**Supplementary Materials**

**The PRONIA COGNITIVE BATTERY (PCB)**

**Recruitment of participants**

In the preliminary screening phase 5547 participants were assessed for eligibility to participate between February 2014 and May 2017 (see also Koutsouleris et al. 2020). After screening, 147 patients at clinical high risk (CHR), 155 patients with recent onset depression (ROD), 163 patients with recent onset psychosis (ROP) and 286 healthy controls (HC) (N=751 participants) entered the dataset between February 2014 and May 2016. Of these, 116 CHR, 123 ROD, 105 ROP and 247 HC entered the analyses, while 31 CHR, 32 ROD, 58 ROP, 39 HC (N=160 participants) were excluded because of the presence of a too high proportion (>20%) of invalid or missing neurocognitive data. Additionally, 178 HC were tested between April 2016 and May 2017. Of those HC, 125 were included in the analyses of the current study, while 53 were excluded because of the presence of more than 20% of invalid or missing neurocognitive data.

**The neurocognitive assessment**

From a broader list of tests that were proposed based on clinical and neuropsychological expertise of the consortium members, ten were shortlisted based on the previous literature in the leading cognitive research studies and cognitive batteries in the field of psychosis, such as the NAPLS 2 consortium of studies. PRONIA Cognitive Battery (PCB) was implemented for administration in the Psychology Experiment Building Language (PEBL) running on Windows8 OS (or later). As PRONIA was a multi centric European project, involving centers located in Germany (Munich, Cologne, Düsseldorf), Italy (Milan, Udine, Bari), Switzerland (Basel), UK (Birmingham), Finland (Turku), the PCB was implemented in 4 different languages (German, English, Italian, Finnish)

The neurocognitive assessment was administered across the PRONIA consortium by research assistants who received ad hoc training and continuous supervision. For ensuring the highest consistency, all the examiners who joined the project attended a webinar concerning the neurocognitive assessment administration and scoring procedures. Also, an additional in-person training was given and written manuals were made available to all the examiners.

At the end of each acquisition, an automated first quality check was run, by mean of a script locally installed into the device used for the tests’ administration, to control the data for the presence of exceptions: i) values exceeding the correct score range; ii) decimals instead of integers or vice versa; iii) missing scores.

The acquired data were then uploaded into the PRONIA portal and checked again by an automated procedure that checked for inconsistencies on the calculated scores (e.g., the value entered as “total accuracy” did not correspond to the actual sum of the accuracies at each single item), and further reported the presence of missing data that could have been caused by possible upload failures. Feedback from this second level quality check was then sent to each site’s case managers who manually corrected the inconsistencies, if any. When corrected, the data were locked into the portal.

The list of tests in

**Inclusion and exclusion criteria**

**General inclusion criteria**

All participants were in the age range between 15 and 40 years, their language comprehension and production were appropriate to participating and could provide their informed consent to the study. The participants' language comprehension and production were evaluated by the rater during the preliminary screening. If a participant was not a native speaker they were required to have completed their high school education (or higher) using the local language. Non-native speakers who did not complete their high school education in the local language but were English native speakers were tested with the English version of the PCB.

**Patients inclusion criteria**

Patients qualified as CHR if meeting the criteria for i) cognitive disturbances (COGDIS) according to the Schizophrenia Proneness Instrument (SPI-A) [1] and/or ii) ultra-high-risk (UHR) criteria for psychosis according to the Structured Interview for Psychosis-Risk Syndromes, Positive scale [2] (SIPS-P, unusual thought content/ delusional ideas, suspiciousness/persecutory ideas, grandiosity, perceptual abnormalities/hallucinations, and disorganized communication). COGDIS criteria were added in addition to UHR criteria as cognitive basic symptoms are frequently reported in individuals at high risk of psychosis. Also, higher levels of subjective cognitive disturbances are observed in patients that later develop first-episode psychosis [3]. The COGDIS criteria are fulfilled Concerning the threshold if the participants report at least 2 among the following basic symptoms: inability to divide attention; thought interference ;thought pressure;thought blockages; disturbance of receptive speech; disturbance of expressive speech; unstable ideas of reference; disturbance of abstract thinking; captivation of attention by details of the visual field. In addition, basic symptoms need to be reported at öleast once a week and present in the past 3 months.

