RUNNING TITLE: Executive Functioning in Bipolar Disorder

Similar Trajectory of Executive Functioning Performance over 5 years among individuals with Bipolar Disorder and Unaffected Controls using Latent Growth Modeling

Kelly A. Ryan, PhD[1], Shervin Assari, MD, MPH[1], Bethany D. Pester, BA[1], Kristin Hinrichs, PhD[1], Kaley Angers, BS[1], Amanda Baker, BS[1], David F. Marshall, PhD[1], Deborah Stringer, PhD[1], Erika F.H. Saunders, MD[2], Masoud Kamali, MD[1,4], Melvin G. McInnis, MD[1], & Scott A. Langenecker, PhD[2]

[1]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan

[2] Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA

[3] Department of Psychiatry, Penn State College of Medicine, Hershey, PA, USA

[4]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

Corresponding Author: Kelly Ryan
2101 Commonwealth Blvd, Suite C
Ann Arbor, MI 48105
Email: karyan@umich.edu
Phone: 734-936-5524
Fax: 734-936-9262

Abstract

Objective: Executive Functioning (EF) deficits in bipolar disorder (BD) are commonly present regardless of mood state and therefore are considered core features of the illness. However, very little is known about the temporal stability of these deficits. We examined the natural course of EF over a five year period in BD and healthy control (HC) samples.

Method: Using a 5-year longitudinal cohort, 91 individuals with BD and 17 HC were administered a battery of neuropsychological tests that captured four main areas of EF: Processing Speed with Interference Resolution, Verbal Fluency with Processing Speed, Inhibitory Control, and Conceptual Reasoning and Set Shifting. Evaluations occurred at study entry, one, and five years later.

Results: Latent Growth Curve Modeling demonstrated that the BD group performed significantly worse in all EF areas than the HC group. Changes in EF from baseline to 5-year follow-up were similar across both diagnostic groups. Older age at baseline, above and beyond education and diagnosis, was associated with worse initial performance in EF. Being of older age was associated with greater decline in Processing Speed with Interference Resolution, and Verbal Fluency with Processing Speed. Higher education was marginally associated with a smaller declining slope for Processing Speed with Interference Resolution.

Conclusions: Executive functioning deficits in BD persist over time, and in the context of normative age-related decline, may place individuals at greater risk for cognitive disability as the disease progresses. Age and having a BD diagnosis together, however, do not accelerate executive functioning decline over time.

Keywords: Bipolar Disorder, neurodegenerative, Cognition, Affective Disorders

Introduction

Cognitive impairments in patients with bipolar disorder (BD) are common and most notable in the areas of attention, verbal learning, emotion perception, and executive functioning [1-5]. Impairments are seen during acute mood episodes, such as depressive, manic, and hypomanic episodes [6], but are also apparent to some degree during clinical remission [6-9]. Executive functioning, defined as one's ability to shift thinking, organize, plan, and solve problems, is a broad area of cognitive dysfunction that seems to persist throughout clinical mood states [9, 10]. However, little is known regarding the temporal stability of executive functioning performance over time/course of illness. Overall, robust findings indicate that executive functioning deficits are chronic and persistent features of BD, and therefore, are promising candidates for intermediate phenotypes [1, 9]. Further, executive functioning deficits have significant impact on overall functional impairment and disability as well as quality of life [3, 11-15]. These findings, however, raise questions regarding the natural course of cognitive functioning in patients with BD.

Despite the range of cognitive dysfunction found in BD and the clear association to functional impairment [2, 16, 17], few studies have examined the longitudinal course of cognitive functioning or the stability of impairments. In the few studies that have examined this relationship, mixed findings have been reported [3, 13, 14, 18]. Some studies have reported no change/progression of cognitive impairments over time [19-22], suggesting a stable, possibly neurodevelopmental, process in BD. In contrast, others have found significant change over time [18], supporting a neurodegenerative model in BD.

Existing longitudinal studies have major methodological limitations that hinder inferences based on their findings. First, most studies include short follow-up periods (3 months to 3 years) [13, 19, 22-27] that may be insufficient in detecting changes in cognitive functioning.

Small sample sizes, ranging from 15 to 30 individuals with BD [22, 28-30], have also been reported. Further, few studies have included a normative comparison group [22, 31] or more than two testing time points [22, 23, 25, 27, 29, 30], which may limit the ability to model a more realistic longitudinal trajectory [32]. Importantly, several studies include brief or narrow measurements of cognition [33, 34], which likely limit the chance of capturing meaningful cognitive changes across the spectrum of neuropsychological test domains. Lastly, traditional statistical approaches, such as multivariate techniques and repeated measures, have been suboptimal and can negatively affect statistical power. Limitations in these statistical approaches include the inability to account for missing data, an inherent issue in longitudinal studies due to attrition, lack of methods to evaluate the possible non-linear relationship of cognitive change, and inability to correlate cognitive trajectories with each other. Therefore, it is critical to investigate whether BD is associated with greater decline in executive functioning compared to a healthy comparison group using longitudinal studies that utilize robust longitudinal analytic methods and include a much larger cohort of individuals with BD who have been followed for multiple years.

Using participants enrolled in the Prechter Longitudinal study of BD [9, 12, 35], we examined the stability of executive functioning performance in a large sample of patients with BD and compared findings to healthy, unaffected controls (HC) across three time points spanning five years. We have strengthened the analytic approach to this pressing question by utilizing Latent Growth Curve Modeling (LGCM), a statistical approach that offers many advantages over traditional methods for longitudinal analyses [32]. Consistent with prior research, we hypothesized that executive functioning performance would be significantly worse in the BD group as compared to the HC at baseline. We further hypothesized that having BD will

negatively be related to the trajectory of executive functioning and, specifically, that individuals with BD will show greater executive functioning decline over time than HCs, supporting an age accelerated decline in functioning in the context of age and disease [36]. As our hypotheses focus on examining the trajectory of executive functioning between a sample of BD and HC, our covariates were limited to variables that are common between individuals with BD and HC (e.g., age, gender, and education) but not clinical variables that represent severity of BD, such as years with illness, type of BD, rapid cycling, etc.

Patients and Methods

Subjects

Participants were enrolled in the Prechter Longitudinal Study of BD, an observational cohort study gathering phenotypic and biological data, at the University of Michigan. This study was approved by the University of Michigan IRB and all participants provided signed informed consents. Participants in this study were enrolled from 2006-2008 and had confirmed diagnosis of BD or were HCs. Out of the total 264 individuals who had been enrolled long enough that a five year follow-up visit could have been completed, 108 completed five year neuropsychological re-testing and entered to the main analyses of this study. This number includes 91 individuals with BD (80 BD I, 9 BD II, 2 BD-NOS) and 17 HC. Eight HC were excluded from analyses as they changed diagnosis over the interim (4 received a new diagnosis of Major Depressive Disorder, 1 Depression NOS, 1 PTSD, 1 BD type II, and 1 BD type I), although there were no significant differences from baseline to Year 5 test scores. Comparison of BD and HC who completed year 5 testing versus those who did not (n=116; 52.2%) are

presented in the results section. In general, there were no differences based on study enrollment. For those included in the main analyses, only 4 individuals did not complete year 1 testing.

Recruitment of participants into the longitudinal study occurred through advertisements on the web, in the newspaper, in an outpatient specialty psychiatric clinic, community mental health centers, community outreach events and in an inpatient psychiatric unit. At study entry, all participants underwent an evaluation using the Diagnostic Interview for Genetic Studies [37] and a best estimate process by at least two of the authors was used to confirm diagnoses. During the best estimate process, two authors reviewed the DIGS information including the psychiatric summary written by the interviewer and medical records, and they came to a consensus agreement on the most appropriate diagnosis. Participants were excluded if they had active substance use at enrollment or neurological disease. Written informed consent was obtained and incentive payment for their participation in the larger longitudinal study was provided. Mood symptoms were obtained via clinician-administered measures (Hamilton Rating Scale for Depression, HRDS [38], Young Mania Rating Scale, YMRS [39]) at study entry and then again at years 1 and 5 during neuropsychological testing (by trained research associates and supervised by one of the study clinicians). Our sample included those with a range of mood scores (Table 1). Clinical variables were collected from the DIGS interview by the clinician and are listed in Supplemental Table 1 (under “completers”). Each participant’s medication using methods adapted from other groups [40-43] allowed us to create medication class and composite load scores whereby higher scores represented higher medication burden. Again, as our study focused on the effect of BD diagnosis (compared to HC) on changes of executive functioning over time, clinical variables could not be included in the main LGCM analyses. However, we have included clinical variables to describe our BD sample.

