Author/s:
Bauer, IE;Suchting, R;Van Rheenen, TE;Wu, MJ;Mwangi, B;Spiker, D;Zunta-Soares, GB;Soares, JC

Title:
The use of component-wise gradient boosting to assess the possible role of cognitive measures as markers of vulnerability to pediatric bipolar disorder

Date:
2019-03-04

Citation:
Bauer, I. E., Suchting, R., Van Rheenen, T. E., Wu, M. J., Mwangi, B., Spiker, D., Zunta-Soares, G. B. & Soares, J. C. (2019). The use of component-wise gradient boosting to assess the possible role of cognitive measures as markers of vulnerability to pediatric bipolar disorder. Cognitive Neuropsychiatry, 24 (2), pp.93-107. https://doi.org/10.1080/13546805.2019.1580190.

HHS Public Access

Author manuscript

Cogn Neuropsychiatry. Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

Cogn Neuropsychiatry. 2019 March ; 24(2): 93–107. doi:10.1080/13546805.2019.1580190.

The use of component-wise gradient boosting to assess the possible role of cognitive measures as markers of vulnerability to pediatric bipolar disorder

Isabelle E. Bauer[1], **Robert Suchting**[1], **Tamsyn E. Van Rheenen**[2,3], **Mon-Ju Wu**[1], **Benson Mwangi**[1], **Danielle Spiker**[1], **Giovana B. Zunta-Soares**[1], **Jair C. Soares**[1]

[1]The University of Texas Health Science Center at Houston, Department of Psychiatry and Behavioral Sciences, Houston (Texas), USA

[2]Melbourne Neuropsychiatry Centre, Level 3, Alan Gilbert Building, 161 Barry St, Carlton, VIC 3053, Australia

[3]Brain and Psychological Sciences Research Centre (BPsyC), Faculty of Health, Arts and Design, School of Health Sciences, Swinburne University, Victoria, Australia

Abstract

Background and Aims: Cognitive impairments are one of the primary hallmarks symptoms of bipolar disorder (BD). Whether these deficits are markers of vulnerability or symptoms of the disease is still an open debate. This study used a component-wise gradient (CGB) machine learning algorithm to identify cognitive measures that could accurately differentiate pediatric BD, unaffected offspring of BD parents, and healthy volunteers.

Methods: 59 healthy controls (HC; 11.19±3.15 yo; 30 girls), 119 children and adolescents with BD (13.31±3.02 yo, 52 girls) and 49 unaffected offspring of BD parents (UO; 9.36±3.18 yo; 22 girls) completed the CANTAB cognitive battery.

Results: After algorithm training, CGB achieved accuracy of 71.4% and an AUROC of 0.797 in classifying individuals as either BD or non-BD on a dataset held out for validation for testing. The strongest cognitive predictors of BD were measures of affective processing and sustained attention. Measures of cognition did not differentiate between unaffected offspring and HC.

Conclusions: Our findings suggest that alterations in affective processing and sustained attention are markers of BD in pediatric populations. Longitudinal studies should determine whether UO with a cognitive profile similar to that of HC in late childhood or early adolescence

Corresponding Author: Isabelle E. Bauer, PhD, University of Texas Health Science Center at Houston, Department of Psychiatry and Behavioral Sciences, 1941 East Rd., Houston TX, 77054, USA, Isabelle.E.Bauer@uth.tmc.edu.

Contributors

DS, GZS and JCS designed the study, wrote the protocol and collected the data. IB wrote the first draft of the manuscript and performed exploratory analyses (descriptives, cluster analyses). RS performed the primary statistical analyses (component-wise gradient boosting, model approximation), co-wrote the data analytic strategy and results, read and revised the first draft of the manuscript. TVR provided conceptual and methodological input, read and revised the first draft of the manuscript, and approved the final manuscript.

Declarations of Interest

Dr Soares has received grants/research support from Forrest, BMS, Merck, Stanley Medical Research Institute, NIH and has been a speaker for Pfizer and Abbott. The other authors have no conflicts of interest to declare.

are at less or equal risk for mood disorders. Future studies should include relevant cognitive measures for BD such as verbal memory and individuals' genetic risk scores.

Keywords

bipolar disorder; high-risk; machine learning; CANTAB; cognitive

Introduction

Cognitive deficits are recognized core deficits of bipolar disorder (BD) alongside mood alterations (Burdick, Goldberg, Harrow, Faull, & Malhotra, 2006). Specifically, impairment in the domain of verbal memory and executive functions have been shown to persist across all phases of BD (Martínez-Arán et al., 2004) and are strong predictors of clinical outcomes such as risk for relapse and global functioning (Bauer, Hautzinger, & Meyer, 2017; Bora, Harrison, Yücel, & Pantelis, 2013; Mora, Portella, Forcada, Vieta, & Mur, 2013; Roux et al., 2017). BD is a highly heritable disease and the risk to develop BD is approximately 8 times higher in first-degree relatives (e.g. offspring, siblings, twins) than in those of patients with non-BD major depression (Tsuang, Tohen, & Jones, 2011; Wilde et al., 2014). First-degree relatives of BD patients are also approximately four times more likely to develop any kind of mood, anxiety and attention deficit hyperactivity disorders (Chang, Steiner, Dienes, Adleman, & Ketter, 2003; Glahn, Bearden, Niendam, & Escamilla, 2004) (Akdemir and Gökler, 2008; Lapalme, Hodgins, & LaRoche, 1997; Mesman, Nolen, Reichart, Wals, & Hillegers, 2013; Robinson et al., 2006).

Whether or not cognitive impairment is a hallmark of the disease or rather the result of the disease is still controversial because of the limited number of family studies including relatives of BD patients and the challenges in recruiting these vulnerable individuals.

Early identification of symptoms for treatment and prevention purposes are primary research targets in BD. Adolescent offspring of BD parents display deficits in processing speed (de la Serna et al., 2017) verbal learning and memory, and cognitive planning (Lin et al., 2017). They also show reduced sensitivity to targets during a sustained attention task (Diwadkar et al., 2011) and a pronounced attentional bias favoring affective over neutral stimuli (Bauer et al., 2015; Gotlib, Traill, Montoya, Joormann, & Chang, 2005). Adult relatives of BD patients displayed deficits in verbal memory, spatial working memory and processing speed (Calafiore, Rossell, & Van Rheenen, 2018). Unaffected adult siblings of BD patients with global premorbid and intellectual quotient (IQ) deficits perform poorly on a verbal memory task when compared to healthy volunteers (Russo et al., 2017). Our previous work using non-verbal tasks found that unaffected adult siblings of BD patients encountered difficulties on a task of visual sustained attention when compared to healthy volunteers but had intermediate levels of performance between healthy volunteers and BD (Bauer et al., 2016). A primary limitation of these studies was the focus on specific cognitive domains and the small sample size the heterogeneity of measures used in these studies. Thus, the reproducibility and generalizability of these findings is debatable.

