HHS Public Access

Author manuscript

*J Affect Disord*. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

*J Affect Disord*. 2018 July ; 234: 276–281. doi:10.1016/j.jad.2018.03.006.

# Inflammatory Markers as Predictors of Depression and Anxiety in Adolescents: Statistical Model Building with Component-Wise Gradient Boosting

**Consuelo Walss-Bass**[1], **Robert Suchting**[1], **Rene L. Olvera**[2], and **Douglas E. Williamson**[3]

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

[2]Division of Child, Adolescent and Adult Psychiatry, University of Texas Health Science Center at San Antonio, TX USA

[3]Division of Translational Neuroscience, Department of Psychiatry and Behavioral Sciences, Duke University and the Durham VA Medical Center Durham, NC USA

## Abstract

**Background—**Immune system abnormalities have been repeatedly observed in several psychiatric disorders, including severe depression and anxiety. However, whether specific immune mediators play an early role in the etiopathogenesis of these disorders remains unknown.

**Methods—**In a longitudinal design, component-wise gradient boosting was used to build models of depression, assessed by the Mood-Feelings Questionnaire-Child (MFQC), and anxiety, assessed by the Screen for Child Anxiety Related Emotional Disorders (SCARED) in 254 adolescents from a large set of candidate predictors, including sex, race, 39 inflammatory proteins, and the interactions between those proteins and time. Each model was reduced via backward elimination to maximize parsimony and generalizability.

**Results—**Component-wise gradient boosting and model reduction found that female sex, growth-regulated oncogene (GRO), and transforming growth factor alpha (TGF-alpha) predicted depression, while female sex predicted anxiety.

*Corresponding Author: University of Texas Health Science Center at Houston Department of Psychiatry and Behavioral Sciences, 1941 East Rd., Houston TX, 77054, USA. Tel.: +713 486 2521. Robert.Suchting@uth.tmc.edu.



**AUTHOR DISCLOSURE**

**CONTRIBUTORS**

Consuelo Walss-Bass conceived the study, performed the cytokine measures and wrote the first draft of the manuscript.
Robert Suchting performed the primary statistical analyses, including the component-wise gradient boosting and model reduction, and authorship of the data analytic strategy and results sections.
Rene L. Olvera contributed to psychological assessments of adolescents.
Douglas E. Williamson provided plasma samples from adolescents and psychological assessments.
All authors have approved the final article.

Conflicts of interest:
none.

**Limitations—**Differential onset of puberty as well as a lack of control for menstrual cycle may also have been responsible for differences between males and females in the present study. In addition, investigation of all possible nonlinear relationships between the predictors and the outcomes was beyond the computational capacity and scope of the present research.

**Conclusions—**This study highlights the need for novel statistical modeling to identify reliable biological predictors of aberrant psychological behavior.

### Keywords

inflammation; cytokines; adolescence; depression; anxiety; machine learning

## INTRODUCTION

Inflammatory signaling molecules such as cytokines and chemokines play important roles in brain processes such as synaptic plasticity, neurogenesis, memory, and cognition. Therefore, a faulty immune signaling mechanism during brain development may have long term consequences that may lead to the onset of psychiatric disorders. Indeed, immune system dysregulation has been repeatedly reported in a variety of mental disorders, including major depressive disorder (MDD) and anxiety markers (Dowlati et al., 2010; Howren et al., 2009; Raison and Miller, 2011). However, most of these studies have been conducted with chronically ill adults, who have suffered from the disorder for many years. Therefore, it is not clear whether abnormalities found in patients precede the onset of illness, emerge during early illness development, or follow disease onset.

Although the etiology of MDD and anxiety are unknown, studies indicate they may have a neurodevelopmental basis, with initial symptoms emerging during childhood and early adolescence (Galecki and Talarowska, 2018; Kalin, 2017). It is therefore important to clearly understand the role that the immune system plays in regulation of behavior during these key developmental years. Blood circulating inflammatory molecules can cross the blood brain barrier and can therefore serve as markers of a high inflammatory state that could potentially affect brain function. A recent study of adolescents diagnosed with bipolar disorder and MDD found increased levels of peripheral inflammatory markers, as well as increased activation of NFkB in peripheral blood mononuclear cells from patients compared to controls (Miklowitz et al., 2016). A large prospective study of children from a birth cohort found that serum levels of IL-6 at 9 years of age are associated with depressive and psychotic symptoms at 18 (Khandaker et al., 2014), and persistent depressive symptoms between 10 and 19 years of age (Khandaker et al., 2017). To our knowledge no previous study has comprehensively examined the relationship between a large panel of peripheral inflammatory markers and behavioral paradigms in adolescents who have not been diagnosed with a psychiatric disorder. We hypothesized that circulating inflammatory molecules could be promising biomarkers to monitor early behavioral alterations. By performing a 2 year longitudinal study of adolescents between the ages of 12 and 15 years at baseline and utilizing novel analytical strategies, we investigated the validity of peripheral levels of inflammatory markers in predicting development of anxiety and depression.