Neuropsychological Assessment

Neuropsychological tests were administered by trained research associates under the training and supervision of licensed clinicians (KAR, SAL, DM) at study entry, and again at years 1 and 5. Similar to our other published work [9, 35], a battery of neuropsychological tasks was administered. The tasks focusing on executive functioning include: the Wisconsin Card Sorting Test [44], the Stroop Color and Word Test [45], the Controlled Oral Word Association Test (FAS task) and Animal Fluency [46], Digit Symbol Coding from the Wechsler Adult Intelligence Scale-III [47], Trail Making Test-Parts A and B [48], and the Parametric Go/No-Go task [49]. We used data reduction methods (a factor analytic procedure) on these executive functioning tests, consistent with our previous work [35, 50], and factor scores derived from that process were used to compute four executive functioning scores using principle axis FA with oblique rotation. These include Processing Speed with Interference Resolution (PSIR), Verbal Fluency with Processing Speed (VFPS), Inhibitory Control (IC), and Conceptual Reasoning and Set Shifting (CRSS). Supplemental Table 2 shows which cognitive subtests scores were included in each factor, including reliability of the factor scores. Estimated verbal intelligence was obtained using the Wechsler Abbreviated Scale of Intelligence vocabulary subtest[51].

Data analysis

IBM SPSS 22 was used for univariate and bivariate analyses. Pearson (or Spearman for non-parametric variables) correlation test or t tests were used for bivariate analysis. We used AMOS [52] to run LGCM, a type of structural equation modeling [53]. We ran a separate LGCM for each executive functioning outcome to test our hypotheses. In the first step, we ran

unconditional LGCMs which suggested intercepts and slopes for all outcomes (see Supplemental Table 3). Due to low sample size, we could not test models with quadratic slope to estimate non-linear slopes, as none of the models with non-linear slopes converged (not enough degree of freedom). In the next step, we tested conditional LGCMs, with BD as the main predictor of interest and age, gender, and education as covariates. Paths were drawn from BD as well as other covariates to intercept, as well as linear slope. As our model was fitted to the pooled sample, with BD as the primary predictor of interest, we could only include covariates that were common between our groups, so we could not include BD specific variables such as type of illness, medications, age of onset, or any measure of severity such as rapid cycling. Outcomes included intercept and linear slope. Fit statistics included the comparative fit index (CFI) larger than 0.90, the root mean squared error of approximation (RMSEA) less than 0.08, and the Chi square to degrees of freedom ratio less than 4.0 [54-58]. Unstandardized and standardized regression coefficients were reported for each path. Full information maximum likelihood (FIML) was applied to handle missing data.

Results

Descriptive statistics

There was a significant group difference for education, t(105)=2.05, p=0.04 and overall with BD having lower education than HC. There were no significant differences between groups in terms of age, t(105)=-1.94, p=0.06, gender, ^{2} (1, N=107)=0.031, p=0.860, and general verbal intelligence (WASI Vocabulary scaled score; t(105)=.909, p=.365 (See Table 1). Table 1 also shows group comparisons with significance values for each of the four executive functioning domains by each of the three time points. The BD group generally underperformed compared to

HC in Verbal Fluency with Processing Speed, Processing Speed with Interference Resolution, and Inhibitory Control (Year 1).

We compared participants who completed 5 year re-testing (completers; BD: n=91, HC: n=17) to those in the entire sample of individuals who had been enrolled long enough that five year testing could have been completed (non-completers; BD: n=91, HC: n=25) to contextualize this sample in terms of attrition (Supplemental Table 1). The completers were of older age in comparison to the non-completers (M(SD)=41.0(12.2) vs. M(SD)=35.24(13.1); t(221)=3.41, *p*=.001). There was no difference between groups in terms of education, *p*=.301 or verbal intellect, *p*=.589. There was a significant difference between completers and non-completers in terms of gender, $\chi^2(1)=15.82$, p=.00007 (completers: 75.5% female, non-completers: 49.6% female). Among just the BD sample, there were no significant differences between completers and non-completers for any clinical variables (*p*>.05). Overall, we did not find evidence suggesting that clinical variables account for attrition in our BD sample who completed 5 year testing. Further, there was no difference in completers and non-completers in terms of employment status.

Bivariate analysis

Table 2a. shows the intercorrelations between the executive functioning factor scores across the three time points for all participants. Based on the oblique factor scores being related to the same cognitive construct (executive functioning), baseline VFPS, CRSS, PSIR, and IC scores were positively correlated with each other, with coefficients ranging from .33 to .78. Intercorrelations were stronger between each specific factor score across the three time points. Baseline VFPS was significantly and strongly correlated with year 1 (r=.83, *p* <.001) and year 5

(r=.80, *p* <.001), suggesting little change over time. Baseline PSIR was significantly and strongly correlated with year 1 (r=.83, *p* <.001) and year 5 (r=.85, *p* <.001). IC also showed a significant correlation with year 1 (r=.47, *p*<.02) and year 5 (r=.52, *p* <.001). Baseline CRSS showed a weaker, albeit still significant correlation with year 1 (r=.48, *p* <.05) but not with year 5 (r=.18, *p*=.11)indicating these this variable was less stable over time.

Age was consistently negatively associated with PSIR at baseline, year 1, and year 5, VFPS at years 1 and 5, and IC at years 1 and 5. Education was only significantly correlated with VFPS at baseline, year 1, year 5 (Table 2a). As this study focuses on the cognitive stability based on diagnosis rather than specific clinical features of BD, we present the correlations between the executive functioning scores, across all three time points, and the clinical features among just the BD sample features in Supplemental Table 4. None of the clinical variables showed a consistent significant association with the executive functioning factor scores, with the exception of chronicity of mood symptoms being significantly negatively correlated with VFPS and PSIR at year 5 (r's=-.27 & -.27, respectively), rapid cycling was positively related to CRSS at baseline (r=.27), and impact of BD illness was negatively related to VFPS at baseline, year 1 and year 5 and with PSIR at year 5 (r's ranged -.23 to -.28, see Table 3).. Depression and YMRS scores (Table 2b) did not sow significant associations with executive functioning scores across all test sessions with the exception of HRDS scores at year 1 was negatively related to IC at year 5. Further, there were a few sporadic correlations between age at onset of specific mood episode (mania, hypomania, and depression episodes), number of mood episodes, and the cognitive scores, but no consistent pattern.

Latent Growth Curve Modeling

Fit of the *PSIR Model* was very good [CFI= .947, Chi-square = 22.776, df = 7, Chi-square /df=3.254, Probability level .002]. Based on this model, diagnosis (BD vs. HC) and age were negatively associated with baseline PSIR scores, suggesting that individuals who had a higher age at baseline had a lower PSIR score at baseline. Age was negatively associated with the slope of the PSIR over the 5 year follow-up, suggesting that having higher age at baseline was also associated with lower PSIR change during the five year period. Education was also associated with PSIR change over time (Table 3 and Figure 1-a).

VFPS Model also showed very good fit [CFI= .999, Chi-square = 8.270, df = 8, Chi-square /df=1.034, Probability level = .408]. As table 3 and Figure 1-b suggest, BD and age were negatively associated with intercept of the VFPS, suggesting lower level of VFPS at baseline among older individuals and those with BD. Education was positively associated with baseline VFPS, suggesting individuals with higher education perform better at baseline for VFPS. Age was negatively associated with the VFPS change over time, suggesting that having higher age at baseline is associated with lower VFPS change over time.

Fit of the *CRSS Model* was also very good [CFI= .983, Chi-square = 22.776, df = 7, Chi-square /df=3.254, Probability level = .002]. According to the results of *Model III*, age was negatively associated with baseline CRSS, suggesting that higher age at baseline is associated with a lower CRSS score at baseline. (Table 3 and Figure 1-c).

Fit of the *IC Model* was also very good [CFI= .925, Chi-square = 14.855, df= 8, Chi-square /df=1.857, Probability level = .062]. As table 3 and Figure 1-d suggest, age was negatively associated with baseline IC scores, with a lower baseline IC level for individuals with higher age. Gender was also negatively associated with change in IC, suggesting smaller IC change over time for females.

Discussion

Our findings show that individuals with BD perform worse than HC on tasks of executive functioning at baseline, consistent with prior literature [1, 2]. Specifically within the complex domain of executive functioning, individuals with BD underperformed compared to the HC across the five years on Verbal Fluency with Processing Speed (VFPS) and Processing Speed with Interference Resolution (PSIR), both executive functioning tasks that include a speeded component. There were no diagnostic group differences in Conceptual Reasoning and Set-Shifting performance (CRSS; except at year 1), and there were no consistent diagnostic group differences for Inhibitory Control (IC). The linear change in performance of individuals with BD across five years is no different from the change in EF of healthy controls, which is against accelerate deterioration associated with BD. decrements in performances among the BD group are seen across five years, indicating that executive functioning deficits are persistent.

