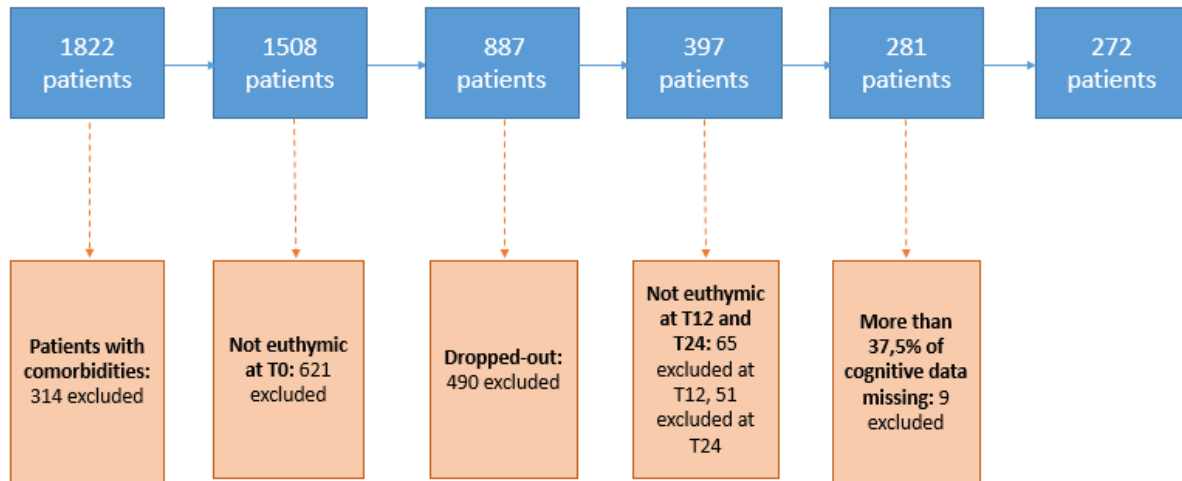## Supplementary methods. Detailed methodology

1 – Selection procedure

2 – Testing longitudinal invariance of the latent variables

We performed multiple group confirmatory factor analyses to test the longitudinal invariance of the latent variables described in the article (1). We examined consensual fit indices and used recommended cutoff criteria (2). The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) should be higher than 0.95 to assume a good fit. The Root Mean Square Error of Approximation (RMSEA) should not be significantly higher than 0.05 and the Standardized Root Mean Square Residual (SRMR) should be lower than 0.08 to assume a good fit.

Nested comparisons were carried out with increasing constraints on the parameters (3). First, the longitudinal invariance of the factorial structure (configural invariance) was tested by assessing the fit indices of the longitudinal CFA with the similar factorial structure across T0, T12, and T24 for functioning and across T0 and T24 for cognition. Then, the factor loadings were constrained to be equal across time (metric model). The metric invariance was tested by comparing Chi² goodness of fit between the configural model and the metric model: a non-significant change in Chi² means that factor loadings were indeed equal across time and that metric invariance was achieved. Finally, the intercepts of each item were constrained to be equal across time (scalar model) and this model was

compared with the metric model. A non-significant change in Chi² means that intercepts were indeed equal across time and that scalar invariance was achieved. Scalar invariance implies that the latent variable was reliable across time and thus could be integrated in the longitudinal structural models.

3 – Required sample size estimation

Velicer & Fava (4) showed that minimum sample size should not be a function of the number of indicators when the model involves latent variables, as these rules of thumb are not supported by evidence and lead to inaccurate estimates. They emphasized the importance of the number of indicators per latent variable. Marsh & Bailey (5) and Boomsma (6) suggested that the ratio of indicator per latent variable is a better basis to calculate adequate sample size. Increasing the number of indicators per latent variable maximizes the amount of information available for estimating parameters, as well the collection of more data. Based on the previous works of Marsh (7–9), Westland (10) suggested the following formula to calculate sample size: **$n >= 50r^2 - 450r + 1100$** (where r is the ratio of indicators per latent variables). For the covariate, which was an observed variable not included into a latent variable, we used the relatively conservative recommendations proposed by Stevens (11) to include 15 observations per exogenous observed variable.

Our model included five latent variables manifested by 30 indicators, thus r = 6. According to the formula cited above, our sample should have been greater than 200 subjects. In addition to the basic model, we added the MADRS depression score as an observed covariate, which was measured at three time points, thus necessitating 45 additional subjects. The total recommended sample size for our model was 245 participants.

References

1. Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika. 1993;58(4):525–43.

2. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999 Jan 1;6(1):1–55.