Machine learning offers a set of tools that may be able to address some of these limitations. Learning algorithms are typically trained in a dataset to identify parameters able to

distinguish individual subjects across groups (BD vs HC). The algorithm with optimal parameters is then tested in an independent dataset to assess its accuracy and generalization ability (Dwyer, Falkai, & Koutsouleris, 2018). Our previous studies successfully applied machine learning algorithms to identify CANTAB cognitive measures identifying BD vs HC with a 71% accuracy (M.-J. Wu et al., 2016). More recently, we showed the high interpretability and suitability of the component-wise gradient boost (CGB) learning algorithm to build a predictive model for the onset of adolescent depression from a longitudinal data set of inflammatory proteins (Walss-Bass et al., 2018) and a model predicting aggression from a set of psychosocial and genetic variables (Suchting, Gowin, Green, Walss-Bass, & Lane, 2018; Walss-Bass, Suchting, Olvera, & Williamson, 2018). To date, however, no study has applied machine learning algorithms to identify cognitive profiles of children and adolescents with BD.

Inspired by our previous CANTAB and CGB work in adults with BD and depression using machine learning methods (M.-J. Wu, et al., 2016), this study aimed to identify the cognitive profile of pediatric BD using a CGB approach. We also decided to extend this work to offspring of BD parents and determine whether CGT could identify CANTAB measures that distinguished unaffected offspring of BD parents from pediatric BD and healthy controls.

Methods and materials

Subjects

Our sample included 59 healthy controls (HC; 11.19±3.15 years; 30 girls), 119 children and adolescents with BD (13.31±3.02 years; 52 girls) and 49 unaffected offspring of BD parents (UO; 9.36±3.18 years; 22 girls). Participants were recruited at the University of North Carolina at Chapel Hill (UNC) and at the University of Texas Health Science Center at Houston. The study protocol was approved by the local institutional review boards and informed consent was obtained from all the participants. Participants included in this study had no current medical disorder including neurological disorders and traumatic brain injury. Children and adolescents with BD and offspring of parents with BD had at least one parent who met criteria for BD as determined via a detailed family history assessment. Unaffected offspring of BD had not taken prescribed psychotropic medication at any point in their lives and were not biologically related to the children and adolescents with BD included in this study. Individuals with BD reported comorbidities such as attentional deficit hyperactivity disorder (ADHD; n=15), Anxiety Disorders including social phobia and generalized anxiety disorder (n=5), and conduct disorders (n=2). 95 out 119 patients with BD reported taking medications including antidepressants (n=26), antipsychotics (n=24), stimulants (n=6), and anti-convulsants (n=4) (Table 1). Healthy controls with an immediate family history of any Axis I disorder and/or who had taken any prescribed psychotropic medication at any point in their lives were excluded. Children and adolescents with history of substance abuse in the six months prior to enrollment, schizophrenia, developmental disorders, eating disorders and intellectual disability were also excluded. Female participants of reproductive age underwent a urine pregnancy test. The primary reasons for excluding pregnant participants were: 1. Cognitive performance was found to be poorer in pregnant women compared to control women, particularly during the third trimester (Davies, Lum, Skouteris, Byrne, & Hayden,

2018). 2. The effects of performing computerized cognitive tasks and associated tasks included in our studies to pregnant women and their unborn children are underexplored. All participants underwent a urine drug screen to exclude illegal drug use.

Clinical assessment

Psychiatric diagnosis was established using the Kiddie Schedule of Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL) interview (Kaufman, Birmaher, Brent, Rao, & Ryan, 1996) based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria, and confirmed subsequently in a clinical evaluation with a research psychiatrist. All parents (of individuals with BD and BD offspring) who reported previous BD diagnosis had their diagnosis ascertained by the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders Axis I (SCID I) (First, Spitzer, Gibbon, & Williams, 2012). All interviews were administered to participants by trained evaluators, and were later reviewed by a board-certified psychiatrist. The affective state was assessed with the Young Mania Rating Scale (YMRS; (Young, Biggs, Ziegler, & Meyer, 1978)) and the Children Depression Rating Scale (CDRS; Poznanski et al., 1984). Both instruments have satisfactory psychometric properties (YMRS: Cronbach α=.80, convergent validity: r=.83 (Fristad, Weller, & Weller, 1995); CDRS: Cronbach α=.85, item-total correlations ranged from .28 to .78, convergent validity: r= .92 (Poznanski and Mokros, 1996).

Cognitive assessment

CANTAB—Participants performed the computerized Cambridge Neurocognitive Test Automated Battery (CANTAB - http://www.cantab.com). This cognitive battery was chosen based on the established sensitivity to cognitive impairment in psychiatric disorders (28). In the current study we focused on tasks that have been previously shown to be impaired in BD when compared to HC (Bauer, et al., 2015; M.-J. Wu, Mwangi, Bauer, et al., 2017; M.-J. Wu, et al., 2016). Specifically, we administered the Affective Go/No-Go task (AGN), the Cambridge Gambling Task (CGT), Rapid Visual Processing (RVP), the Motor Screening (MOT), Big/Little Circle (BLC), Intra-Extradimensional Set shift (IED), working memory (Spatial Span task - SSP), and spatial memory (Spatial Recognition Memory – SRM). Variables of interest across tasks included reaction times and accuracy (number of correct responses, commission and omission errors). The tasks of the CANTAB included in the present study are described in Tables 2 and 2S.

Data Analytic Strategy

Exploratory analyses—Statistical analyses were performed using IBM SPSS statistics (Version 21.0) and the caret package (v. 6.0-79) in the R statistical computing environment. Normality assumptions were examined. One-way ANOVAs and χ^2 analyses were used to compare demographic and clinical differences between groups. All predictors were z-scored and all variables were on the same metric in terms of standard deviations (SD).