This study was focused on comparing BD vs HC for trajectories of executive functioning using LGCM, and as such, we could only control for potential demographic and socio-economic confounders unique to both groups. In our study, clinical variables (e.g., illness severity, hospitalizations) could not confound the effect of BD on cognitive trajectory. Contrary to expectations based on prior literature and our hypothesis, having BD did not negatively relate to the trajectory of executive functioning performance over five years. There was no difference in the linear decline of executive functioning scores over time between the BD and HC samples, suggesting that the longitudinal course of executive functioning over five years is not dependent on having a BD diagnosis and that the rate of change in BD is equivalent to healthy, psychiatrically unaffected individuals. Further, these results suggest that the cognitive deficits in

BD are not age accelerated, age-compounded, or neurodegenerative, extending earlier work using smaller samples [19-22, 59].

Our study makes a contribution to the existing literature, which is mostly cross-sectional, uses small sample sizes if using a longitudinal design, includes short and infrequent follow-up periods, and utilizes a broad view of cognition and statistical methods that do not allow for the possible non-linear relationship of cognitive change. We included a healthy, unaffected sample of adults who were similar in education and gender, and a uniquely large sample of individuals diagnosed with BD who were followed for 5 years at three time points and healthy controls who did not demonstrate any change in their healthy status over time or have any other conditions. We uniquely focused our analyses on specific aspects of executive functioning that have representatively shown to be affected by the illness.

There is a significant effect of age on executive functioning performance. Despite expectations of potential disease compounding age effects, having BD does not affect the trajectory of executive functioning performance over time, and clinical variables did not show a consistent significant association with the executive functioning factor scores. Those who are older, regardless of diagnosis, performed worse across all areas of executive functioning at baseline. Further, those who are older, irrespective of diagnosis, show a larger decline over time on executive tasks that have a speeded component (VFPS and PSIR), consistent with the literature that indicates that normal aging is associated with changes in processing speed [60]. Also consistent with literature on normal aging and the protective effects of education on cognitive functioning [61], those who have more years of education had less decline in some areas of executive functioning (significant effect for VFPS and CRSS and a marginally

significant effect for IC). Females showed greater linear decline in IC compared to males, indicating one area where there may be some unknown gender effect.

Our study has addressed many limitations found in prior longitudinal studies. However, there are still limitations that reduce the generalizability of the results and warrant follow-up investigation. As is common in naturalistic studies of this kind, we could not control for medication use and the impact that medications may have on the cognitive trajectories [62]. It is possible that medication usage impacts diagnostic group differences on executive tasks; however, given no difference in cognitive trajectory, we suspect that there may be no relationship to change in functioning. We focused on specific aspects of executive functioning because of the large prevalence of such deficiencies in BD and the identification that these may be intermediate endophenotypes of the disorder [1, 9] and good markers for neuroprogression. However, other aspects of cognition, such as memory, may be more appropriate to look at to determine disease-related progressive change as memory has been shown to be the only cognitive area subject to change over time in one longitudinal study [63]. Further, due to sample size and exclusion of 8 controls who had a chance in psychiatric diagnosis over time, we could not include quadratic slope in addition to linear slope. Additional sample size via our longitudinal study in the future will enable us to model non-linear slope. We plan to expand these results to look at trajectories of performance in memory, fine motor speed, and emotion processing. Further, as this study focused on trajectories of executive functioning using Latent Growth Curve analysis focusing on the influence of merely having the BD diagnosis or not, our future work will examine the specific relationships between clinical indices and cognitive performance change in just our BD sample to examine if specific scar or illness burden factors are related to cognitive trajectories. We could not include any covariate that was specific to BD in our models, which was a

consequence of our aim to compare BD and HC. Future research needs to explore if any bipolar-specific variables (e.g. type of illness, medications) predict trajectories of cognitive function within individuals with BD. Lastly, it is unclear to what degree our sample, who maintained enrollment in the longitudinal study for five years, differs from those who were lost to follow-up or chose not to continue with the study. We attempted to address this potential concern (differential attrition) by comparing those who completed re-testing after being in the study for five years to those who were eligible for re-testing but did not complete it. These two samples were not different in illness-related variables (for those with BD), education, or employment status, but they were different based on age and gender (completers were older, more likely to be female, and have had the illness longer).

In summary, our findings suggest that there are persistent executive functioning deficits in BD, in the context of similar age related declines, which supports the theory that this specific cognitive area is a trait marker for the disease - a cognitive intermediate phenotype. However, having BD does not appear to be a risk factor for neurodegradation or age-compounding effects upon executive functioning skills, as our sample of BD experienced the same slopes of cognitive changes over time as did our healthy control sample. Declines in executive functioning performance, regardless of a BD diagnosis, seem related to increasing age and lower education. These findings may have significant implications on how the illness is managed with findings suggesting that BD may be a chronic, but relapsing-remitting psychiatric illness rather than a neuroprogressive illness in terms of cognitive functioning. However, further longer term follow up is needed to determine to rule out the neuroprogression hypothesis.

Acknowledgements:

This research was supported by the Heinz C. Prechter Bipolar Research Fund at the University of Michigan Depression Center and the Richard Tam Foundation (KR, SA, BP, KH, DM, AB, KA, DS, MK, MM), and grant support from the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number 2KL2TR000434 (KR). We would like to express appreciation to our research participants in the Prechter Longitudinal Study of Bipolar Disorder. We would also like to acknowledge and thank our research team consisting of Holli Bertram, Christine Brucksch, Brent Doil, Valerie Foster, Laura Gabriel, Nicole Greer, Lauren Grove, Brennan Haase, Gloria Harrington, Alexander Hayek, Michelle Kassel, Marisa Kelly, Katie Lavin, Kortni Meyers, Jennifer Montgomery, Lisa O'Donnell, Philip Presnell, Anne Weldon, and the rest of the staff of the Prechter Bipolar Research Team for their contributions to this project.

Declaration of Interest:
Dr. Ryan, Ms. Angers, Ms. Pester, Ms. Baker, Dr. Marshall, and Dr. Hinrichs report no competing interests.

Dr. Kamali has received research grant support from Janssen Pharmaceutical and Assurex Health. Dr. Saunders has been a consultant for Projects In Knowledge. Dr. Langenecker has served as a consultant for Cogstate, Ltd and Easter Seals, Inc, in work unrelated to the present work. Dr. McInnis has affiliations with Janssen Pharmaceuticals.

References

1. Bora, E., M. Yucel, and C. Pantelis, *Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives*. Journal of Affective Disorders, 2009. **113**(1–2): p. 1-20.
2. Mann-Wrobel, M.C., J.T. Carreno, and D. Dickinson, *Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: An update and investigation of moderator variables*. Bipolar Disorders, 2011. **13**(4): p. 334-342.
3. Robinson, L.J. and I.N. Ferrier, *Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence.* Bipolar Disord, 2006. **8**(2): p. 103-16.
4. Quraishi, S. and S. Frangou, *Neuropsychology of bipolar disorder: a review.* Journal of Affective Disorders, 2002. **72**(3): p. 209-226.
5. Kohler, C.G., et al., *Facial emotion perception in depression and bipolar disorder: a quantitative review.* Psychiatry Res, 2011. **188**(3): p. 303-9.
6. Martínez-Arán, A., et al., *Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder.* The American Journal of Psychiatry, 2004. **161**(2): p. 262-70.
7. Clark, L., et al., *Sustained attention-deficit confirmed in euthymic bipolar disorder but not in first-degree relatives of bipolar patients or euthymic unipolar depression.* Biological Psychiatry, 2005. **57**(2): p. 183-187.
8. El-Badri, S.M., et al., *Electrophysiological and cognitive function in young euthymic patients with bipolar affective disorder.* Bipolar Disorders, 2001. **3**(2): p. 79-87.
9. Ryan, K.A., et al., *Differential executive functioning performance by phase of bipolar disorder.* Bipolar Disorders, 2012. **14**(5): p. 527-536.
10. Dixon, T., et al., *Effect of symptoms on executive function in bipolar illness.* Psychological Medicine, 2004. **34**(5): p. 811-21.
11. SANCHEZ-MORENO, et al., *Neurocognitive dysfunctions in euthymic bipolar patients with and without prior history of alcohol use*. Vol. 70. 2009, Memphis, TN, ETATS-UNIS: Physicians Postgraduate Press. 8.
12. Ryan, K.A., et al., *Emotion perception and executive functioning predict work status in euthymic bipolar disorder.* Psychiatry Research, 2013.
13. Mur, M., et al., *Long-term stability of cognitive impairment in bipolar disorder: a 2-year follow-up study of lithium-treated euthymic bipolar patients.* J Clin Psychiatry, 2008. **69**(5): p. 712-9.
14. Mur, M., et al., *Neuropsychological profile in bipolar disorder: a preliminary study of monotherapy lithium-treated euthymic bipolar patients evaluated at a 2-year interval.* Acta Psychiatr Scand, 2008. **118**(5): p. 373-81.
15. Wingo, A.P., P.D. Harvey, and R.J. Baldessarini, *Neurocognitive impairment in bipolar disorder patients: functional implications.* Bipolar Disorders, 2009. **11**(2): p. 113-125.
16. Bas, T.O., et al., *The impact of cognitive impairment, neurological soft signs and subdepressive symptoms on functional outcome in bipolar disorder.* J Affect Disord, 2015. **174**: p. 336-41.