3. Brown TA. Confirmatory factor analysis for applied research. Second edition. New York ; London: The Guilford Press; 2015. 462 p. (Methodology in the social sciences).

4. Velicer WF, Fava JL. Effects of variable and subject sampling on factor pattern recovery. Psychol Methods. 1998;3(2):231–51.

5. Marsh HW, Bailey M. Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models. Appl Psychol Meas. 1991 Mar;15(1):47–70.

6. Boomsma A. The robustness of LISREL against small sample sizes in factor analysis models. Syst Indirect Obs Causality Struct Predict. 1982;149–73.

7. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychol Bull. 1988;103(3):391–410.

8. Marsh HW, Hau K-T, Balla JR, Grayson D. Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. Multivar Behav Res. 1998 Apr;33(2):181–220.

9. Marsh HW, Balla JR, Hau KT. An Evaluation of Incremental Fit Indexes: A Clarification of Mathematical and Empirical Properties. In: Advanced Structural Equation Modeling Techniques : Issues and Techniques. Lawrence Erlbaum Associates. Mahwah, N.J: G.A. Marcoulides and R.E. Schumaker (eds); 1996. p. 315–53.

10. Westland CJ. Lower bounds on sample size in structural equation modeling. 2010 Nov [cited 2018 Dec 11]; Available from: http://dspace-prod.lib.uic.edu//handle/10027/7655

11. Stevens J. Applied multivariate statistics for the social sciences. Mahwah (New Jersey): L. Erlbaum Associates; 1996.

Supplementary Table 1. Comparison between completers and non-completers at baseline

Variable	Mean (sd) or distribution				Comparison		Effect size
	Completers (n = 397)		Non-completers (n = 490)		Test	p	d or V
Socio-demographic							
Sex (% Male)	40.6		39.2		X² (1) = 0.12	0.73	0.012
Age (year)	42.35	(11.32)	39.9	(12.13)	t (867.4) = 3.1	**0.002**	0.207
Illness duration	17.76	(11.29)	15.63	(10.16)	t (775.8) = 2.75	**0.006**	0.192
Education (year)	14.63	(2.54)	14.2	(2.55)	t (836.1) = 2.42	**0.016**	0.165
Marital status (% Yes)	56.7		49.2		X² (1) = 4.16	**0.041**	0.068
Employed (% Yes)	76.7		71.8		X² (1) = 1.94	0.164	0.047
Individual housing (% Yes)	82.4		75.2		X² (1) =4.67	**0.031**	0.073
Judiciary protection (% Yes)	1.4		2		X² (1) = 0.12	0.728	0.012
Functioning (score 0-12)							
FAST Autonomy	1.21	(2.02)	1.4	(2.28)	t (835.9) = -1.26	0.207	0.086
FAST Occupational	5.7	(6.07)	5.8	(6.02)	t (809) = -0.24	0.811	0.017
FAST Cognition	2.48	(3)	2.94	(3.28)	t (831.1) = -2.1	**0.036**	0.144
FAST Finances	0.56	(1.19)	0.86	(1.59)	t (829.5) = -3.07	**0.002**	0.206
FAST Interpersonal	3.27	(3.3)	3.89	(3.74)	t (836.8) = -2.54	**0.011**	0.174
FAST Leisure time	1.31	(1.59)	1.39	(1.71)	t (829) = -0.68	0.494	0.047
Clinical							
Type of BD (% 1/2/Not otherwise specified)	55.4 / 37.8 / 6.8		49.2 / 36.1 / 14.7		X² (2) = 14.04	**0.001**	0.089
Age at first episode (year)	24.55	(9.24)	24.14	(9.68)	t (818.2) = 0.63	0.529	0.044
Psychotic history (% Yes)	50.4		46.7		X2 (1) =0.85	0.357	0.034
Number of episodes	10.16	(10.78)	10.25	(11.1)	t(556.4) = -0.1	0.917	0.009
Predominant polarity (% Depressive/ Intermediate/Manic)	32.3 / 49.8 / 17.9		29.6 / 50.3 / 20.1		X² (2) = 0.69	0.707	0.035
Depression (MADRS: 0-60)	3.8	(3.1)	4.21	(3.25)	t (837.6) = -1.85	0.064	0.127
Mania (YMRS: 0-60)	1.84	(2.85)	1.75	(2.77)	t (815.1) = 0.45	0.652	0.031
Severity (CGI: 0-7)	2.48	(1.38)	2.57	(1.33)	t (795.1) = -0.9	0.368	0.063
Current treatment (% Yes)							
Antidepressants	17.3		30.2		X² (1) = 16.91	**<0.001**	0.138
Anticonvulsants	29.3		40.5		X² (1) = 10.03	**0.002**	0.106
Lithium	23.6		24.7		X² (1) = 0.09	0.765	0.01
Antipsychotics	14.9		31.4		X² (1) = 28.53	**<0.001**	0.179
Anxiolytics	14.9		21.4		X² (1) = 4.99	**0.025**	0.075
Cognitive measures (z score)							
Verbal memory	-0.11	(0.95)	-0.18	(0.93)	t (777.2) = 1.11	0.269	0.079
Working memory	-0.13	(0.7)	-0.21	(0.71)	t (791.8) = 1.54	0.124	0.109
Executive functions	-0.17	(0.77)	-0.18	(0.81)	t (795.7) = 0.34	0.738	0.024
Processing speed	-0.01	(0.77)	-0.12	(0.81)	t (798.1) = 1.9	0.057	0.134
Attention	-0.25	(0.76)	-0.32	(0.83)	t (610.9) = 1.08	0.283	0.086
Verbal and perceptual reasoning	0.26	(0.8)	0.12	(0.87)	t (788.8) = 2.23	**0.026**	0.157