Patients reporting overt and stable psychotic symptoms (at the SIPS-P) which did not remit spontaneously and emerged within one year of baseline, were assigned to the ROP clinical group.

ROD patients fulfilled the criteria for major depression at the Structured Clinical Interview for DSM-IV-TR (SCID) [4][5] within the past three months. Also, they never had a previous depressive episode, and the current episode had not lasted for more than 24 months.

**General exclusion criteria**

Exclusion criteria for all participants in the main study were: i) an IQ lower than 70, as estimated by mean of the Vocabulary and Matrix Reasoning subtests from the WAIS-IV [6]; ii) the existence of neurological or somatic health conditions that could affect the brain’s functions; iii) past or current head trauma with a loss of consciousness longer than 5 minutes; iv) substance or alcohol dependence within the past 6 months; and v) impossibility to carry out an MRI exam. Also, HC were excluded if: i) they were ever diagnosed with a DSM-V axis disorder; ii) they had a first degree relative who ever received a diagnosis for affective or non-affective psychosis; iii) psychotropic medications intake for more than 5 times/year and/or in the month before participation to the study.

**Patients exclusion criteria**

Finally, additional exclusion criteria applied to CHR and ROD: i) antipsychotic medication intake for more than 30 cumulative days at the minimum dosage in the past year; ii) antipsychotic medication intake within the 3 months preceding the baseline evaluation.

**Confirmatory Factor Analysis**

An analysis of the factorial structure was carried out in order to highlight the main domains measured by the PCB. Moreover, the PCB scores were fitted to the appropriate factorial model so to obtain a reduce number of dimensions, instead of a larger number of measures, to be used for later data analyses.

The confounding effects of i) age, ii) site, iii) sex, iv) years of education, v) language, were removed at this stage (i.e. the four factors SP, AWM, VL, SL  scores have the effects of age, site, as well as sex, years of education, language, already removed)

16 measures drawn from the PCB (see Supplementary Table 1) were entered into a Confirmatory Factors Analysis (CFA).

At first the CFA tested a model with 5 factors. The choice of the initial number of factors was decided based on the existing literature concerning the MATRICS Consensus Cognitive Battery [7]. Although the original model from the MATRICS battery included 7 factors (i.e. speed of processing, attention/vigilance, working memory, verbal learning, visual learning, reasoning and problem solving, social cognition), we started from a 5 factors model that did not include i) reasoning and problem solving, ii) social cognition (which are in the 7 factors model of MATRICS). Indeed, the CFA on the MATRICS battery included in “reasoning and problem solving” the following tests: i) block design sub-test of the WAIS-III, ii) the Tower of London from BACS, iii) the mazes subtest from the Neuropsychological Assessment. The PCB did not include any of those tests, nor tests that can be considered as their equivalent. For this reason the factor “reasoning and problem solving” was excluded a-priori from the initial model. Also, we did not include “social cognition” as the measure of social cognition of the PCB (i.e. the DANVA) turned out to be rather easy for the participants and showed a ceiling effect.

Progressively simpler CFA models were compared in order to choose the simplest model leading to the least loss of information (i.e. the model that allows the variance to be explained by the smallest number of factors). So, Then the model was reduced to 4 and 3 factors, progressively excluding the factor that proved to less affecting the general model fit as compared to the model including it.

The following procedures apply to all the models:

-  Full information maximum likelihood used for handling of missings.

-  Correlated (non-orthogonal) latent factors.

-  Latent factors standardized (fixing the variance to 1).

-  Observed univariate-outliers were excluded in fitting the models (ta have better distributions).

-  Observed measures standardized in the sample (to avoid different-scale effects).

-  Observed measures with intercepts fixed to 1.

-  Observed measures with independent residuals, with the exception of same-test correlations:

Supplementary Table 1 about here

**CFA with 5 factors**

The PCB tests were grouped according to 5 of the 6 factors that were previously used for grouping the tests of the MATRICCS Consensus Cognitive Battery:

1) SP, Speed of processing:

* Verbal Fluency (VF), total phonemic score
* VF, total semantic score
* Trail Making Task (TMA) - A, time
* CPT-IP, Hit reaction time
* Digit Symbol Substitution Test - total score

2) AV, Attention / Vigilance:

* Auditory Digit Span (ADS)- Forward, Correct responses
* CPT-IP, random accuracy.