17. Depp, C.A., et al., *Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder.* Bipolar Disord, 2012. **14**(3): p. 217-26.
18. Goodwin, G.M., et al., *Cognitive impairment in bipolar disorder: neurodevelopment or neurodegeneration? An ECNP expert meeting report.* Eur Neuropsychopharmacol, 2008. **18**(11): p. 787-93.
19. Arts, B., et al., *A 2-year naturalistic study on cognitive functioning in bipolar disorder.* Acta Psychiatr Scand, 2011. **123**(3): p. 190-205.
20. Mora, E., et al., *Persistence of cognitive impairment and its negative impact on psychosocial functioning in lithium-treated, euthymic bipolar patients: a 6-year follow-up study.* Psychol Med, 2013. **43**(6): p. 1187-96.
21. Burdick, K.E., J.F. Goldberg, and M. Harrow, *Neurocognitive dysfunction and psychosocial outcome in patients with bipolar I disorder at 15-year follow-up.* Acta Psychiatr Scand, 2010. **122**(6): p. 499-506.
22. Balanza-Martinez, V., et al., *Persistent cognitive dysfunctions in bipolar I disorder and schizophrenic patients: a 3-year follow-up study.* Psychother Psychosom, 2005. **74**(2): p. 113-9.
23. Chaves, O.C., et al., *Association of clinical symptoms and neurocognitive performance in bipolar disorder: a longitudinal study.* Bipolar Disord, 2011. **13**(1): p. 118-23.
24. Gildengers, A., et al., *Two-year course of cognitive function and instrumental activities of daily living in older adults with bipolar disorder: evidence for neuroprogression?* Psychological Medicine, 2013. **43**(04): p. 801-811.
25. Depp, C.A., et al., *Short-term course of neuropsychological abilities in middle-aged and older adults with bipolar disorder.* Bipolar Disord, 2008. **10**(6): p. 684-90.
26. Gildengers, A.G., et al., *The longitudinal course of cognition in older adults with bipolar disorder.* Bipolar Disorders, 2009. **11**(7): p. 744-52.
27. Schouws, S.N., et al., *Cognitive decline in elderly bipolar disorder patients: a follow-up study.* Bipolar Disord, 2012. **14**(7): p. 749-55.
28. Engelsmann, F., et al., *Lithium and memory: a long-term follow-up study.* J Clin Psychopharmacol, 1988. **8**(3): p. 207-12.
29. Delaloye, C., et al., *Longitudinal analysis of cognitive performances and structural brain changes in late-life bipolar disorder.* Int J Geriatr Psychiatry, 2011. **26**(12): p. 1309-18.
30. Moorhead, T.W., et al., *Progressive gray matter loss in patients with bipolar disorder.* Biol Psychiatry, 2007. **62**(8): p. 894-900.
31. Torrent, C., et al., *Long-term outcome of cognitive impairment in bipolar disorder.* J Clin Psychiatry, 2012. **73**(7): p. e899-905.
32. Curran, P.J., K. Obeidat, and D. Losardo, *Twelve Frequently Asked Questions About Growth Curve Modeling*. J Cogn Dev, 2010. **11**(2): p. 121-136.
33. Gildengers, A.G., et al., *Cognitive Functioning in Late-Life Bipolar Disorder.* American Journal of Psychiatry, 2004. **161**(4): p. 736-738.
34. Dhingra, U. and P.V. Rabins, *Mania in the elderly: a 5-7 year follow-up.* J Am Geriatr Soc, 1991. **39**(6): p. 581-3.
35. Langenecker, S.A., et al., *Intermediate: Cognitive phenotypes in bipolar disorder.* Journal of Affective Disorders, 2010. **122**(3): p. 285-293.

36. Weisenbach, S.L., Marshall, D., Weldon, A.L., Ryan, K.A., Vederman, A.C., Kamali, M., Zubieta, J-K., McInnis, M.G., Langenecker, S.A. , *The double burden of age and disease on cognition and quality of life in bipolar disorder.* . International Journal of Geriatric Psychiatry, 2014. **29**: p. 952-961.
37. Nurnberger, J.I., Jr., et al., *Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative.* Archives of General Psychiatry, 1994. **51**(11): p. 849-59; discussion 863-4.
38. Hamilton, M.A.X., *Development of a rating scale for primary depressive illness.* British Journal of Social and Clinical Psychology, 1967. **6**(4): p. 278-296.
39. Young, R.C., et al., *A rating scale for mania: Reliability, validity and sensitivity.* The British Journal of Psychiatry, 1978. **133**(5): p. 429-35.
40. Almeida, J.R.C., et al., *Reduced gray matter volume in ventral prefrontal cortex but not amygdala in bipolar disorder: Significant effects of gender and trait anxiety.* Psychiatry Research - Neuroimaging, 2009. **171**(1): p. 54-68.
41. Davis, J.M. and N. Chen, *Dose Response and Dose Equivalence of Antipsychotics.* Journal of Clinical Psychopharmacology, 2004. **24**(2): p. 192-208.
42. Hassel, S., et al., *Elevated striatal and decreased dorsolateral prefrontal cortical activity in response to emotional stimuli in euthymic bipolar disorder: No associations with psychotropic medication load.* Bipolar Disorders, 2008. **10**(8): p. 916-927.
43. Sackeim, H.A., *The definition and meaning of treatment-resistant depression.* Journal of Clinical Psychiatry, 2001.
44. Heaton, R.K., *A Manual for the Wisconsin Card Sorting Test*. 1981, Odessa, FL: Psychological Assessment Resources.
45. Golden, C., *Stroop Color and Word Test*. 1978, Chicago, IL: Stoelting.
46. Benton, A.L. and K.D. Hamsher, *Multilingual Aphasia Examination*. 1976, Iowa City, IA: The University of Iowa.
47. Wechsler, D., *WAIS-III: Wechsler adult intelligence scale*. 1997: Psychological Corporation San Antonio.
48. War Department Adjutant General's, O., *Army Individual Test Battery; Manual of Directions and Scoring*. 1944, Washington, D.C.: War Department, Adjutant General's Office.
49. Langenecker, S.A., et al., *A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test-retest reliability of the Parametric Go/No-Go Test.* Journal of Clinical and Experimental Neuropsychology, 2007. **29**(8): p. 842-853.
50. Ryan, K.A., et al., *Emotion perception and executive functioning predict work status in euthymic bipolar disorder.* Psychiatry research, 2013. **Advance online publication**.
51. Wechsler, D., *Wechsler Abbreviated Scale of Intelligence (WASI) Manual*. 1999, The Psychological Corporation: San Antonio, TX.
52. Arbuckle, J.L., *Amos*. 2009, SPSS: Chicago.
53. Kline, R.B., *Principles and Practice of Structural Equation Modeling*. 2011: Guilford Press.
54. Hu, L.t. and P.M. Bentler, *Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.* Structural Equation Modeling: A Multidisciplinary Journal, 1999. **6**(1): p. 1-55.

55. Lei, M. and R.G. Lomax, *The Effect of Varying Degrees of Nonnormality in Structural Equation Modeling*. Structural Equation Modeling: A Multidisciplinary Journal, 2005. **12**(1): p. 1-27.
56. Tabachnick, B.G. and L.S. Fidell, *Using multivariate statistics* 3rd ed. 1996, New York, NY: HarperCollins College Publishers.
57. Schumacker, R.E. and R.G. Lomax, *A Beginner's Guide to Structural Equation Modeling*. 2004: Lawrence Erlbaum Associates.
58. Bollen, K.A. and J.S. Long, *Testing Structural Equation Models*. 1993: SAGE Publications.
59. Xu, G., et al., *Neuropsychological performance in bipolar I, bipolar II and unipolar depression patients: a longitudinal, naturalistic study.* J Affect Disord, 2012. **136**(3): p. 328-39.
60. Salthouse, T., *The processing speed theory of adult age differences in cognition.* Psychological Review, 1996. **103**: p. 403-428.
61. Stern, Y., et al., *Brain networks associated with cognitive reserve in healthy young and old adults.* Cereb.Cortex, 2005. **15**(4): p. 394-402.
62. Lopez-Jaramillo, C., et al., *Lithium treatment effects on the neuropsychological functioning of patients with bipolar I disorder.* J Clin Psychiatry, 2010. **71**(8): p. 1055-60.
63. Santos, J.L., et al., *A five-year follow-up study of neurocognitive functioning in bipolar disorder.* Bipolar Disorders, 2014. **16**(7): p. 722-731.