Significant p values are reported in bold. MADRS: Montgomery-Asberg Depression Rating Scale; YMRS: Young Mania Rating Scale; FAST: Functional Assessment Short Test.

Effects sizes (V: Cramer's V; d: Cohen's d) are reported in absolute values.

Completers are patients who completed both neuropsychological assessments. Non-completers are patients who did not complete all three visits. The completers group is not similar to the final included group, as it does account for exclusion criteria at T12 and T24 (euthymia and missing data cut-off).

Supplementary Table 2. Comparison of the variables across time (linear mixed models)

	Mean (sd)						Comparisons		Effect size
	T0		T12		T24		Test	p	d
Verbal memory	-0.11	(0.94)			0.23	(0.87)	F (1,205) = 39.38	**< 0.001**	0.45
Working memory	-0.11	(0.66)			0.09	(0.68)	F (1,205) = 18.4	**< 0.001**	0.28
Executive functions	-0.19	(0.75)			-0.01	(0.69)	F (1,203) = 17.38	**< 0.001**	0.3
Processing speed	-0.01	(0.78)			0.22	(0.74)	F (1,201) = 45.78	**< 0.001**	0.49
Attention	-0.21	(0.76)			-0.12	(0.85)	F (1,174) = 2.91	0.09	0.15
Reasoning	0.29	(0.77)			0.38	(0.79)	F (1,194) = 4.88	**0.03**	0.16
MADRS	3.31	(3.01)	2.31	(2.81)	2.37	(2.85)	F (2, 497) = 12.88	**< 0.001**	
FAST - Autonomy	0.91	(1.61)	0.46	(1.13)	0.43	(1.12)	F (2, 490) = 13.63	**< 0.001**	
FAST - Work	5.36	(6.04)	3.87	(5.65)	3.93	(5.74)	F (2, 492) = 12.6	**< 0.001**	
FAST - Cognition	1.98	(2.55)	1.21	(1.81)	1.23	(1.74)	F (2, 488) = 16.9	**< 0.001**	
FAST - Finances	0.38	(0.98)	0.23	(0.69)	0.22	(0.72)	F (2, 499) = 3.98	**0.02**	
FAST - Interpersonal	2.84	(3.05)	2.11	(2.74)	1.98	(2.4)	F (2, 490) = 12.06	**< 0.001**	
FAST - Leisure	1.16	(1.47)	1.04	(1.47)	0.88	(1.2)	F (2, 493) = 4.41	**0.01**	

Bonferroni post-hoc pairwise comparisons (performed for variables on which the main effect of visit was significant)

	T0-T12		T12-T24		T0-T24	
	p-value	Cohen's d	p-value	Cohen's d	p-value	Cohen's d
MADRS	**< 0.001**	0.34	1	0.02	**< 0.001**	0.32
FAST - Autonomy	**< 0.001**	0.32	1	0.03	**< 0.001**	0.35
FAST - Work	**0.01**	0.26	1	0.01	**0.01**	0.24
FAST - Cognition	**< 0.001**	0.34	1	0.02	**< 0.001**	0.34
FAST - Finances	0.12	0.18	1	0.01	0.08	0.18
FAST - Interpersonal	**0.01**	0.25	0.54	0.05	**< 0.001**	0.31
FAST - Leisure	1	0.08	0.59	0.12	0.06	0.21

p-values of significant differences are written in bold.