Component-Wise Gradient Boosting (CGB)—The present study utilized the component-wise gradient boosting (CGB; (Bühlmann and Hothorn, 2007)) machine learning algorithm as implemented in the *mboost* package (Hothorn et al., 2017) in the R statistical

computing environment (R Core Team, 2018) to build models that may classify our individuals as HC, BD, or UO based on a set of 30 cognitive test predictors and accounted for the covariates age, sex, education, and site. Specifically, for each comparison, the CGB algorithm was "free" to select from the entire set of cognitive predictors as well as the demographic covariates (i.e., age, sex, education, site). In other words, the algorithm could, for instance, determine whether age was relevant to the classification and choose to include it as a covariate or choose not to include it.

Table 1S includes a list of the candidate predictors. Following an initialization step, the CGB algorithm develops a series of models, each of which explains variability that was not explained by previous models (Friedman, 2001, 2002). Each constructed model in the series selects one predictor at a time, and after a certain number of iterations (an algorithm parameter called "*mstop*," tuned via K-fold cross validation using 75% of the original data for algorithm training), a final model is established and subsequently evaluated using the remaining 25% of the original data held out for algorithm testing. A shrinkage parameter, nu, was held at 0.1 by convention (Hofner, Mayr, Robinzonov, & Schmid, 2014). As the number of training iterations is finite, and any given predictor may be chosen as many times as the necessary, the tuned algorithm contains only those predictors that best explain variability in the outcome. Further, this iterative process is inherently penalized and addresses potential collinearity effects (Hofner, et al., 2014).

CGB models—CGB was used to develop four models, each with a different split of the diagnosis outcome: *model 1* examined three-class diagnosis (HC vs BD vs UO) using a multinomial logistic function, *model 2* classified HC vs UO, *model 3* classified BD vs UO, and *model 4* classified combined HC/UO vs BD. This modeling strategy was developed *a posteriori* based on the success of each model to classify based on diagnosis. Algorithm performance was evaluated by making predictions on the held-out test set. Classification success was measured using performance metrics including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC). Accuracy was assessed for each model by comparing frequencies of observed classification to predicted classification in a table called a confusion matrix. This matrix compared each model's accuracy (% correctly classified vs % incorrectly classified) to a baseline no information rate (NIR) that would be achieved from simply selecting the most frequent category in the observed data. Predicted classification accuracy was assessed for each model using a confusion matrix of predicted classification vs observed classification. This matrix compared each model's predicted accuracy to a given NIR that would be achieved from selecting the most frequent category in the observed data. The NIR is defined as the percentage of a classification variable that represents the most frequent category. A significance test established the degree to which the predicted accuracy was or was not superior to the NIR. For instance, when comparing 119 BD to 59 HC and 49 UO the NIR would be equal to 119 divided by the sum of 59, 119 and 49.

Results

Group characteristics

Demographics and clinical features of the participants included in this study are reported in Table 1. The three groups were comparable in terms of gender and ethnicity. As expected, BD's YMRS and CDRS scores were elevated compared to those of HC. BD reported higher education levels than HC and UO ($p<.05$). UO were younger than HC and BD ($p<.01$). Based on this result, age differences were accounted for in all CGB analyses. As illustrated in Table 2S correlation coefficients between the cognitive variables were overall below Pearson's r=.5 thus showing a weak linear relationship between variables. Means and standard deviations of the cognitive variables included in our analyses are presented in Table 3S.

Component-Wise Gradient Boosting

Model 1 – Multinomial model classifying HC *vs* BD *vs* UO—The CGB algorithm retained 17 predictors of the three-level diagnosis outcome. Optimization of the mstop parameter occurred at 946 iterations. HC was the reference category for this comparison and regression coefficients and odds ratios for UO and BD are provided in Table 3. The strongest retained cognitive predictors were Affective Go/No-Go (AGN) Mean Correct Latency (UO O.R. = 0.78; BD O.R. = 0.54) and Big Little Circle (BLC) Mean Correct Latency (UO O.R. = 0.91, BD O.R. = 1.27). In other words, one standard deviation increase in Affective Go/No Go Mean Correct Latency (i.e. slower cognitive reaction times) was related to a 22% decrease in the odds of being UO relative to HC, and a 46% decrease in the odds of being BD relative to HC.

One standard deviation increase in BLC mean correct latency (e.g. slower motor reaction times) reflected a 9% decrease in the odds of being UO and a 27% increase in the odds of being BD relative to HC.

53% of observations were correctly classified, with correct classification rates for HC = 33%, BD = 76%, and UO = 21%. This overall classification accuracy was, however, not significantly greater than the NIR of 52.7% ($p = 0.554$). Based on this result, selected pairwise splits of the three-level outcome variable were tested to further understand diagnosis.

Model 2 – Model classifying HC *vs* UO—The CGB algorithm retained 3 predictors of the binary outcome. Optimization of the mstop parameter occurred at 14 iterations. Table 4 provides the regression coefficients and odd ratios for predicting UO relative to the reference category HC.

The strongest retained predictor was AGN Total commission errors (O.R. = 0.77). This means that one standard deviation increase in AGN Total Commission errors was associated with a 23% decrease in the odds of being UO relative to HC.

Classification performance found that 61.5% of observations were correctly classified with an area under the receiver operating characteristic curve (AUROC) = 0.643. The predicted

accuracy was not greater than the NIR of 53.9% ($p = 0.279$). The algorithm was more successful in identifying HC (71.4%) than UO (50%).

Model 3 – Model classifying BD vs UO—The CGB algorithm retained 7 predictors of the binary outcome. Optimization of the *mstop* parameter occurred at 86 iterations. The strongest retained cognitive predictors were Affective Go/No-Go total commission errors (O.R. = 0.79) and Spatial Recognition Memory (SPM) Mean Correct Latency (O.R. = 1.05).

This means that one standard deviation increase in AGN total commission errors and SPM mean correct latency was linked to a 21% decrease and a 5% increase, respectively, in the odds of being UO relative to BD. Classification performance on the held-out test set found that 78% of observations were correctly classified, with AUROC = 0.902. This classification accuracy was not significantly greater than the NIR of 70.7% ($p = 0.197$). The algorithm was more successful in identifying BD (86.2%) than UO (58.3%) (Table 4).

Model 4 – Model classifying HC+UO vs BD—Given the lack of success differentiating between HC and UO by previous algorithms, the categories were combined and classified against BD. The CGB algorithm retained 12 predictors of the binary outcome. Optimization of the *mstop* parameter occurred at 64 iterations. HC+UO was the reference category for this comparison.