Abstract

Objective: Executive Functioning (EF) deficits in bipolar disorder (BD) are commonly present regardless of mood state and therefore are considered core features of the illness. However, very little is known about the temporal stability of these deficits. We examined the natural course of EF over a five year period in BD and healthy control (HC) samples.

Method: Using a 5-year longitudinal cohort, 91 individuals with BD and 17 HC were administered a battery of neuropsychological tests that captured four main areas of EF: Processing Speed with Interference Resolution, Verbal Fluency with Processing Speed, Inhibitory Control, and Conceptual Reasoning and Set Shifting. Evaluations occurred at study entry, one, and five years later.

Results: Latent Growth Curve Modeling demonstrated that the BD group performed significantly worse in all EF areas than the HC group. Changes in EF from baseline to 5-year follow-up were similar across both diagnostic groups. Older age at baseline, above and beyond education and diagnosis, was associated with worse initial performance in EF. Being of older age was associated with greater decline in Processing Speed with Interference Resolution, and Verbal Fluency with Processing Speed. Higher education was marginally associated with a smaller declining slope for Processing Speed with Interference Resolution.

Conclusions: Executive functioning deficits in BD persist over time, and in the context of normative age-related decline, may place individuals at greater risk for cognitive disability as the disease progresses. Age and having a BD diagnosis together, however, do not accelerate executive functioning decline over time.

Table 1. Demographic characteristics and executive functioning factor scores for the bipolar and healthy control groups. Data are presented as Mean (SD).

	Baseline				**1 Year**				**5 Year**			
	Bipolar n=91	Healthy Controls n=17	*t*	*p*	Bipolar n=91	Healthy Controls n=17	*t*	*p*	Bipolar n=91	Healthy Controls n=17	*t*	*p*
Age	42.06(11.30)	34.96(14.29)	-2.62	0.01	--	--	--	--	--	--	--	--
Education	15.53(2.18)	16.04(2.42)	1.00	0.32	--	--	--	--	--	--	--	--
Gender[a]												
% Females	74.44	76.00	0.03	0.87	--	--	--	--	--	--	--	--
Verbal Intelligence[b]					--	--	--	--	--	--	--	--
HRDS	8.88(6.22)	2.29(2.93)	-5.03	<0.001	7.74(5.94)	1.80(1.96)	-4.91	<0.001	7.58(5.32)	2.12(3.60)	-4.92	<0.001
YMRS	4.26(4.70)	0.86(1.57)	-1.87	0.71	2.64(3.97)	0.64(1.43)	-2.32	0.02	3.12(3.88)	0.96(1.67)	-2.76	.007
VFPS	-0.41(0.82)	0.05(0.60)	2.36	0.02	-0.31(0.74)	0.12(0.79)	2.54	0.01	-0.50(0.88)	0.10(0.79)	3.08	0.003
CRSS	-0.11(0.73)	-0.01(0.76)	0.57	0.57	-0.32(0.62)	-0.33(0.63)	-0.06	0.95	-0.30(0.36)	-0.32(0.49)	-0.26	0.80
PSIR	-0.45(0.68)	0.34(0.57)	4.72	<.001	-0.27(0.60)	0.26(0.61)	3.86	<.001	-0.40(0.68)	0.17(0.62)	3.76	<.001
IC	-0.38(0.85)	0.01(0.90)	1.81	0.07	-0.25(0.84)	0.12(0.71)	2.05	0.04	-0.41(0.77)	-0.34(1.11)	0.33	0.74

Notes: [a]Chi-Square analyses. [b]WASI IQ = Wechsler Abbreviated Scale of Intelligence Intelligence Vocabulary scaled score; HRDS = Hamilton Rating Scale for Depression; YMRS = Young Mania Rating Scale; VFPS = Verbal Fluency and Processing Speed; CRSS = Conceptual Reasoning and Set-Shifting; PSIR = Processing Speed with Interference Resolution; IC = Inhibitory Control

Table 2. Correlations between executive functioning domains, demographics and mood rating variables.

2a. Entire Sample

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1 Gender	1.00																					
2 Age	-.07	1.00																				
3 Education	.14	.21*	1.00																			
4 HRDS_0	.17	.01	-.33**	1.00	.																	
5 YMRS_0	.04	-.04	-.18	.38**	1.00																	
6 HRDS_1	.16	-.09	-.30**	.67**	.35**	1.00																
7 YMRS_1	.02	-.01	-.13	.24*	.33**	.39**	1.00															
8 HRDS_5	.01	-.05	-.35**	.46**	.41**	.56**	.21*	1.00	.													
9 YMRS_5	-.01	-.05	-.06	.17	.25*	.25**	.16	.33**	1.00													
10 VFPS_0	.17	-.14	.27**	-.28**	-.07	-.22*	-.10	-.19	.03	1.00												
11 CRSS_0	.02	-.21	.14	-.13	-.11	-.05	-.19	-.04	-.05	.37**	1.00											
12 PSIR_0	.08	-.41**	.10	-.21*	-.07	-.16	-.20	-.13	-.10	.78**	.34**	1.00										
13 IC_0	.07	-.11	.13	-.14	-.06	-.09	-.12	.07	-.10	.43**	.34**	.58**	1.00									
14 VFPS_1	.20*	-.27**	.22*	-.22*	-.02	-.17	-.06	-.12	-.02	.83**	.38**	.83**	.47**	1.00								
15 CRSS_1	.08	-.15	.03	-.12	-.20	-.15	-.05	.03	-.06	.28*	.48**	.27*	.13	.36**	1.00	.						
16 PSIR_1	.18	-.49**	.05	-.17	-.04	-.10	-.06	-.14	-.09	.60**	.31**	.85**	.40**	.78**	.27*	1.00	.					
17 IC_1	.19	-.29**	.07	-.09	-.07	-.04	.09	-.04	-.07	.52**	.28*	.56**	.47**	.62**	.13	.65**	1.00	.				
18 VFPS_5	.12	-.32**	.27**	-.28**	-.11	-.30**	.02	-.34**	-.05	.80**	.29**	.73**	.37**	.80**	.27*	.66**	.47**	1.00	.			
19 CRSS_5	.05	-.09	.09	-.06	-.07	-.10	-.00	-.06	-.02	.50**	.18	.38**	.19	.44**	.21	.33**	.28**	.46**	1.00			
20 PSIR_5	.00	-.50**	.19	-.23*	-.09	-.20*	-.17	-.26**	-.08	.68**	.33**	.85**	.41**	.71**	.18	.80**	.53**	.80**	.44**	1.00		
21 IC_5	-.10	-.25*	.07	-.11	-.13	-.07	-.27**	-.13	-.01	.57**	.21	.61**	.52**	.48**	-.04	.50**	.58**	.46**	.37**	.65**	1.00	
22. Diagnosis	-.02	.18	-.20*	.46**	.32**	.47**	.21*	.53**	.28**	-.27**	.04	-.37**	-.18	-.24*	-.14	-.36**	-.18	-.34**	-.11	-.42**	-.15	1.00