# METHODS

## Subjects

This study was approved by the University of Texas Health Science Center at San Antonio (UTHSCSA) and carried out in accordance with the Declaration of Helsinki. All participants provided written informed consent. Adolescents ages 12–15 at baseline were enrolled at the UTHSCSA Department of Psychiatry and assessed in two subsequent annual visits. Blood samples were collected at each time point, followed by centrifugation at 3,000 rpm for 10 min for isolation of plasma, which was aliquoted and stored at −80°C until analysis.

## Behavioral Measures

**Schedule for Affective Disorders and Schizophrenia for School Aged Children - Present and Lifetime Version (K-SADS -PL) (Kaufman et al., 1997)—**The K-SADS-PL diagnostic interview was used to provide assessments of present episode and lifetime history of psychiatric illness according to DSM-IV criteria. Subjects were excluded it they had a diagnosis of any psychiatric disorder, or present behavioral episodes, or had prior history of significant neurological disorder, head trauma, mental retardation or recent substance abuse. The K-SADS-PL was administered by interviewing the parent(s) first, then interviewing the child alone, and finally achieving summary ratings which included all sources of information.

**Mood-Feelings Questionnaire-Child (MFQC) (Angold et al., 1995)—**The MFQC is a 32-item child and parent-report scale was used to assess depressive symptomatology. Each item is scored on a scale from 0 (not true) to 2 (true).

**Screen for Child Anxiety Related Emotional Disorders (SCARED) (Birmaher et al., 1997)—**The SCARED is a 41-item child and parent self-report instrument was used to assess DSM IV symptoms of panic, separation anxiety, social phobia, general anxiety disorders, and symptoms of school refusal. Each item is scored on a scale from 0 (not true/hardly ever true) to 2 (very true/often true).

**Measurement of cytokine levels in plasma—**Fifty microliters of plasma were used for the assessment of 39 cytokines and chemokines using Millipore bead-based flow immunoassays (Billerica, MA) in the Luminex FlexMap 3D system (Austin, TX), according to the manufacturer's instructions. The cytokines assessed were the following: EGF, eotaxin, FGF2, Flt3L, fractalkine, GCSF, GMCSF, GRO, IFNα2, IFNγ, IL1α, IL1β, IL1RA, IL2, IL3, IL4, IL5, IL6, IL7, IL8, IL9, IL10, IL12p40, IL12p70, IL13, IL15, IL17, IP10, MCP1, MCP3, MDC, MIP1α, MIP1β, sCD40L, SIL-2R, TGFα, TNFα, TNFβ, and VEGF. All samples were analyzed in duplicate. CV's for each analyte were under 20%. Interplate calibration was performed for each analyte by normalizing the average of each sample to the average of a reference standard included in each plate. Normalized Cytokine values were z-scored by subtracting the overall sample mean of each analyte from each sample value and then dividing by the standard deviation. Outliers in the present study were defined, identified, and handled as outliers following recommendations from literature (Aguinis et al., 2013). As the outliers here were likely accurate but not interesting to the present context

(these values are so extreme they are thought to represent a distinct subpopulation outside the scope of the present research), they were defined as influential model fit outliers. Extreme outliers may result from somatic processes other than psychopathology (e.g., infection). These individuals could potentially constitute a distinct subpopulation not represented by the indices of central tendency in this sample's dependent variables. Follow-up analyses revealed that models including these values demonstrated worse fit. Outliers were handled via deletion: z-scores outside the range of −3.5 to +3.5 (94 of 780 observations) were considered outliers and removed. Outliers removed from the data were more male (n = 54) than female (n = 40) with extreme maximum z-scores (z > 9) found across the predictors. Analyses were repeated using the full data for comparison (see supplementary material).

## Data Analytic Strategy

**Overview—**The present study adopted a data-driven exploration of depression and anxiety in adolescents. This exploration utilized a data science approach to model building, including guided automation of pattern discovery using machine learning algorithms. Data science techniques have two primary goals: optimizing prediction and maximizing knowledge discovery; the present research emphasizes the latter. Results of such research bolster understanding of the nature of the relationships between variables, and may be viewed as hypothesis generating. The workflow of the present study used a two-step process to identify the strongest predictors of each outcome: (1) fit an optimized predictive model of each outcome using a boosting algorithm and (2) reducing that model via backward elimination. This two-step workflow has demonstrated utility in recent exploratory research to find the strongest predictors of time to lapse during a cigarette smoking quit attempt in a sample of rural smokers (Suchting et al., 2017).