The strongest retained cognitive predictors were Affective Go/No-Go mean correct latency (O.R. = 0.78) and Rapid Visual Processing total correct responses (O.R. = 0.93). This means that one standard deviation increase in AGN mean correct latency reflected a 22% decrease in the odds of being BD. An increase in RVP total correct responses was associated with a 7% decrease in the odds of being BD.

Classification performance on the held-out test set found that 71% of observations were correctly classified, with AUROC = 0.797 (Figure 1). This classification accuracy was significantly greater than the NIR of 51.8% ($p \leq 0.002$). The algorithm was slightly worse at classifying HC+UO (59.3%) than BD (82.8%) (Table 4).

Discussion

In this study we used component-wise gradient boosting (CGB) to determine whether cognitive variables from the CANTAB battery could differentiate pediatric BD from unaffected offspring (UO) of BD parents and HC. This study aimed to expand on our work on how to better identify cognitive profiles of BD youth using machine learning algorithms (M.-J. Wu, Mwangi, Bauer, et al., 2017; M.-J. Wu, et al., 2016).

To our knowledge, the CGB approach has never been applied to cognitive measures in the field of psychiatry (Bühlmann and Hothorn, 2007). One of the advantages of CGB over other machine learning algorithms is the high interpretability of the model. This is due to the high degree of variable screening performed by the algorithm. Further, the “significance tests” in the abovementioned models rely on the accuracy rates of tuned algorithms as compared to the baseline no information rate (NIR) and odds ratios provide the best index of the unique effect of each variable (Hofner, et al., 2014). CGB has been previously used to

predict biological and psychosocial markers of psychiatric disorders (Suchting, et al., 2018; Walss-Bass, et al., 2018) an appears, therefore, to be a valuable tool to identify vulnerability markers in psychiatry.

Due to the low accuracy of Models 1, 2, and 3 we decided to explore the data further and combined HC and UO to classify them against BD. We made this decision for the following reasons. The lack of significant cognitive predictors able to distinguish UO and HC were in line with previous evidence showing that a subcluster of UO perform comparably to HC on a range of cognitive tasks (Peredo, Jomphe, Maziade, Paccalet, & Merette, 2018). Further, another study showed that UO may differ from HC on verbal rather than visuo-spatial tasks (Calafiore, et al., 2018) such as those included in the CANTAB. We therefore concluded that, from a cognitive viewpoint, our HC and UO were relatively similar and could be included in the same comparison group.

As shown in Model 4, the primary variables of interest were AGN Mean Correct Latency and Rapid Visual Processing total correct responses. The algorithm provided moderate ability to discriminate between HC+UO and BD across all possible probability thresholds (AUC = 0.797). The latter finding compares well to machine learning efforts in other studies in psychiatry. Some examples include AUC = 0.77 for discriminating between suicide attempters and non-attempters (Passos et al., 2016); AUC = 0.74 for predicting methamphetamine use relapse (Gowin, Ball, Wittmann, Tapert, & Paulus, 2015), and AUC = 0.76 for predicting heart failure six months before a clinical diagnosis (J. Wu, Roy, & Stewart, 2010). Further, our previous work using a least absolute shrinkage operator (LASSO) machine learning algorithm on CANTAB measures correctly distinguished adults with BD from HC with an accuracy of up to 92% (M.-J. Wu, et al., 2016)(M.-J. Wu, Mwangi, Bauer, et al., 2017). Partially consistent with the present findings, in our previous studies, the most relevant neurocognitive included the AGN correct latency times and AGN commission errors. Overall, these findings suggest that latencies in response to affective stimuli are promising markers of BD in both adult and pediatric populations.

Although CGB could not accurately classify UO *vs* HC, UO *vs* BD, and BD *vs* HC+UO above the required NIR level (see Models 1 to 3), it is noteworthy mentioning that Model 1 found that slower response times to affective stimuli reduced the odds of being BD relative to HC. However, slow motor processing speed (on the BCL) increased the odds of being BD rather than HC. These findings were partially in line with our previous work in BD offspring (Bauer, et al., 2015) which found that pediatric BD displayed impulsive and inaccurate responses the AGN task but not on a non-affective task (RVP). Thus, Taken together, findings from Models 1 and 4 support the hypothesis that impaired processing of emotionally salient information is a marker of vulnerability to BD (Murphy and Sahakian, 2001; M.-J. Wu, Mwangi, Passos, et al., 2017).

Further, Model 2 showed that a greater number of AGN commission errors reflected a decrease in the odds of being UO relative to HC. Along the same line, in our 2015 study, the performance of healthy BD offspring was comparable to that of HC. It has been previously shown that healthy high-risk individual (aged 14.0±2.4 years) succeeded in labelling face emotions but displayed a stronger amygdala response compared with BD and healthy

controls in response to fearful stimuli (Olsavsky et al., 2012). Thus, one could argue that healthy offspring have a unique pattern of response to affective stimuli due to potentially stronger fronto-limbic connectivity. There is, however, little longitudinal data available to provide evidence that this type of functional profile is associated with less or no cognitive deficits and offers protection against the development of mood disorders.

The current study has a number of limitations. UO were younger (9.36±3.18 years) than BD and HC. This means that these children's neurodevelopment and education (see Table 1) were not in line with that of the other groups. We addressed this issue by including the covariate "education" in our analyses. Further, their cognitive performance did not distinguish them from HC and BD. Imbalances across groups (e.g. 119 BD, 49 offspring) may have hindered the CGB performance to identify differences between BD and UO. However, boosting techniques (including CGB) are known to handle imbalanced classes better than traditional classifiers (García-Pedrajas and García-Osorio, 2013).

Another potential limitation is that 35 of our BD participants did not specify whether they were on psychotropic medication and we had no specific IQ information for all our participants. Hence, this type of information could not be suitably integrated in our analyses. It is important to note that IQ scores vary dramatically in the teenage years. In a 4-year follow-up study, up to 39% of adolescents aged 12 to 16 years showed shifts of +/− 23 points in verbal and performance IQ. These fluctuations were found to be closely related to maturational changes in the sensorimotor and cerebellar brain regions (Ramsden et al., 2011). These findings are important because they show that, in adolescents, cognitive performance may fluctuate due to age-related neural changes (Price, Ramsden, Hope, Friston, & Seghier, 2013). Although we had no specific information on the IQ of our participants, the educational achievement of our BD participants was greater than that of HC. Given the strong positive correlation between education and IQ (Brinch, Bratsberg, & Raaum, 2012)(Jonsson et al., 2017), it could be hypothesized that our BD participants had high IQ scores and the ability to perform comparably to, if not better than, HC. The cognitive differences observed between BD and HC and UO are therefore likely to originate from biological differences between groups rather than differences in intelligence *per se*.