2b. BD Upper Right, HC Lower Left

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 Gender	1.00	-.03	.20	.18	.02	.17	.05	.03	-.03	.15	.03	.02	-.02	.19	.07	.12	.19	.08	.07	-.04	-.11
2 Age	-.30	1.00	.23*	-.09	-.09	-.18	-.06	-.17	-.13	-.08	-.20	-.38**	-.04	-.23*	-.13	-.50**	-.30**	-.27*	-.05	-.49**	-.24*
3 Education	-.18	.38	1.00	-.30**	-.15	-.29**	-.14	-.33**	-.03	.27*	.13	.09	.13	.24*	.02	-.01	.01	.27*	.10	.15	.05
4 HRDS_0	.47	.12	-.12	1.00	.32**	.62**	.19	.35**	.08	-.21	-.14	-.10	-.09	-.15	-.10	-.03	-.02	-.19	-.04	-.09	-.04
5 YMRS_0	.31	-.19	-.08	.18	1.00	.29**	.30**	.34**	.20	-.01	-.12	.01	-.02	.04	-.18	.06	-.03	-.03	-.09	.00	-.10
6 HRDS_1	.38	-.30	.23	.08	-.20	1.00	.37**	.47**	.19	-.16	-.07	-.05	-.05	-.10	-.12	.03	.03	-.21	-.16	-.07	-.00
7 YMRS_1	-.03	.23	.35	.19	.04	-.03	1.00	.15	.13	-.05	-.17	-.12	-.07	.00	-.04	.03	.15	.09	.01	-.09	-.25*
8 HRDS_5	.09	-.01	.21	.18	-.27	.15	.32	1.00	.26*	-.11	-.05	.02	.16	-.01	.07	.03	.06	-.24*	-.05	-.11	-.07
9 YMRS_5	.08	.20	.20	.11	-.20	-.03	.05	.39	1.00	.11	-.06	-.00	-.08	.07	-.06	.03	-.01	.05	.06	.02	.05
10 VFPS_0	.42	-.42	-.20	.04	-.01	.08	-.60*	.15	-.32	1.00	.37**	.76**	.38**	.82**	.27*	.54**	.47**	.78**	.52**	.64**	.56**
11 CRSS_0	.00	-.52	.29	-.32	--	.22	-.75	-.20	.24	.87	1.00	.33**	.33**	.39**	.50**	.30**	.26*	.31**	.22	.32**	.21
12 PSIR_0	.54	-.48	-.28	.14	.25	.29	-.76*	-.06	-.30	.89**	.66	1.00	.58**	.81**	.27*	.81**	.51**	.69**	.46**	.82**	.60**
13 IC_0	.54	-.36	-.12	-.05	-.30	.33	-.60	.27	.16	.81**	.72	.46	1.00	.47**	.10	.36**	.45**	.35**	.27*	.38**	.51**
14 VFPS_1	.28	-.28	-.19	.09	.35	.19	-.43	.08	-.32	.83**	.23	.85**	.34	1.00	.37**	.75**	.59**	.77**	.46**	.68**	.46**
15 CRSS_1	.49	-.20	.03	.47	--	.04	-.95*	.44	.90*	.98*	.53	.25	.81	.13	1.00	.27*	.13	.26*	.31**	.16	-.05
16 PSIR_1	.54*	-.34	-.10	.11	.24	.40	-.41	.09	-.35	.84**	.57	.92**	.44	.81**	-.05	1.00	.64**	.60**	.32**	.78**	.49**
17 IC_1	.23	-.12	.22	-.04	.13	.19	-.41	-.16	-.44	.77**	.58	.72**	.47	.64**	-.11	.66**	1.00	.42**	.24*	.48**	.54**
18 VFPS_5	.34	-.36	-.14	.19	.39	.17	-.31	.12	-.23	.76**	-.15	.75**	.25	.90**	.13	.66**	.52*	1.00	.54**	.78**	.44**
19 CRSS_5	.03	-.19	.08	-.31	.19	.25	.10	-.41	-.74**	.32	-.10	.35	-.06	.44	-.78	.51*	.53*	.35	1.00	.51**	.36**
20 PSIR_5	.17	-.47	-.09	-.04	.31	.26	-.56*	-.04	-.23	.78**	.69	.92**	.41	.83**	.25	.75**	.74**	.74**	.45	1.00	.65**
21 IC_5	.00	-.20	-.05	-.11	.08	-.02	-.27	-.23	-.27	.56	.50	.59*	.45	.45	-.10	.45	.77**	.37	.46	.60*	1.00

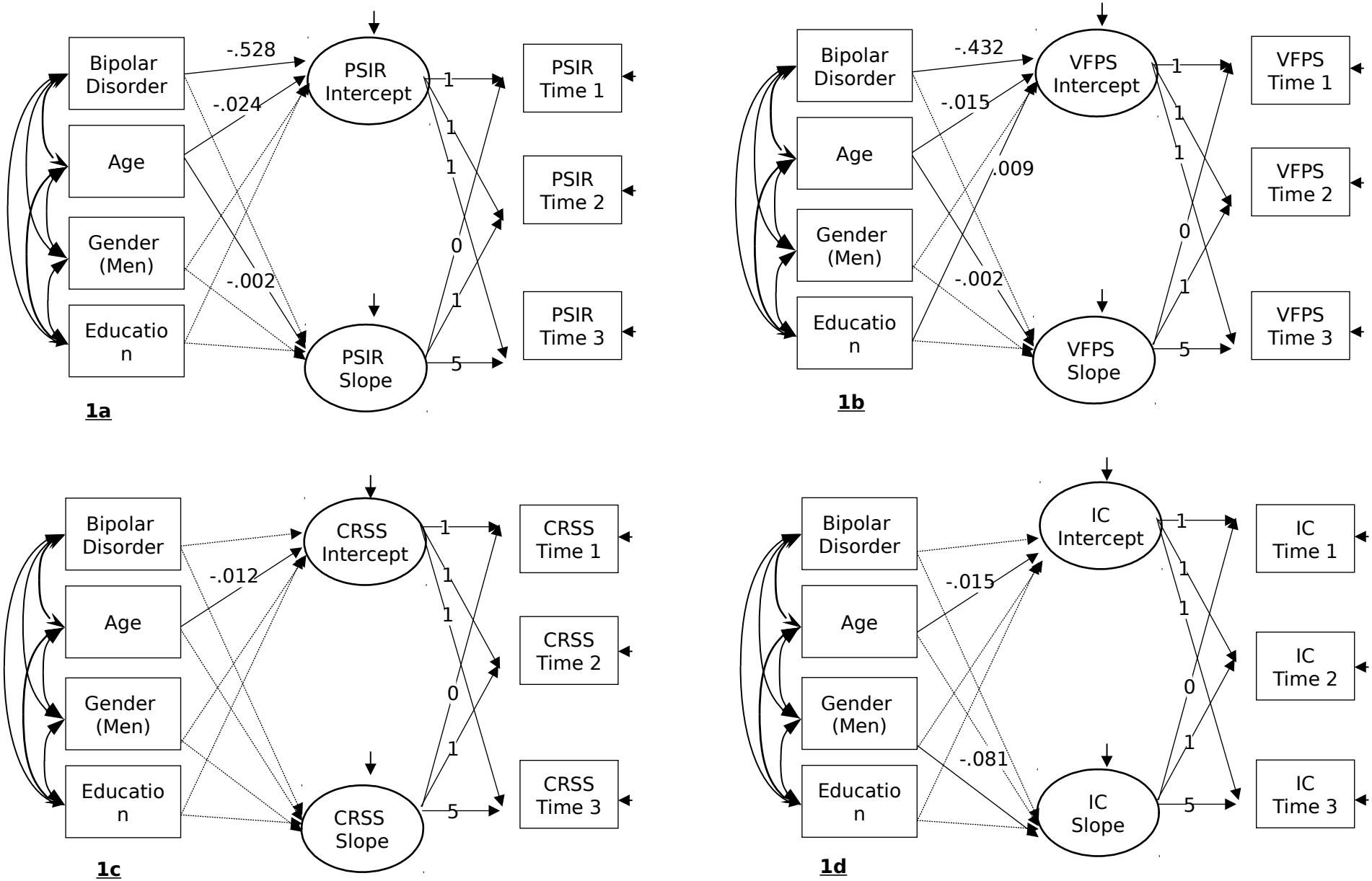
Notes. p value < .05*, p value <.001**; HRDS = Hamilton Rating Scale for Depression; YMRS = Young Mania Rating Scale; VFPS = Verbal Fluency and Processing Speed; CRSS = Conceptual Reasoning and Set-Shifting; PSIR = Processing Speed with Interference Resolution; IC = Inhibitory Control; _0=Baseline; _1=1 year follow-up; _5=5 year follow-up; Diagnosis=BD vs HC.

Table 3: Summary of path models to test the effect of bipolar disorder on baseline and trajectory of PSIR, VFPS, CRSS, and IC over a 5 year period

			PSIR		VFPS		CRSS		IC		
			B(SE)	P	B(SE)	P	B(SE)	P	B(SE)	P	B(SE)
BD		Intercept	-.528(.160)	0.000	-0.432(0.204)	0.034	-.067(.175)	0.704		-.180(.223)	0.419
BD	□	Slope	.014(.023)	.538	-0.01(0.031)	0.743	.034(.041)	0.405		.041(.045)	0.364
Age	□	Intercept	-.024(.005)	0.000	-0.015(0.006)	0.016	-.012(.005)	0.019		-.015(.007)	0.031
Age	□	Slope	-.002(.001)	.030	-0.002(0.001)	0.023	.002(.001)	0.235		-.002(.001)	0.246
Gender	□	Intercept	.158(.131)	.231	0.258(0.167)	0.123	.115(.142)	0.418		.216(.183)	0.239
Gender	□	Slope	-.030(.019)	.114	-0.039(0.025)	0.125	-.037(.034)	0.281		-.081(.037)	0.027
Education	□	Intercept	.045(.027)	.100	0.09(0.035)	0.01	.047(.030)	0.117		.061(.038)	0.112
Education	□	Slope	.008(.004)	.057	0.006(0.005)	0.223	-.004(.007)	0.538		-.003(.008)	0.701

Notes: BD=Bipolar Disorder; PSIR= Processing Speed with Interference Resolution; VFPS = Verbal Fluency with Pressing Speed; CRSS = Conceptual Reasoning and Set Shifting; IC = Inhibitory Control

Figure 1. Trajectory of Executive Functioning Performance over 5 years among individuals with Bipolar Disorder and Controls using Latent Growth Modeling. 1a) Processing Speed with Interference Resolution (PSIR), 1b) Verbal Fluency with Processing Speed (VFPS), 1c) Inhibitory Control (IC), and 1d) Conceptual Reasoning and Set Shifting (CRSS).