Supplementary Table 3. Zero-order correlation (Pearson's) matrix between variables included in the models

Cognition: verbal memory (VM), working memory (WM), executive functions (EF), processing speed (PS), attention (A), and verbal and perceptual reasoning (VPR). Functioning (FAST sub-scores): F-A autonomy, F-W work, F-C cognition, F-IR interpersonal relationships, and F-L leisure time.

Significant correlations are indicated bold, significance level: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Cognition T0						MADRS1	FAST T0						FAST T12		
	VM1	WM1	EF1	PS1	A1	VPR1	MADRS1	F-A1	F-W1	F-C1	F-F1	F-IR1	F-L1	MADRS2	F-A2	F-W2
WM1	0.17**															
EF1	0.22***	0.33***														
PS1	0.19**	0.27***	0.63***													
A1	-0.02	0.24***	0.15*	0.15*												
VPR1	0.3***	0.47***	0.34***	0.34***	0.12											
MADRS1	-0.05	0	0.02	-0.01	0.06	0.08										
F-A1	-0.05	0.05	0.08	0.05	0.06	0.09	0.37***									
F-W1	-0.07	0.04	-0.03	-0.05	-0.04	-0.01	0.2**	0.3***								
F-C1	-0.2**	-0.06	-0.04	-0.03	0.04	-0.01	0.29***	0.28***	0.26***							
F-F1	0.03	-0.08	0.07	0.1	-0.06	-0.03	0.1	0.3***	0.19**	0.09						
F-IR1	-0.15*	-0.02	-0.02	0.11	0.06	0.1	0.36***	0.42***	0.23***	0.4***	0.16**					
F-L1	-0.05	0.03	-0.01	-0.05	0.09	0.11	0.27***	0.35***	0.24***	0.16**	0.18**	0.41***				
MADRS2	-0.01	-0.02	0.15*	0.09	0.07	0.08	0.27***	0.16*	0.13	0.28***	0.07	0.23***	0.16*			
F-A2	-0.14*	-0.03	-0.09	-0.17*	0.01	-0.03	0.05	0.09	0.14*	0.09	-0.01	0.09	0.15*	0.24***		
F-W2	-0.13	0.05	-0.04	-0.12	-0.03	-0.12	0.15*	0.2**	0.57***	0.34***	0.18**	0.18**	0.21**	0.23***	0.2**	
F-C2	-0.15*	-0.16*	-0.17*	-0.22***	-0.12	-0.16*	0.08	0.05	0.04	0.34***	-0.03	0.13	0.08	0.27***	0.27***	0.29***
F-F2	-0.03	-0.1	-0.03	-0.04	-0.08	0	0.1	0.18**	0.23***	0.2**	0.37***	0.26***	0.12	0.13	0.11	0.34***
F-IR2	-0.08	-0.05	-0.08	-0.08	-0.06	-0.05	0.19**	0.13	0.15*	0.28***	-0.01	0.47***	0.12	0.35***	0.43***	0.28***
F-L2	-0.03	-0.03	-0.09	-0.15*	-0.01	-0.04	-0.05	0.12	0.16*	0.08	0.01	0.19**	0.3***	0.23***	0.3***	0.27***
VM3	0.67***	0.05	0.19**	0.23**	-0.01	0.15*	0	-0.1	-0.08	-0.09	-0.05	-0.11	-0.1	0.08	-0.11	-0.12
WM3	0.09	0.57***	0.23**	0.25***	0.2*	0.42***	0.07	0.11	-0.01	-0.12	-0.07	0.04	-0.01	-0.01	0.01	-0.04
EF3	0.17*	0.27***	0.67***	0.45***	0.05	0.3***	0.01	0.11	-0.11	-0.14	0.05	-0.02	-0.06	0.02	-0.08	-0.12
PS3	0.22**	0.23**	0.46***	0.77***	0.1	0.29***	0.05	0.12	-0.17*	-0.03	0.06	0.09	-0.07	0.08	-0.18*	-0.12
A3	-0.04	0.13	0.12	0.13	0.58***	0.09	-0.11	0.1	-0.12	-0.02	0	0.06	-0.03	-0.13	-0.04	-0.12
VPR3	0.29***	0.45***	0.31***	0.29***	0.11	0.72***	0.14	0.15*	-0.01	0.02	0	0.2**	0.06	0.11	0	-0.08
MADRS3	-0.03	-0.04	0	-0.01	0.04	-0.02	0.28***	0.1	-0.01	0.17**	-0.01	0.2**	0.08	0.33***	0.16*	0.15*
F-A3	-0.02	-0.07	-0.04	-0.11	0.1	0.01	0.2**	0.24***	0.12	0.12	0.08	0.12	0.13*	0.27***	0.26***	0.06
F-W3	-0.13*	-0.01	-0.07	-0.12	-0.03	-0.14*	0.15*	0.15*	0.45***	0.26***	0.07	0.15*	0.21***	0.26***	0.19**	0.67***
F-C3	-0.13*	-0.2**	-0.09	-0.02	-0.02	-0.12	0.13*	0.04	0.09	0.39***	0.08	0.17**	0.06	0.17*	0.08	0.13
F-F3	-0.05	-0.03	-0.03	0.01	0.13	0.02	0.05	0.01	0.01	0.09	0.16*	0.09	0.14*	0.05	0.08	0.11
F-IR3	-0.05	-0.06	-0.04	0.01	-0.01	0.03	0.17**	0.09	0.05	0.16*	0.02	0.36***	0.08	0.24***	0.16*	0.12
F-L3	-0.11	0.03	-0.1	-0.04	-0.02	0.07	0.1	0.12	0.1	0.04	0	0.2**	0.24***	0.18**	0.23***	0.16*