From a methodological viewpoint, the CANTAB battery does not include tasks of verbal learning and memory, which are cognitive domains that have been found to be impaired in adult BD populations (Burdick et al., 2011; Van Rheenen and Rossell, 2014). Thus, the tasks used in the current study may have not been sufficiently specific and sensitive to detect subtle differences in performance between UO, HC and BD. One could also speculate that, as observed in adults with BD, pediatric BD and high-risk individuals, such as offspring and siblings of BD patients, may display a number of cognitive subprofiles (Russo, et al., 2017; Van Rheenen et al., 2017). Similarly, there is cognitive variability even among HC (Rabinowitz and Arnett, 2013) so outlining differences between high-risk and HC may pose a challenge.

A major strength of our study is that, unlike previous studies using samples of individuals vulnerable to go on to develop BD, our work included only unaffected offspring of BD parents *and* unrelated BD patients, thus circumventing the confounding effect of shared

environmental factors. Further, it used a novel machine learning algorithm, CGB, which provides highly interpretable and replicable findings. Our sample size was relatively large compared to previous offspring studies (n=49) and, given the young age of our participants, long-term medication effects were likely to be minimal.

In sum, although cognitive measures may not have “diagnostic power” as such, clinicians may benefit from these preliminary findings to focus on the strongest predictors in borderline cases where diagnostic criteria may be unclear. Future studies should consider adopting a longitudinal design to determine whether cognitive deficits appear just prior to the onset of the disease, and whether there is a critical age to predict the onset of the disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding

This work was partly supported by the Stanley Medical Research Institute, the Dunn Foundation, NIH grant MH085667 (JCS), and by the Pat Rutherford Jr. Chair in Psychiatry (UTHealth).

References

Akdemir D, & Gökler B (2008). Psychopathology in the Children of Parents with Bipolar Mood Disorder. Turkish Journal of Psychiatry, 19(2), pp. 133–140. [PubMed: 18561045]

Bauer IE, Frazier TW, Meyer TD, Youngstrom E, Zunta–Soares GB, & Soares JC (2015). Affective processing in pediatric bipolar disorder and offspring of bipolar parents. Journal of child and adolescent psychopharmacology, 25(9), pp. 684–690. [PubMed: 26468988]

Bauer IE, Hautzinger M, & Meyer TD (2017). Memory performance predicts recurrence of mania in bipolar disorder following psychotherapy: A preliminary study. Journal of psychiatric research, 84, pp. 207–213. [PubMed: 27764692]

Bauer IE, Wu M-J, Frazier T, Mwangi B, Spiker D, Zunta-Soares GB, & Soares JC (2016). Neurocognitive functioning in individuals with bipolar disorder and their healthy siblings: A preliminary study. Journal of affective disorders, 201, pp. 51–56. [PubMed: 27179338]

Bora E, Harrison B, Yücel M, & Pantelis C (2013). Cognitive impairment in euthymic major depressive disorder: a meta-analysis. Psychological medicine, 43(10), pp. 2017–2026. [PubMed: 23098294]

Brinch CN, Bratsberg B, & Raaum O (2012). The effects of an upper secondary education reform on the attainment of immigrant youth. Education Economics, 20(5), pp. 447–473.

Bühlmann P, & Hothorn T (2007). Boosting algorithms: Regularization, prediction and model fitting. Statistical Science, pp. 477–505.

Burdick KE, Goldberg JF, Harrow M, Faull RN, & Malhotra AK (2006). Neurocognition as a stable endophenotype in bipolar disorder and schizophrenia. The Journal of nervous and mental disease, 194(4), pp. 255–260. [PubMed: 16614546]

Burdick KE, Goldberg TE, Cornblatt BA, Keefe RS, Gopin CB, DeRosse P, … Malhotra AK (2011). The MATRICS consensus cognitive battery in patients with bipolar I disorder. Neuropsychopharmacology, 36(8), pp. 1587–1592. [PubMed: 21451499]

Calafiore D, Rossell SL, & Van Rheenen TE (2018). Cognitive abilities in first-degree relatives of individuals with bipolar disorder. Journal of affective disorders, 225, pp. 147–152. [PubMed: 28829959]

Chang K, Steiner H, Dienes K, Adleman N, & Ketter T (2003). Bipolar offspring: a window into bipolar disorder evolution. Biological psychiatry, 53(11), pp. 945–951. [PubMed: 12788239]

Davies SJ, Lum JA, Skouteris H, Byrne LK, & Hayden MJ (2018). Cognitive impairment during pregnancy: a meta-analysis. Medical Journal of Australia, 208(1), pp. 35–40. [PubMed: 29320671]

de la Serna E, Sugranyes G, Sanchez-Gistau V, Rodriguez-Toscano E, Baeza I, Vila M, … Moreno D (2017). Neuropsychological characteristics of child and adolescent offspring of patients with schizophrenia or bipolar disorder. Schizophrenia research, 183, pp. 110–115. [PubMed: 27847227]

Diwadkar VA, Goradia D, Hosanagar A, Mermon D, Montrose DM, Birmaher B, … Amirsadri A (2011). Working memory and attention deficits in adolescent offspring of schizophrenia or bipolar patients: comparing vulnerability markers. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 35(5), pp. 1349–1354. [PubMed: 21549798]

Dwyer D, Falkai P, & Koutsouleris N (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. Annual review of clinical psychology, 14, pp. 91–118.

First MB, Spitzer RL, Gibbon M, & Williams JB (2012). Structured Clinical Interview for DSM-IVo Axis I Disorders (SCID-I), Clinician Version, Administration Booklet: American Psychiatric Pub.

Fristad MA, Weller RA, & Weller EB (1995). The Mania Rating Scale (MRS): further reliability and validity studies with children. Annals of Clinical Psychiatry, 7(3), pp. 127–132. [PubMed: 8646272]

García-Pedrajas N, & García-Osorio C (2013). Boosting for class-imbalanced datasets using genetically evolved supervised non-linear projections. Progress in Artificial Intelligence, 2(1), pp. 29–44.