Supplemental Table 1. Demographic data for those who completed vs did not complete 5 year retesting (for both BD and HC) and clinical variables for BD participants who completed vs did not complete 5 year retesting.

	Non-completers	**Completers**	***t*/^{2}**	***p***
Age	35.24 (13.01)	41.02 (12.23)	3.41	.001
Education	15.36 (2.32)	15.67 (2.14)	1.04	.301
Gender (% female)	49.6	75.5	15.82	.00007
Verbal Intelligence[a]	12.39	12.61	.54	.589
Chronicity mood (% w/ mood symptoms most of the time)	60.7	48.2	2.71	.10
Rapid cycling (% rapid cyclers)	46.9	52.5	.39	.53
Psychosis (% with history of)	45.0	52.9	1.35	.25
Suicide history (% attempted)	37.4	31.1	4.68	.10
Illness impact (% impacted)	52.6	62.5	2.17	.14
Years with illness	18.99 (12.70)	22.24 (11.32)	1.81	.07
Mood NOE	72.27 (128.88)	47.82 (91.58)	-1.46	.15
Medload	2.78 (2.25)	3.04 (1.90)	.80	.42

[a]Weschler Abbreviated Scale of Intelligence Vocabulary scaled score; mood NOE=number of lifetime mood episodes (depression, mania/hypomania); Illness Impact=clinician's rating of how much the illness have impacted their life via functioning; Medload=total medication burden based on number, dosage, and time taking a psychiatric medication.

Supplemental Table 2. Executive Functioning Factors, Tests and Reliability Estimates

factor analysis and reliability	Test	Cronbach's alpha r
Verbal Fluency and Processing Speed	Phonemic Fluency Category Fluency Digit Symbol Stroop Color Word Test (SCWT) Word Condition T Color Condition T Trail Making Test (TMT) Form B	0.81
Conceptual Reasoning and Set-Shifting	Wisconsin Card Sort Test Correct Perseverative errors (inverted) Parametric Go/No-Go (PGNG) Mean accuracy for target trials	0.79
Processing Speed with Interference Resolution	TMT Form A Form B Digit Symbol SCWT Interference T PGNG Mean target response time (inverted)	0.76
Inhibitory Control	PGNG Mean target response time (inverted) Mean accuracy for inhibitory trials	N / N/A

SCWT=Stroop Color Word Test, PGNG=Parametric Go No-Go Test, TMT=Trail Making Test

Supplemental Table 3: Summary of fit statistics of the unconditional models

	IC					PSIR					CRSS					VFPS			
Fit																			
	CFI	CMIN	DF	P	CMIN/DF	CFI	CMIN	DF	P	CMIN/DF	CFI	CMIN	DF	P	CMIN/DF	CFI	CMIN	DF	P
	0.941	7.32	3	0.062	2.44	0.931	19.732	3	0	6.577	0.9	4.993	3	0.172	1.664	0.994	4.261	3	0.235
	Mean	SE	P			Mean	SE	P			Mean	SE	P			Mean	SE	P	
Means																			
Intercept	-0.288	0.081	<0.001			-0.254	0.068	<0.001			-0.389	0.064	<0.001			-0.263	0.079	<0.001	
Slope	-0.026	0.016	0.106			-0.009	0.009	0.302			0.01	0.015	0.48			-0.02	0.011	0.078	
Covariance																			
	0.008	0.016	0.606			0.002	0.006	0.728			-0.031	0.013	0.015			0.005	0.01	0.592	
CRSS_fctr_ICEPT	0.437	0.102	<0.001			0.431	0.068	<0.001			0.226	0.06	<0.001			0.577	0.093	<0.001	
CRSS_fctr_SLOPE	-0.002	0.005	0.666			0.001	0.001	0.64			0.005	0.004	0.237			0.003	0.002	0.204	
e1	0.385	0.056	<0.001			0.089	0.013	<0.001			0.22	0.035	<0.001			0.133	0.019	<0.001	
e2	0.385	0.056	<0.001			0.089	0.013	<0.001			0.22	0.035	<0.001			0.133	0.019	<0.001	
e3	0.385	0.056	<0.001			0.089	0.013	<0.001			0.22	0.035	<0.001			0.133	0.019	<0.001	

Supplemental Table 4. Correlations between executive functioning domains and clinical features among the BD participants.

Variable	VFPS_0	CRSS_0	PSIR_0	IC_0	VFPS_1	CRSS_1	PSIR_1	IC_1	VFPS_5	CRSS_5	PSIR_5	IC_5
Chronicity mood	-.19	-.09	-.20	.07	-.17	.06	-.12	.06	-.27*	-.07	-.27*	-.17
Rapid Cycling	-.05	.27*	-.09	.09	-.19	.24	-.06	.08	-.03	-.01	-.08	-.10
Psychosis history	-.06	.04	-.01	-.05	.01	.12	.12	.00	.11	-.18	.02	.05
Suicide History	-.03	.13	.05	.03	-.03	.01	.09	.13	-.13	-.04	-.04	.01
Illness Impact	-.28*	-.11	-.18	-.16	-.28**	-.09	-.14	-.15	-.28*	-.04	-.23*	-.19
Age of BD onset	-.04	-.15	-.16	.03	-.06	.06	-.22*	-.17	-.02	-.01	-.17	-.05
Years with Illness	-.06	-.08	-.26*	-.07	-.18	-.18	-.32**	-.17	-.26*			
	-.04	-.36**	-.20									
Mood NOE	-.08	-.04	-.05	-.02	-.11	-.20	-.11	-.10	-.21	-.01	-.16	.01

Notes. *p* value < .05*, *p* value <.001**; Chronicity of Mood (Mood symptoms most of time, relapsing-remitting, illness duration <2 years); Rapid Cycling (Yes/No); Psychosis History (Yes/No); Suicide History (Prior attempts/no attempts); Illness Impact=clinician's rating of how much the illness have impacted their life via functioning; mood NOE=number of lifetime mood episodes (depression, mania/hypomania); VFPS=Verbal Fluency with Processing Speed; HRDS=Hamilton Depression Rating Scale; YMRS=Young Mania Rating Scale; CRSS=Conceptual Reasoning with Set-Shifting; PSIR=Processing Speed with Interference Resolution; IC=Inhibitory Control (_0=Baseline; _1=1 year follow-up; _5=5 year follow-up)

Highlights

- Deficits in executive functioning among individuals with bipolar illness are present and are substantially worse compared to unaffected individuals.
- Executive functioning deficits are stable over time with a normal age-related decline on top of the baseline difficulties.
- There is equivalence of slope in executive functioning decline over five years in the bipolar and control samples - those with bipolar disorder start out and continue to show lower executive functioning performance over time.
- Older age, regardless of diagnosis, is related to greater decline over five years and higher education is related to smaller decline.

<u>Title:</u> would the authors consider referring to "latent growth modelling" in their title?

We have updated the title to reflect Reviewer 2's comments regarding the findings and have incorporated the phrase "Latent Growth Modeling" into it.

<u>Table 1:</u> I would recommend that the authors provide mean/SD values for HRDS, YMRS, education, IQ, number of episodes/onset of BD and comorbidities.

We have added the Mean/SD for IQ, HRDS, and YMRS to Table 1. The table already included education. We intentionally did not include BD illness-specific variables such as number of episodes/onset, or comorbidities to the table or paper as this is beyond the scope of our LGCM models. However, we looked at these specific variables when examining if there are any differences between participants in the longitudinal study who completed 5 years of testing vs those who were eligible to complete testing, so this data is presented in Supplemental Table 1. We hope this is sufficient.

<u>Data:</u> 1.I was wondering if the authors could please clarify which outcome measures they had access to prior to calculating the four executive functioning scores with the factor analyses. For instance, did they use both reaction times and accuracy for all the tasks or only TMT-B, Stroop, and Wisconsin Card Sorting Test? As far as I know for the other tasks there are time-discontinuation limits.