3) WM, Working memory:

* ADS-Backward, Correct responses
* TMT switch
* CPT-IP, commission accuracy
* Self-Ordered Pointing Test (SOPT), perseverations

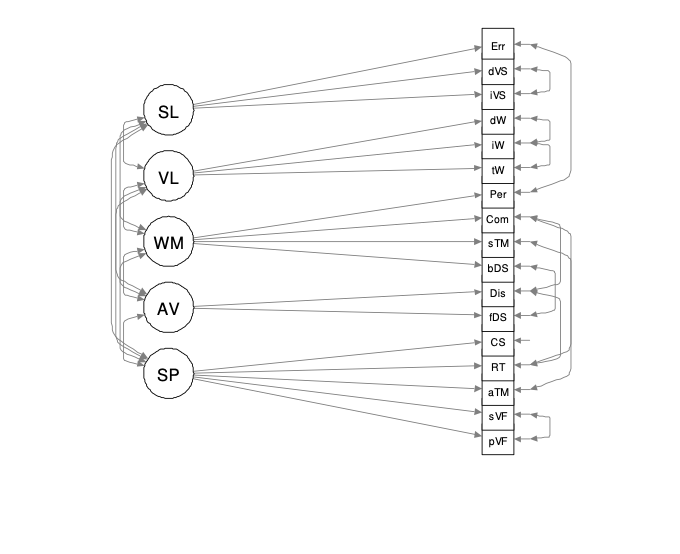
4) VL, Verbal learning and memory:

* Rey Auditory Verbal Learning Test (RAVLT) correct Immediate whole:
* RAVLT, correct Immediate 1st trial
* RAVLT, delayed

5) SL, Visual learning and memory:

* Rey Osterrieth Figure Copy (ROCF), immediate memory
* ROCF, delayed memory

Note: with respect to the MATRICCS the factor “Social Cognition” was not included as the DANVA-2-AF was the only test of the PCB measuring this construct. Also the DANVA-2-AF scores were at ceiling.

Visual representation of the model:

Model fit:

* χ2 (99) = 221.056, p < 0.001 (the model do not fit perfectly).
* χ2/df = 2.23 (acceptable)
* Root Mean Square Error of Approximation (RMSEA) = 0.045, 95% ci [0.037, 0.053] (adequate)
* Root mean square of the residuals (SRMR): 0.045 (adequate)
* Goodness of Fit Index (GFI): 0.958 (adequate)
* Comparative Fit Index (CFI): 0.962 (adequate)
* Tucker Lewis Index of factoring reliability (TLI): 0.948 (adequate)

Compared with a 1-factor model:

* AIC(5F) = 25357 vs AIC(1F) = 25617
* BIC(5F) = 25670 vs BIC(1F) = 25885
* χ2-test: Δχ2(10) = 279.5, p < 0.001 (adequate)

**CFA with 4 factors**

The 5 factors model was then reduced to 4 merging the domains AV and WM into the domain “Attention” (A). So, the factorial structure of the model was the following:

1) SP, Speed of processing:

* VF, total phonemic score
* VF, total semantic score
* TMT, time in part A
* CPT-IP, Hit reaction time
* Digit Symbol Substitution Test - total score

2) A, Attention (AV, Attention / Vigilance + WM, Working memory):

* ADS- F, Correct responses
* CPT-IP, random accuracy
* ADS-B correct responses
* TMT switch
* CPT-IP, commission accuracy
* SOPT, perseverations

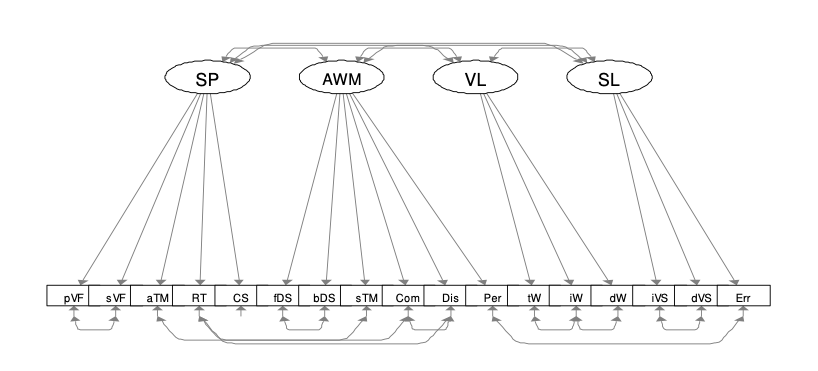
3) VL, Verbal learning and memory:

* RAVLT correct Immediate whole
* RAVLT, correct Immediate 1st trial
* RAVLT, delayed

4) SL, Visual learning and memory:

* ROCF, immediate memory
* ROCF, delayed memory

Visual representation of the model:



Model fit:

* χ2 (103) = 221.425, p < 0.001 (the model do not fit perfectly)
* χ2/df = 2.15 (acceptable)
* Root Mean Square Error of Approximation (RMSEA) = 0.044, 95% ci [0.036, 0.052] (adequate)
* Root mean square of the residuals (SRMR): 0.045 (adequate)
* Goodness of Fit Index (GFI): 0.958 (adequate)
* Comparative Fit Index (CFI): 0.963 (adequate)
* Tucker Lewis Index of factoring reliability (TLI): 0.952 (adequate)

Compared with a 1-factor model:

* AIC(4F) = 25350 vs AIC(1F)= 25617
* BIC(4F) = 25645 vs BIC(1F) = 25885
* χ2-test: Δχ2 (6)= 279.13, p < 0.001 (adequate)

Compared with a 5-factor model:

* AIC(4F) = 25350 vs AIC(5F) = 25357
* BIC(4F) = 25645 vs BIC(5F) = 25670
* χ2-test: Δχ2(4) = 0.369, p < 0.985 (the model with 4-factor is preferable to the model with 5-factors)

**CFA with 3 factors**

The model was reduced further to 3 factors merging the domains VL and SL into an inclusive “Learning” (L) factor. So, the factorial structure of the model was the following:

1) SP, Speed of processing:

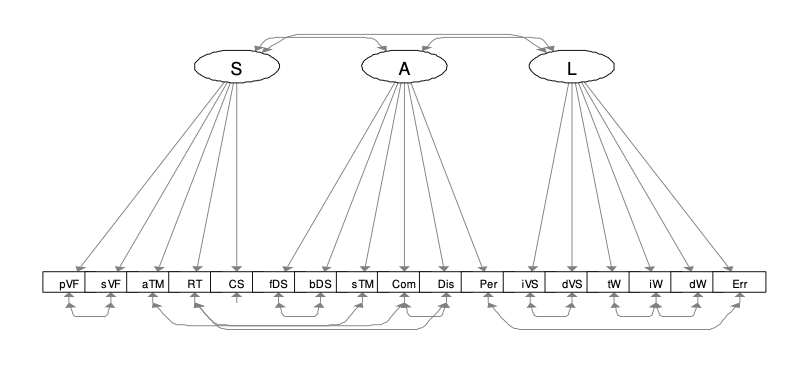
* VF, total phonemic score
* VF, total semantic score
* TMT, time in part A
* CPT-IP, Hit reaction time
* Digit Symbol Substitution Test - total score

2) A, Attention (AV, Attention / Vigilance + WM, Working memory):

* ADS- F, Correct responses
* CPT-IP, random accuracy
* ADS-B, correct responses
* TMT switch
* CPT-IP, commission accuracy
* SOPT, perseverations

3) L, Learning (VL, Verbal learning and memory + SL, Visual learning and memory):

* RAVLT correct Immediate whole
* RAVLT, correct Immediate 1st trial
* RAVLT, delayed
* ROCF, immediate memory
* ROCF, delayed memory

Visual representation of the model:

Model fit:

* χ2(106) = 247.222, p < 0.001 (the model do not fit perfectly)
* χ2/df = 2.33 (acceptable)
* Root Mean Square Error of Approximation (RMSEA) = 0.047, 95% ci [0.039, 0.055] (adequate)
* Root mean square of the residuals (SRMR): 0.052 (acceptable)
* Goodness of Fit Index (GFI): 0.953 (adequate)
* Comparative Fit Index (CFI): 0.956 (adequate)
* Tucker Lewis Index of factoring reliability (TLI): 0.944 (acceptable)

Compared with a 1-factor model:

* AIC3F = 25370 vs AIC1F = 25617
* BIC3F = 25651 vs BIC1F = 25885
* χ2-test: Δχ2(3) = 253.33, p < 0.001 (adequate)

Compared with a 4-factor model:

* AIC3F = 25370 vs AIC4F = 25350. o BIC3F = 25651 vs BIC4F = 25645
* χ2-test: Δχ2 (3)= 25.80, p < 0.001 (the model with 4-factor is preferable to the model with 3-factors)

**Conclusions**

The 4 factors model proved to being equivalent but simpler than the 5 factors model. Instead the 3 factors model was not comparable or simpler than the one with 4 factors. For this reason the 4 factors model was the one representing the PCB’s factorial structure more appropriately.