	FAST T12				Cognition T24							FAST T24				
	F-C2	F-F2	F-IR2	F-L2	VM3	WM3	EF3	PS3	A3	VPR3	MADRS3	F-A3	F-W3	F-C3	F-F3	F-IR3
F-F2	0.12															
F-IR2	**0.32***	**0.21**														
F-L2	**0.23***	0.12	**0.44***													
VM3	**-0.18***	-0.09	-0.1	-0.07												
WM3	-0.09	-0.09	-0.02	-0.04	0.1											
EF3	**-0.17***	-0.11	-0.14	**-0.21**	**0.21**	**0.32***										
PS3	**-0.18***	**-0.17***	-0.13	**-0.22**	**0.27***	**0.3***	**0.55***									
A3	-0.07	**-0.2***	-0.03	0.06	-0.01	**0.16***	0.05	0.15								
VPR3	**-0.16***	-0.09	-0.04	-0.05	**0.25***	**0.43***	**0.42***	**0.31***	0.04							
MADRS3	**0.17***	0.1	**0.26***	0.06	**-0.18***	-0.09	**-0.16***	-0.08	**-0.17***	-0.1						
F-A3	0.07	0.1	0.12	**0.18**	**-0.22**	-0.07	-0.13	**-0.19**	-0.06	0	**0.26***					
F-W3	**0.25***	**0.22**	**0.24***	**0.27***	**-0.2**	**-0.16***	**-0.16***	**-0.15***	**-0.19***	**-0.17***	**0.22***	**0.2**				
F-C3	**0.35***	0.1	**0.23***	**0.16***	**-0.25***	**-0.19***	**-0.18***	-0.12	-0.13	**-0.21**	**0.33***	**0.28***	**0.27***			
F-F3	0.03	**0.23***	0.06	0.04	**-0.22**	0.02	-0.04	-0.03	-0.04	-0.05	**0.15***	**0.18**	0.1	0.11		
F-IR3	**0.19**	0.12	**0.56***	**0.2**	-0.1	-0.01	-0.05	-0.01	-0.14	-0.01	**0.31***	**0.26***	**0.17**	**0.34***	0.06	
F-L3	0.1	0.09	**0.21**	**0.44***	-0.13	-0.01	**-0.15***	-0.09	0.07	0.04	**0.16**	**0.37***	**0.17**	**0.23***	0.07	**0.33***

Supplementary Table 4. Path statistics for the retained model

Path	Non-standardized coefficients	Standard error	z-value	p-value	Standardized coefficients
Regressions					
Functioning T0 → Functioning T12	0.57	0.19	2.95	**0.003**	0.44
Functioning T12 → Functioning T24	0.57	0.25	2.27	**0.023**	0.57
Cognition T0 → Cognition T24	2	0.47	4.29	**< 0.001**	0.89
MADRS T0 → Cognition T0	-0.02	0.03	-0.62	0.54	-0.05
MADRS T24 → Cognition T24	0.04	0.05	0.97	0.33	0.06
MADRS T0 → Functioning T0	-0.21	0.04	-5.89	**< 0.001**	-0.54
MADRS T12 → Functioning T12	-0.24	0.06	-4.26	**< 0.001**	-0.44
MADRS T24 → Functioning T24	-0.21	0.05	-3.89	**< 0.001**	-0.39
Cognition T0 → Functioning T12	0.57	0.16	3.68	**< 0.001**	0.37
Cognition T0 → Functioning T24	0.05	0.14	0.35	0.73	0.03
Covariances					
Functioning T0 ←→ Cognition T0	0.01	0.11	0.12	0.91	0.01
Functioning T24 ←→ Cognition T24	0.48	0.22	2.23	**0.026**	0.48