Glahn DC, Bearden CE, Niendam TA, & Escamilla MA (2004). The feasibility of neuropsychological endophenotypes in the search for genes associated with bipolar affective disorder. Bipolar Disorders, 6(3), pp. 171–182. [PubMed: 15117396]

Gotlib IH, Traill SK, Montoya RL, Joormann J, & Chang K (2005). Attention and memory biases in the offspring of parents with bipolar disorder: indications from a pilot study. Journal of Child Psychology and Psychiatry, 46(1), pp. 84–93. [PubMed: 15660646]

Gowin JL, Ball TM, Wittmann M, Tapert SF, & Paulus MP (2015). Individualized relapse prediction: Personality measures and striatal and insular activity during reward-processing robustly predict relapse. Drug & Alcohol Dependence, 152, pp. 93–101. [PubMed: 25977206]

Hofner B, Mayr A, Robinzonov N, & Schmid M (2014). Model-based boosting in R: a hands-on tutorial using the R package mboost. Computational Statistics, 29(1-2), pp. 3–35.

Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A, Scheibe S, & Hothorn MT (2017). Package 'multcomp': Obtained from http://cran.stat.sfu.ca/web/packages/multcomp/multcomp.pdf.

Jonsson B, Waling M, Olafsdottir AS, Lagström H, Wergedahl H, Olsson C, … Gunnarsdottir I (2017). The Effect of Schooling on Basic Cognition in Selected Nordic Countries. Europe's journal of psychology, 13(4), pp. 645–666.

Kaufman J, Birmaher B, Brent D, Rao U, & Ryan N (1996). Kiddie-Sads-present and Lifetime version (K-SADS-PL). Pittsburgh, University of Pittsburgh, School of Medicine

Lapalme M, Hodgins S, & LaRoche C (1997). Children of parents with bipolar disorder: a metaanalysis of risk for mental disorders. The Canadian Journal of Psychiatry, 42(6), pp. 623–631. [PubMed: 9288425]

Lin K, Lu R, Chen K, Li T, Lu W, Kong J, & Xu G (2017). Differences in cognitive deficits in individuals with subthreshold syndromes with and without family history of bipolar disorder. Journal of psychiatric research, 91, pp. 177–183. [PubMed: 28521253]

Martínez-Arán A, Vieta E, Reinares M, Colom F, Torrent C, Sánchez-Moreno J, … Salamero M (2004). Cognitive function across manic or hypomanic, depressed, and euthymic states in bipolar disorder. American Journal of Psychiatry, 161(2), pp. 262–270. [PubMed: 14754775]

Mesman E, Nolen WA, Reichart CG, Wals M, & Hillegers MH (2013). The Dutch bipolar offspring study: 12-year follow-up. American Journal of Psychiatry, 170(5), pp. 542–549. [PubMed: 23429906]

Mora E, Portella M, Forcada I, Vieta E, & Mur M (2013). Persistence of cognitive impairment and its negative impact on psychosocial functioning in lithium-treated, euthymic bipolar patients: a 6-year follow-up study. Psychological medicine, 43(6), pp. 1187–1196. [PubMed: 22935452]

Murphy F, & Sahakian B (2001). Neuropsychology of bipolar disorder. The British Journal of Psychiatry, 178(S41), pp. s120–s127. [PubMed: 11388950]

Olsavsky AK, Brotman MA, Rutenberg JG, Muhrer EJ, Deveney CM, Fromm SJ, … Leibenluft E (2012). Amygdala Hyperactivation During Face Emotion Processing in Unaffected Youth at Risk for Bipolar Disorder. Journal of the American Academy of Child & Adolescent Psychiatry, 51(3), pp. 294–303. [PubMed: 22365465]

Passos IC, Mwangi B, Cao B, Hamilton JE, Wu M-J, Zhang XY, … Kapczinski F (2016). Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. Journal of affective disorders, 193, pp. 109–116. [PubMed: 26773901]

Peredo R, Jomphe V, Maziade M, Paccalet T, & Merette C (2018). Cluster analysis identifies two cognitive profiles among offspring of patients with a major psychiatric disorder: The healthy and impaired profiles. Journal of Child and Adolescent Psychiatry, 2(2), pp. 6–11.

Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, & Gibbons R (1984). Preliminary studies of the reliability and validity of the Children's Depression Rating Scale. Journal of the American Academy of Child Psychiatry, 23(2), pp. 191–197. [PubMed: 6715741]

Poznanski EO, & Mokros HB (1996). Children's depression rating scale, revised (CDRS-R): manual: Western Psychological Services.

Price C, Ramsden S, Hope T, Friston K, & Seghier M (2013). Predicting IQ change from brain structure: a cross-validation study. Developmental cognitive neuroscience, 5, pp. 172–184. [PubMed: 23567505]

R Core Team. (2018). A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL https://www.R-project.org/. Retrieved

Rabinowitz AR, & Arnett PA (2013). Intraindividual cognitive variability before and after sports-related concussion. Neuropsychology, 27(4), pp. 481–490. [PubMed: 23876120]

Ramsden S, Richardson FM, Josse G, Thomas MS, Ellis C, Shakeshaft C, … Price CJ (2011). Verbal and non-verbal intelligence changes in the teenage brain. Nature, 479(7371), p 113. [PubMed: 22012265]

Robinson LJ, Thompson JM, Gallagher P, Goswami U, Young AH, Ferrier IN, & Moore PB (2006). A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. Journal of affective disorders, 93(1-3), pp. 105–115. [PubMed: 16677713]

Roux P, Raust A, Cannavo A-S, Aubin V, Aouizerate B, Azorin J-M, … Cussac I (2017). Associations between residual depressive symptoms, cognition, and functioning in patients with euthymic bipolar disorder: results from the FACE-BD cohort. The British Journal of Psychiatry, 211(6), pp. 391–387.

Russo M, Van Rheenen T, Shanahan M, Mahon K, Perez-Rodriguez M, Cuesta-Diaz A, … Burdick K (2017). Neurocognitive subtypes in patients with bipolar disorder and their unaffected siblings. Psychological medicine, 47(16), pp. 2892–2905. [PubMed: 28587689]

Suchting R, Gowin JL, Green CE, Walss-Bass C, & Lane SD (2018). Genetic and Psychosocial Predictors of Aggression: Variable Selection and Model Building with Component-Wise Gradient Boosting. Frontiers in Behavioral Neuroscience, 12, p 89. [PubMed: 29867390]

Tsuang MT, Tohen M, & Jones P (2011). Textbook of psychiatric epidemiology: John Wiley & Sons.