The PCB demonstrated to represent the domains: speed of processing, attention, verbal learning and memory, visual learning and memory.

**Supplementary Figure\_1a\_1d**

Supplementary Figure\_1a\_1d about here

**Supplementary Table 2**

Supplementary Table 2 about here

**References**

1 Schultze-Lutter F, Addington J, Ruhrmann S, Klosterkötter J. Schizophrenia proneness instrument, adult version (SPI-A). Rome: Rome: Giovanni Fioriti; 2007.

2 Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Ventura J, McFarlane W, et al. Prodromal Assessment with the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms: Predictive Validity, Interrater Reliability, and Training to Reliability. Schizophr Bull 2003;29:703–15. doi:10.1093/oxfordjournals.schbul.a007040.

3 Schultze-Lutter F, Klosterkötter J, Picker H, Steinmeyer EM, Ruhrmann S. Predicting First-Episode Psychosis by Basic Symptom Criteria. Clin Neuropsychiatry. 2007;11–22.

4 First MB, Gibbon M. The Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In: Hilsenroth MJ, Segal DJ, editors. Compr. Handb. Psychol. Assessment, Vol. 2, Personal. Assess., Hoboken, NJ, USA: John Wiley & Sons, Inc; 2004, p. 134–43.

5 First MB, Spitzer RL, Gibbon M, Williams JBW. User’s Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders. SCID-I. Clinician Version. Washington, DC, USA and London, UK: American Psychiatric Publishing, Inc; 1997.

6 Wechsler D. Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV). 2008; San Antonio, TX, USA: The Psychological Corporation.

7 Nuechterlein KH, Ph D, Green MF, Kern RS, Baade LE, Barch DM, et al. Test Selection , Reliability , and Validity. Psychiatry Interpers Biol Process. 2008;(February):203–13.

8 Osterrieth PA. Le test de copie d’une figure complexe. Arch Psychol (Geneve) 1944;30:2016–356.

9 Nowicki SJ. Manual for the Receptive Rests of the Diagnostic Analysis of Nonverbal Accuracy 2. vol. 2. Atlanta, GA, USA: Department of Psychology, Emory University; 2000.

10 Harrison JE, Buxton P, Husain M, Wise R. Short Test of Semantic and Phonological Fluency: Normal Performance , Validity and Test-Retest Reliability. Br J Clin Psychol 2000;39:181–91. doi:10.1348/014466500163202.

11 Rey A. L’examen clinique en psychologie. Paris, France: Universitaires de France; 1964.

12 McMinn MR, Wiens AN, Crossen JR. Rey Auditory-Verbal Learning Test: Development of Norms for Healthy Young Adults. Clin Neuropsychol 1988;2:67–87. doi:10.1080/13854048808520087.

13 Reitan RM. Trail Making Test: Manual for Administration and Scoring. Tucson, AZ, USA: Reitan Neuropsychology Laboratory; 1979.

14 Reitan RM, Wolfson D. The Halstead–Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. Tucson, AZ, USA: Neuropsychology Press; 1993.

15 Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L. The Continuous Performance Test, Identical Pairs Version (CPT-IP): I. New Findings About Sustained Attention in 16 Normal Families. Psychiatry Res 1988;26:223–38. doi:http://dx.doi.org/10.1016/0165-1781(88)90076-5.

17 Gillett R. Assessment of Working Memory Performance in Self-Ordered Selection Tests. Cortex 2007;43:1047–56. doi:10.1016/S0010-9452(08)70702-0.

18 Keefe RSE, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: Reliability, Sensitivity, and Comparison With a Standard Neurocognitive Battery. Schizophr Res 2004;68:283–97. doi:10.1016/j.schres.2003.09.011.