We included the specific variables that made up each executive functioning factor score in Supplemental Table 2, which also includes reliability estimates for the executive functioning factors. This table indicates that we used "seconds" rather than number of errors for TMT (these variables were included in the processing speed factors) and it also indicates which two variables from the WCST loaded in the CRSS factor during the factor analysis, etc. In our original work (Langenecker et al, 2010, JAD), we included all variables from the presented executive functioning tasks and the specific variables that loaded within each executive functioning factor via the factor analysis. The subsequent confirmatory factor analyses settled on the presented variables in Supplemental Table 2. We included reference to this table in the Method section, page 8.

-Reviewer #2

The authors want to build a better case as to why their question is important. Moreover, some alternative explanations need to be considered. For example, do we know to which extent BD status remained stable over the course of the study? How would the results be affected if there was remission within the BD sample?

Since this is a longitudinal study, we do monitor the diagnoses over time so we investigated the question regarding stability in diagnosis over time. For our sample, we had no BD individuals

that changed from a BD diagnosis over the course of the study. We had 8 control participants who changed diagnosis in the interim between baseline and year 5 testing. Four of the controls received a new diagnosis of Major Depressive Disorder, single episode, 1 Depression NOS, 1 PTSD, 1 BD type II, with single depression, and 1 BD type I. Given this, we re-ran all analyses including the LGCM without these individuals.

Limiting the sample to those whose diagnosis remained stable over the course of the study did not influence effect of having BD on baseline and the slope of the outcomes. However, it caused some change in the effects of covariates. Main changes were related to the following two outcomes; the slope of PSIR/VFPS was associated with higher age. We have made the changes throughout the manuscript.

We still find the results compelling and the overall conclusions regarding findings have not changed. Further, we investigated the change in cognitive performance overtime for these 8 individuals and there were no significant changes in neuropsychological test performance.

We also looked at the mood scores (HRDS and YMRS) in the BD sample over time to assess stability in mood symptoms and found no overall difference in depression or mania symptom scores from baseline to year 5. However, similar to our response to Review 1 (Comment 1), we did not include BD-illness specific variables in this paper as it was beyond the scope (looking at the relationship between diagnosis-HC vs BD-and cognitive trajectories over time) and space limitations would not allow a paper that included tables and figures for 8-10 additional LGCM figures and tables.

More data need to be presented to evaluate basic claims of the paper. Specifically, can the authors provide a full table of intercorrelations (including all four EF measures, BD status, age, gender, education)? Can they provide results from the unconditional LGCM (e.g., intercept and slope mean and variances; intercept-slope associations; model fit)? Why did they settle on linear models and did not test nonlinear models as well? Table 1 shows that there is actually nonlinear change in PSIR (e.g., -.45 to -.27 to -.40 in BD), which likely accounts for the suboptimal fit of the PSIR LGCM (I don't think the fit of this model was "very good"). Did the authors include intercorrelations between the independent variables in their conditional LGCM (i.e., age, gender, education, BD status) and, if not, why not? I don't see them in the figures.

We have now included a full table of intercorrelations between diagnosis, covariates, and outcomes at all three time points (See Table 2 and also referenced in Results section, page 10 & 11). We added intercept and slope mean and variances; intercept-slope associations; model fit as another table (Table 3). In the unconditional model, adding nonlinear slopes required adding 6 new parameters (1 slope, 3 paths, and 2 covariances) which worsened our fit indices (added Supplemental Table 3; in data analysis section). Thus, even if the nonlinear slope was significant (not the effect of bipolar disorder but its difference from zero), due to the poor fit, we did not interpret or report any models with non-linear slope. As our models are limited to those with linear slopes, we do not make any interpretation about the non-linear trajectories over time (secondary to BD). We have mentioned not enough sample size to have a robust model with non-linear fit as a limitation.

The kind of LGCM approach chosen is somewhat at odds with the goal of the study. In order to examine "stability of executive functioning" (e.g., title), one would have expected a multi-group LGCM. This kind of approach would be able to tell us to what extent mean levels of executive functioning [e.g., within the BD group or the controls] remain stable over time. For example, in Table 1, we can see that the VPFS means actually decrease over time in BD, whereas they appear to slightly increase in controls. This pattern might support the authors' hypothesis --a multi-group approach would tell us whether these different change trajectories are actually significantly different from each other. Moreover, this approach would allow for modeling trajectories separately for high-burden BD and low-burden BD (contrasting these two subgroups), something that the authors can't accomplish with their current approach. If they worry about controlling for age, gender, and education, there are also ways to do that using a multi-group approach (e.g., using residualized scores). In addition, it would be interesting to learn more about rank-order stability (e.g., by examining wave-to-wave correlations).

Thank you for this very important point. First, similar to multi-group LGCM, we ran our models for BDs and controls and observed that mean slope is not statistically significant (not different from zero) which suggests there is no linear decline in the performances, regardless of the domain of the test.

In the previous version, we had considered "no difference between our patients and controls over time" as indicator of "stability of the tests" over time, which is not very accurate, as the reviewer has suggested. Based on the comment, we have changed the interpretation, and have avoided use of the term "stability of the tests". Instead, we are referring to no difference with the controls.

Finally, is there really stability in executive functioning? With the current data, we cannot say for sure and it would be important to conduct analyses that are better suited to examine this question (see 2). Whatever the result is, it is important to keep in mind that the sample size (although certainly very respectable) does not afford statistical power to detect small effects. It would be helpful to present a power analysis and mention this as a limitation.

We looked at the significance (and variation) of the linear slope to detect if there is stability of not. We have added this to our note, and added the results. In our limitation we discussed our statistical limited power, we added that we cannot rule out that no significant effect of BD is not due to low sample size.

More detail is needed on how the four executive functioning scores were derived.

Please see response to Reviewer 1. We have included the specific EF variables in Supplemental Table 2. We are happy to put in the specific details regarding how the four EF scores were derived from our Journal of Affective Disorders paper in 2010 if the editor finds it necessary.

The authors use very liberal cut off values to evaluate model fit (e.g., chi square to degrees of freedom ratio less than 4.0; see Kline).

Different references have suggested various thresholds for goodness of fit in SEM. We have provided some references supporting use of these fit indices. For Chi square to degree of freedom, the recommended threshold range from 2 (Ullman, 2001) to 5 (Schumacker & Lomax, 2004), CFI, thresholds .90 (Byrne, 1994) and .95 (Schumacker & Lomax, 2004) are both accepted. For RMSEA, .08 (Browne & Cudeck, 1993; Hu & Bentler, 1998) and .05 (Stieger, 1990). The relative chi-square should be less than 2 or 3 (Kline, 1998; Ullman, 2001) or 5 (Schumacker & Lomax, 2004).

We added Schumacker & Lomax, 2004 and Browne & Cudeck, 1993 to our citations.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newsbury Park, CA: Sage.

Byrne, B. M. (1994). Structural equation modeling with EQS and EQS/Windows. Thousand Oaks, CA: Sage Publications.

Kline, R. B. (1998). Principles and practice of structural equation modeling. NY: Guilford Press.

Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling, Second edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioural Research, 25, 173-180.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (2001). Using Multivariate Statistics (4th ed; pp 653- 771). Needham Heights, MA: Allyn & Bacon.

Please avoid causal language (e.g., "impact"; "influence").

We have changed the wording throughout the manuscript to avoid using such causal language.

What does "best estimate process" mean?

We have clarified this in the Method section, page 7.

What does "multivariable analysis" mean (p. 8)?

Despite multivariable and multivariate are being used interchangeably, multivariable analysis seems to be a better term in this case; however, we changed this to LGCM to avoid any confusion. Here is a definition and a source:

This is a verbatim of a paper by Hidalgo and Goodman (2013) published in AJPH:

> *Statistically speaking, multivariate analysis refers to statistical models that have 2 or more dependent or outcome variables,1 and multivariable analysis refers to statistical models in which there are multiple independent or response variables.2*

1. Van Belle G. Biostatistics: A Methodology for the Health Sciences. Hoboken, NJ: Wiley-Interscience; 2004
2. Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med. 2003;138(8):644–650

Hidalgo B, Goodman M. Multivariate or multivariable regression?. American journal of public health. 2013 Jan;103(1):39-40.

The last sentence in the Discussion seems a bit far-reaching.

We have tempered this statement a bit and added a follow-up sentence to indicate that we still need longer-term follow up studies to rule out a neuroprogressive course.

In the Figures (p. 26), it would be helpful to indicate the construct under study for the latent variables (e.g., "PSIR Baseline" instead of "Baseline")

We have redrawn our figures with the name of the construct added to the latent factors (slope and intercept).