Van Rheenen TE, Lewandowski K, Tan E, Ospina L, Ongur D, Neill E, … Rossell S (2017). Characterizing cognitive heterogeneity on the schizophrenia–bipolar disorder spectrum. Psychological medicine, 47(10), pp. 1848–1864. [PubMed: 28241891]

Van Rheenen TE, & Rossell SL (2014). Investigation of the component processes involved in verbal declarative memory function in bipolar disorder: utility of the Hopkins Verbal Learning Test-Revised. Journal of the International Neuropsychological Society, 20(7), pp. 727–735. [PubMed: 24870365]

Walss-Bass C, Suchting R, Olvera RL, & Williamson DE (2018). Inflammatory markers as predictors of depression and anxiety in adolescents: Statistical model building with component-wise gradient boosting. Journal of affective disorders, 234, pp. 276–281. [PubMed: 29554616]

Wilde A, Chan H-N, Rahman B, Meiser B, Mitchell P, Schofield P, & Green M (2014). A meta-analysis of the risk of major affective disorder in relatives of individuals affected by major depressive disorder or bipolar disorder. Journal of affective disorders, 158, pp. 37–47. [PubMed: 24655763]

Wu J, Roy J, & Stewart WF (2010). Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical care, 48(6), pp. S106–S113. [PubMed: 20473190]

Wu M-J, Mwangi B, Bauer IE, Passos IC, Sanches M, Zunta-Soares GB, … Soares JC (2017). Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. Neuroimage, 145, pp. 254–264. [PubMed: 26883067]

Wu M-J, Mwangi B, Passos IC, Bauer IE, Cao B, Frazier TW, … Soares JC (2017). Prediction of vulnerability to bipolar disorder using multivariate neurocognitive patterns: a pilot study. International journal of bipolar disorders, 5(1), p 33. [PubMed: 28921165]

Wu M-J, Passos IC, Bauer IE, Lavagnino L, Cao B, Zunta-Soares GB, … Soares JC (2016). Individualized identification of euthymic bipolar disorder using the Cambridge Neuropsychological Test Automated Battery (CANTAB) and machine learning. Journal of affective disorders, 192, pp. 219–225. [PubMed: 26748737]

Young R, Biggs J, Ziegler V, & Meyer D (1978). A rating scale for mania: reliability, validity and sensitivity. The British Journal of Psychiatry, 133(5), pp. 429–435. [PubMed: 728692]

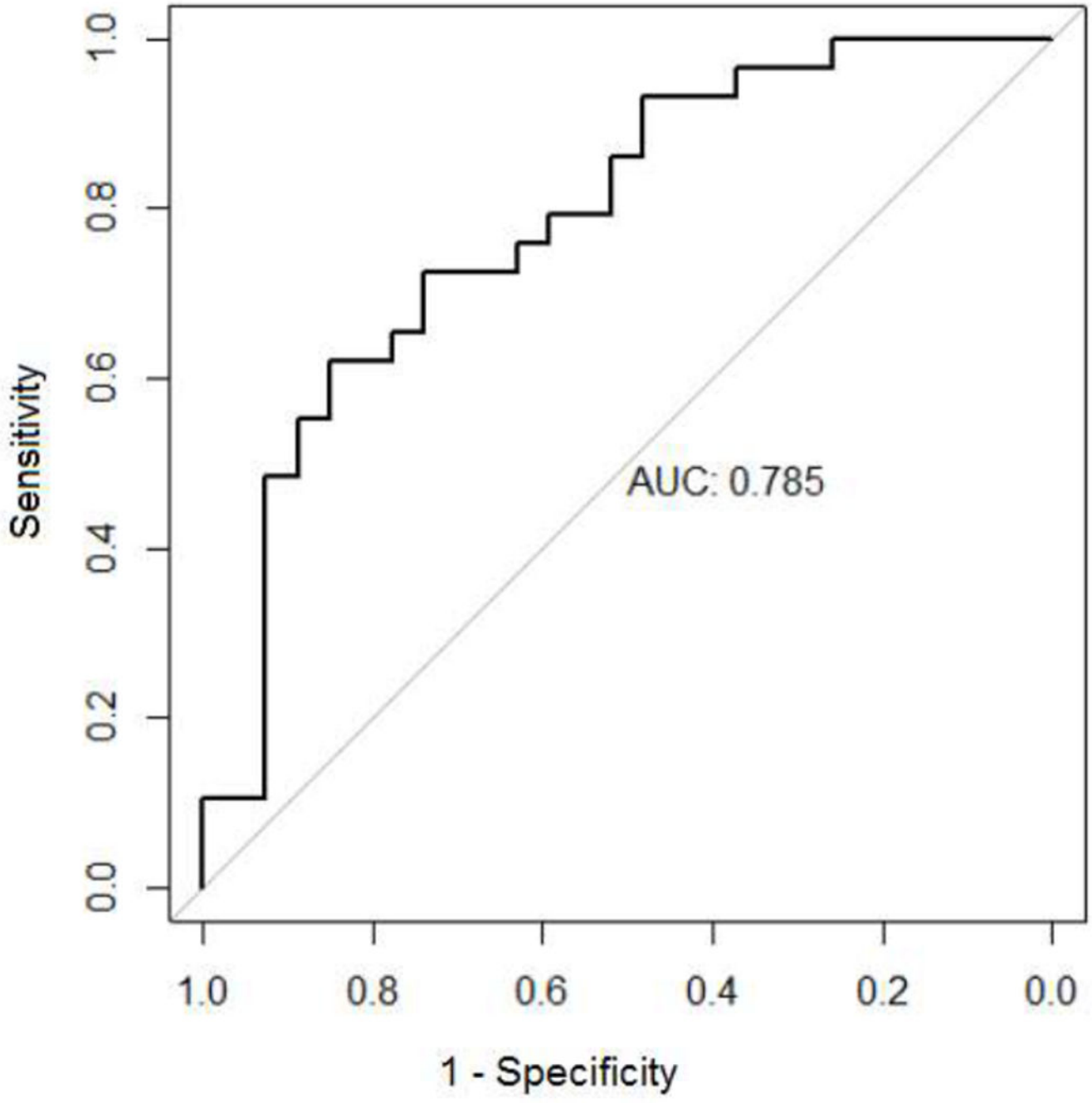

Figure 1.
Area Under the Receiver Operating Characteristic Curve (AUROC) for the CGB model classifying HC+UO vs BD

Table 1:

Demographic and Clinical Characteristic of the Sample (mean ± standard deviation).