**Component-wise Gradient Boosting—**We used the component-wise gradient boosting (CGB) algorithm as implemented in the mboost package in R, version 2.6-0 (Hothorn et al., 2016) to generate predictive models of depression and anxiety in adolescents. Detailed and accessible descriptions of CGB exist elsewhere (Bühlmann and Hothorn, 2007; Hofner et al., 2014); here we provide a brief overview. The algorithm was designed as an alternative formulation of boosting algorithms (Friedman, 2001, 2002), a technique from the machine learning literature that iteratively builds a strong predictive model from a collection (ensemble) of weak models via gradient descent. After an initialization step, the CGB algorithm constructs a series of models, each of which explains the variability that was not explained by prior models. Predictors in CGB models may take multiple functional forms including linear fixed and random effects as well as nonlinear (i.e., smoothing spline) effects. Selection of predictors in these various functional forms occurs in the course of the iterative procedure with resulting coefficients made more robust by penalization; such shrinkage techniques may stabilize effect estimates and reduce potential complications from multicollinearity (Hofner et al., 2014). While the current outcome is lognormal, the algorithm is equipped to utilize a variety of distributions. K-fold cross validation determines the number of iterations resulting in the optimized model. Mboost was chosen over other machine learning algorithms (e.g., elastic net, random forest) for its ability to account for

correlations between repeated observations by explicitly including random effects for participant and time.

Algorithm parameters were chosen by convention, such as shrinkage parameter nu = 0.1 (Hofner et al., 2014), or 10-fold cross-validation to prevent overfitting (stopping parameter mstop = 1919 for depression and 2387 for anxiety). Each of the base-learners, including a global intercept term, were fit as linear components (as opposed to nonlinear components, i.e., splines), with the exception of two random components (participant ID and time).

**Model Reduction—**The final optimized model chosen via component-wise gradient boosting features regularized parameter estimates and inherent variable selection. This model may then be reduced to maximize parsimony (and therefore generalizability) by engaging in backwards elimination (James et al., 2013; Kuhn and Johnson, 2013). This technique was chosen for its ability to reduce the model attained via CGB, rather than building a new model, and for its ability to provide a model with a more attractive parameter-to-sample size ratio. Base-learners selected by the CGB algorithm were fit using the lme() function of the nlme R package (Pinheiro et al., 2016) for multilevel modeling in R to establish a baseline model fit for reduction and ensure inclusion of main effects for the interaction effects selected by the boosting procedure. Backward elimination was performed using the StepAIC() function of the MASS package in R (Venables and Ripley, 2002; Zhang, 2016). All analyses were performed in the R statistical computing environment (R Core Team, 2016).

## RESULTS

### Demographic Characteristics

This study included psychometric and blood measures from adolescents measured across three annual (roughly) time points. The sample size, average age, and frequencies of sex and race at each time point are described in Table 1.

### Descriptive Statistics

Our statistical modeling included an exploration of 83 predictors (sex, race, time, 39 cytokines/chemokines, the interaction of each inflammatory marker with time, and random effects of participant id and time) to predict depression and anxiety across time. Participant age was used as a continuous measure of time in all analyses. The log forms of depression and anxiety were used as outcomes. Means and standard deviations are described in supplementary Table 1 (Table S1).

### Models for Depression and Anxiety

After tuning, the component-wise gradient boosting algorithm derived regularized models of both depression and anxiety using the full set of base-learners. This model was then reduced via backward elimination using the stepAIC() function of the MASS R package (Venables and Ripley, 2002). The squared correlation between predicted and observed values was then calculated for each model by the squared correlation between predicted and observed values (model comparison). The squared correlation as well as Akaike information criteria (AIC, a

measure of model fit that accounts for model complexity i.e., the number of parameters in the model) values for the full baseline and reduced multilevel models suggest that the models provide comparable fit. Relevant statistics for the depression and anxiety models are included in Tables 2 and 3, including penalized coefficients for the selected base-learners from the boosted model, fixed effects from the full baseline and reduced models, and model fit statistics for comparison. Plots of predicted versus observed values of the final reduced models for depression and anxiety are shown in Supplemental Figures S1A and B, respectively. The reduced models found that female sex predicted higher levels of depression and anxiety. Transforming growth factor alpha (TGF-alpha) predicted higher levels of depression, while growth-regulated oncogene (GRO) predicted lower levels of depression (Table 2, Figure 1). No statistically reliable effects of inflammatory markers were found for the prediction of anxiety (Table 3).

