.34 | 0.15 to 0.54 | 3.5 | <0.001 | 2.16 | 0.34 | 0 | 7 |  |
| Digit Backwards | |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 198 | 0.17 | -0.04 to 0.37 | 1.60 | 0.11 | 1.99 | 0.37 | 0 | 0 |  |
| CVLT |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 320 | 0.29 | 0.12 to 0.45 | 3.37 | 0.001 | 2.3 | 0.32 | 0 | 13 |  |
| HVLT |  |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 94 | -0.08 | -0.59 to 0.43 | -0.31 | 0.76 | 5.56 | 0.06 | 0.12 | 64 |  |
| MSCEIT-ME Total | | | | | | | | | | | |
| Patients | 3 | 221 | 0.28 | 0.04 to 0.52 | 2.26 | 0.02 | 2.46 | 0.29 | 0.01 | 19 |  |
| COWAT Total | |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 117 | 0.20 | -0.16 to 0.56 | 1.08 | 0.28 | 3.53 | 0.17 | 0.05 | 43 |  |
| WAIS Vocabulary | |  |  |  |  |  |  |  |  |  |  |
| Patients | 3 | 104 | 0.29 | 0.00 to 0.58 | 1.98 | 0.05 | 0.16 | 0.93 | 0 | 0 |  |
| Estimated effect (g) = improvement in performance from first assessment to follow-up  *Q =* measure of the heterogeneity of the distribution of effect size  I2 = quantifies of the percentage of total variation across studies due to heterogeneity  Bias = p value of Egger’s test  Qbet = comparison of change in patients vs healthy controls  Results reported before any exclusion of outliers | | | | | | | | | | | |

***Patient change in cognition – Tasks***

The patient samples significantly improved on 9 out of 19 tasks, with no improvement on the TMTa, WCST perseverative errors, LNS total, VF categories or letters, digit span backwards or total, HVLT, COWAT total, or WAIS vocabulary. Significant heterogeneity was observed for several tasks and outliers were detected for TMTa (González-Ortega et al., 2013), WCST perseverative errors (González-Ortega et al., 2013), LNS total (Peña et al., 2011), LMI (González-Ortega et al., 2013) and VF letters (Zhou et al., 2017) tasks. Leave-one-out analyses resulted in increases in effect sizes for the TMTa [g = 0.21, p = 0.001], WCST perseverative [g = 0.29, p = 0.03], LNS total [g = 0.14, p = 0.07], and VF letters [g = 0.09, p = 0.14], and reduced effect sizes for LMI [g = 0.24, p = 0.001], and Stroop interference [g = 0.20, p = <0.001]. After removal of outliers, improvement on TMTa and WCST perseverative errors was significant (p = 0.002 and p = 0.003), and samples were largely homogenous except for WCST perseverative errors where significant heterogeneity remained (q = 0.01, I2 = 71).

***Healthy controls change in cognition – Tasks***

HCs significantly improved on 6 out of 11 tasks for which data were available. No improvement was found on the TMTb, WCST categories or perseverative errors, LMI, or VF categories. HC samples were largely homogenous, except for WCST perseverative errors where there was significant heterogeneity and no outliers detected.

**Supplementary materials references**

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Dickinson, D., Ramsey, M. E., & Gold, J. M. (2007). Overlooking the obvious: A meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Archives of General Psychiatry*, *64*(5), 532–542. https://doi.org/10.1001/archpsyc.64.5.532

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-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. *Neuroscience and Biobehavioral Reviews*, *35*(3), 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001

Green, M. F., Kern, R. S., Braff, D. L., & Mintz, J. (2000). Neurocognitive Deficits and Functional Outcome in Schizophrenia: Are We Measuring the “Right Stuff”? *Schizophrenia Bulletin*, *26*(1), 119–136. https://doi.org/10.1093/oxfordjournals.schbul.a033430

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology*, *12*(3), 426–445. https://doi.org/10.1037//0894-4105.12.3.426

Kern, R. S., Green, M. F., & Marder, S. R. (2007). The NIMH MATRICS Initiative: Development of a Consensus Cognitive Battery. *Progress in Neurotherapeutics and Neuropsychopharmacology*, *2*(1), 173–186. https://doi.org/10.1017/S1748232106000127

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment, 4th ed* (pp. xiv, 1016). Oxford University Press.

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. *Neuropsychology*, *23*(3), 315–336. https://doi.org/10.1037/a0014708

Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., & Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research*, *72*(1), 29–39. https://doi.org/10.1016/j.schres.2004.09.007