	HC mean (SD)	BD Mean (SD)	Offspring Mean (SD)	F/chi-square	p-value
Age (years)	11.19±3.15	13.31±3.02	9.36±3.18	30.58	.00 **#, ##*
Female/total	30/59	52/119	22/49	.83	.66
Bipolar type	-	53 BDI 12 BDII 54 BD-NOS	-		
Education (years)	5.17±3.024	6.92±2.93	3.47±3.163	24.27	.00 **#,##*
Ethnicity	25 C 7 H 18 AA 3 A 6 Multiracial	44 C 25 H 37 AA 13 Multiracial	18 C 8 H 17 AA 6 Multiracial	11.1	.2
YMRS	.34±.76	8.4±7	.48±1.03	58.04	.00 **#*
CDRS	17.44±1.12	27.33.72±7.4	17.56±1.42	14.1	.00 **#*
Age of onset (years)	-	10.75±3.14	-		
Most recent mood episode	-	41 depressed 5 mixed 4 manic 8 hypomanic	-		
Currently or previously taken any psychotropic medicine (N/total)	-	95/119 ***** 4 anti-convulsant 26 antidepressant 24 antipsychotics 6 stimulants.	-		
Primary Comorbidities	-	15 ADHD 2 Conduct Disorder 5 Anxiety Disorders	-		

Abbreviations: A:Asian; AA: African American; BD: Bipolar Disorder; C: Caucasian; GAF: Global Assessment of Functioning; GAD: Generalized Anxiety Disorder; H: Hispanic; HC: Healthy Controls;

Comparisons:

#: BD vs HC,

##: BD vs offspring;

*** p<.05,

**** p<.01

***** 35 participants were taking medication at the time of testing but did not specify whether it was a psychotropic treatment.

Table 2.

Cognitive tasks and measurements.

No.	CANTAB Task	Evaluation	Measurements
1	Affective Go/No-Go (AGN)	Affective and Cognitive control	Reaction time, accuracy
2	Big/Little Circle (BLC)	Motor speed	Reaction time, accuracy
3	Cambridge Gambling Task (CGT)	Decision-making	Reaction time, accuracy, proportion bets across trials with more/equally/less likely outcome
5	Motor Screening (MOT)	Motor processing speed	Reaction time
6	Match to Sample Visual Search (MTS)	Visuo-motor speed	Reaction time, accuracy
7	Rapid Visual Processing (RVP)	Sustained attention	Reaction time, accuracy
8	Spatial Recognition Memory (SRM)	Spatial memory	Reaction time, accuracy
9	Intra-Extradimensional Set shift (IED)	Attention, cognitive flexibility	Accuracy, number of trials and stages completed
10	Spatial Span task (SST)	Spatial working memory	Span length, number of attempts, reaction times

Table 3.

List of the most relevant cognitive variables in predicting membership status (Healthy Controls (HC), Bipolar Disorder (BD), Unaffected Offspring (UO)) sorted by magnitude from highest to lowest. The predicted category is indicated by column heading compared to the reference category (HC). Positive regression coefficients indicate higher scores in HC patients as compared to BD and UO. Negative coefficients suggest lower scores in HC compared to UO and BD patients.

	Regression Coefficient		Odds Ratio	
Variable	**UO**	**BD**	**UO**	**BD**
Age	−0.542	1.000	0.582	2.718
Site	−0.094	0.652	0.910	1.920
Affective Go/No-Go Mean Correct Latency	−0.278	−0.621	0.757	0.538
Big Little Circle Mean Correct Latency	−0.093	0.238	0.912	1.268
Motor Screening Mean Latency	−0.103	0.234	0.902	1.263
Cambridge Gambling Task Deliberation Time	−0.028	0.206	0.972	1.228
Rapid Visual Processing Mean Latency	0.162	−0.017	1.176	0.983
Rapid Visual Processing Total Correct Responses	−0.149	−0.298	0.862	0.743
Spatial Recognition Memory Mean Correct Latency	0.113	−0.046	1.119	0.955
Big Little Circle Percent Correct	0.111	−0.134	1.118	0.875
Affective Go/No-Go Total Commission Errors	−0.361	0.110	0.697	1.116
Cambridge Gambling Task Risk Adjustment	−0.108	−0.124	0.897	0.883
Intra/Extradimensional Completed Stage Trials	−0.009	0.103	0.991	1.109
Spatial Recognition Memory Percent Correct	−0.094	−0.237	0.910	0.789
Sex	−0.055	0.056	0.946	1.058
Education	−0.020	0.030	0.980	1.030
Cambridge Gambling Task Quality of Decision Making	−0.015	−0.095	0.985	0.910

Table 4.

List of the most relevant cognitive variables in predicting membership status (Healthy Controls (HC), Bipolar Disorder (BD), Unaffected Offspring (UO)) sorted by magnitude from highest to lowest. The predicted category is indicated by column heading compared to the reference category (HC). Positive boosted regression coefficients indicate higher scores in HC patients as compared to BD and UO. Negative boosted coefficients suggest lower scores in HC compared to UO and BD patients.

Comparison	Variable	Regression Coefficient	Odds Ratio
HC vs **UO**	Affective Go/No-Go Total Commission Errors	−0.268	0.765
	Intra/Extradimensional Stages Completed	−0.098	0.907
	Cambridge Gambling Task Risk Adjustment	−0.049	0.952
BD vs **UO**	Age	−0.333	0.717
	Affective Go/No-Go Total Commission Errors	−0.235	0.790
	Education	−0.089	0.915
	Spatial Recognition Memory Mean Correct Latency	0.046	1.047
	Rapid Visual Processing Mean Latency	0.031	1.032
	Affective Go/No-Go Mean Correct Latency	0.027	1.028
	Motor Screening Mean Latency	−0.012	0.988
HC/UO vs **BD**	Age	0.471	1.601
	Affective Go/No-Go Mean Correct Latency	−0.251	0.778
	Rapid Visual Processing Total Correct Responses	−0.074	0.929
	Cambridge Gambling Task Quality of Decision Making	−0.070	0.932
	Big Little Circle Mean Correct Latency	0.063	1.065
	Education	0.061	1.063
	Intra/Extradimensional Pre-ED Errors	0.044	1.045
	Spatial Span Task Span Length	−0.037	0.964
	Motor Screening Mean Latency	0.035	1.036
	Spatial Recognition Memory Mean Correct Latency	−0.023	0.978
	Rapid Visual Processing “A” Responses	−0.022	0.978
	Site	0.012	1.012