## DISCUSSION

There is significant evidence demonstrating elevated levels of peripheral cytokines in patients with depression and anxiety (Dantzer, 2006; Dantzer and Kelley, 2007; Garcia-Bueno et al., 2008; Raison et al., 2006; Young et al., 2014). In regards to depression, some studies suggest that pro-inflammatory cytokines may contribute to maladaptation to adverse life events that trigger the disorder (Rantala et al., 2017). However, the precise interpretation of the role of cytokines in development of depression and anxiety is limited by the fact that most studies have been conducted cross-sectionally in individuals who have been suffering from these disorders for a large number of years. In attempts to partially address this limitation in knowledge, we performed a comprehensive longitudinal evaluation of levels of a large panel of inflammatory molecules in adolescents who have not yet developed a psychiatric disorder. Utilizing a novel analytical strategy of component-wise gradient boosting to build predictive models of depression and anxiety we found that increased levels of TGF-alpha predicted higher levels of depression, while increased GRO predicted lower levels of depression (Table 2, Figure 1). TGF-alpha is a ligand for the epidermal growth factor receptor and is involved in neurogenesis in the CNS (Cooper and Isacson, 2004). Deficits in neurogenesis have been suggested to play a role in development of depression (Pascual-Brazo et al., 2014). As far as we know, TGF-alpha has not been previously reported to influence risk for depression, although other members of the epidermal growth factor family such as VEGF and TGF-beta have been implicated in depression and other psychiatric disorders (Galvez-Contreras et al., 2016; Sharma et al., 2016). TGF-alpha has been suggested as a biomarker for cocaine abuse (Maza-Quiroga et al., 2017). GRO (CXCL1), a member of the CXC family of chemokines, is a ligand for the CXCR2 receptor that plays a role in neuronal electrical activity, neurotransmitter release, and synaptic plasticity in the CNS (Semple et al., 2010). Interestingly, CXCR2 is thought to contribute to the trafficking of neuronal processes to form appropriate synapses during brain development (Luan et al., 2001). GRO brain levels are increased in an animal model of chronic stress (Girotti et al., 2011). This is of interest given the above mentioned hypothesis regarding the role of inflammatory molecules in response to stress in depression (Rantala et al., 2017).

We found that female sex predicted higher levels of depression and anxiety. This is in line with previous findings of a “female preponderance” in depression (Wang et al., 2016). In

regards to immunological factors, males and females are known to differ in their innate and adaptive immune responses, with adult females mounting stronger responses than males, which renders them more susceptible to inflammatory and autoimmune diseases (Klein and Flanagan, 2016). Whether this increased immune response also renders females more susceptible to depression and anxiety is an intriguing question that warrants further studies in light of our current findings. Importantly, sex hormones contribute to the differential immune response between sexes. In general, low estrogen concentrations promote production of inflammatory cytokines and cell-mediated immunity, while high estrogen concentrations reduce production of inflammatory cytokines and promote humoral immunity (Bouman et al., 2005; Straub, 2007). On the other hand, testosterone, found at high levels in post-pubertal men and women, generally suppresses immune cell activity (Roberts et al., 2001). Notably, menarchal status has been shown to be a strong predictor of depression and anxiety in girls (Patton et al., 1996). Given that levels of sex hormones are highly influenced by pubertal status, and given that our study was performed during adolescence, our findings of female sex predicting depression and anxiety may be due to differences in the onset of puberty which occurs between the ages of 11–14 in females, and between 13–16 in males (Parent et al., 2003). Unfortunately, the pubertal status of the participants was not measured as a part of this study, rendering it impossible to control for this important factor. In addition, we did not assess the menstrual phase of female subjects in this study. As levels of circulating cytokines are influenced by the menstrual cycle (Hatta et al., 2009; O'Brien et al., 2007), this factor may have influenced our results. An additional limitation of the present data lies in the decrease of participation following each subsequent time point. While we did not find reason for any systematic dropping out, the present sample may reflect a certain degree of convenience, and caution should be exercised in interpreting parameter estimates at higher age values. Additionally, as noted in the data analytic strategy, the portion of outliers removed from the analyses represented 12% of the available observations. While these values were removed based on a philosophical choice to investigate adolescents under normalized inflammatory processes, the removal of outliers represents loss of information. This study is limited in its ability to represent adolescents that are undergoing acute inflammation. The proportion of males to females was also higher in the set of removed outliers than in the sample represented by the analyses, potentially influencing the impact of sex on the presented models. Further longitudinal studies should be performed to validate our present results.

There are several other limitations to this study. First, the component-wise gradient boosting algorithm used here is capable of investigating nonlinear spline functions of the predictors and interactions between covariates; however, investigation of all possible nonlinear functions and interactions was beyond the computational capacity (and scope) of the present research. Further, the algorithm as implemented here was used to model the means of the outcome variables; future research should examine the utility of boosted quantile regression for these outcomes to provide predictive models at different percentiles of the outcome. It also bears note that the bootstrapped standard errors (and resulting confidence intervals) provided here were directed at resolving issues related to heteroscedasticity. The present findings are limited to the extent that the bootstrap fails to provide robust standard errors in this context.

In summary, the present study suggests that female sex is a strong predictor of depression and anxiety. Levels of TGF-alpha and GRO may be useful for prediction of manifestation of symptoms in subjects at risk for depression. These studies could lead to development of strategies for prevention or early treatment in susceptible populations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank participating families for their contributions to this work. This work was supported by the National Institute on Alcohol Abuse and Alcoholism grant R01AA016274 (DEW) and the National Institute of Mental Health grant R01MH087493 (CWB).

## References

Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. Organ Res Meth. 2013; 16:270–301.

Angold A, Costello EJ, Messer SC, Pickles A, Winder F, Silver D. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Methods Psychiatr Res. 1995; 5:237–249.

Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, Neer SM. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997; 36:545–553. [PubMed: 9100430]

Bouman A, Heineman MJ, Faas MM. Sex hormones and the immune response in humans. Hum Reprod Update. 2005; 11:411–423. [PubMed: 15817524]

Bühlmann P, Hothorn T. Boosting algorithms: Regularization, prediction and model fitting. Stat Sci. 2007; 22:477–505.

Cooper O, Isacson O. Intrastriatal transforming growth factor alpha delivery to a model of Parkinson's disease induces proliferation and migration of endogenous adult neural progenitor cells without differentiation into dopaminergic neurons. J Neurosci. 2004; 24:8924–8931. [PubMed: 15483111]

Dantzer R. Cytokine, sickness behavior, and depression. Neurol Clin. 2006; 24:441–460. [PubMed: 16877117]

Dantzer R, Kelley KW. Twenty years of research on cytokine-induced sickness behavior. Brain Behav Immun. 2007; 21:153–160. [PubMed: 17088043]

Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK, Lanctot KL. A meta-analysis of cytokines in major depression. Biol Psychiatry. 2010; 67:446–457. [PubMed: 20015486]

Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001; 29:1189–1232.

Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal. 2002; 38:367–378.

Galecki P, Talarowska M. Neurodevelopmental theory of depression. Prog. Neuropsychopharmacol Biol Psychiatry. 2018; 80:267–272. [PubMed: 28571776]

Galvez-Contreras AY, Campos-Ordonez T, Lopez-Virgen V, Gomez-Plascencia J, Ramos-Zuniga R, Gonzalez-Perez O. Growth factors as clinical biomarkers of prognosis and diagnosis in psychiatric disorders. Cytokine Growth Factor Rev. 2016; 32:85–96. [PubMed: 27618303]

Garcia-Bueno B, Caso JR, Leza JC. Stress as a neuroinflammatory condition in brain: damaging and protective mechanisms. Neurosci Biobehav Rev. 2008; 32:1136–1151. [PubMed: 18468686]

Girotti M, Donegan JJ, Morilak DA. Chronic intermittent cold stress sensitizes neuro-immune reactivity in the rat brain. Psychoneuroendocrinology. 2011; 36:1164–1174. [PubMed: 21411230]

Hatta K, Bilinski MJ, Haladyn JK, Roy JJ, Horrocks J, van den Heuvel MJ, Han VK, Croy BA. Cytokine array comparisons of plasma from cycling fertile women on cycle day 5 and ovulation. Am J Reprod Immunol. 2009; 62:158–164. [PubMed: 19694641]

Hofner B, Mayr A, Robinzonov A, Schmid M. Model-based boosting in R: A hands-on tutorial using the R package mboost. Comput Stat. 2014; 29:3–35.

Hothorn T, Buehlmann P, Kneib T, Schmid M, Hofner B. mboost: Model-Based Boosting R package version 2.6-0. 2016

Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. Psychosom Med. 2009; 71:171–186. [PubMed: 19188531]

James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning: With applications in R. Springer; New York, NY: 2013.

Kalin NH. Mechanisms underlying the early risk to develop anxiety and depression: A translational approach. Eur Neuropsychopharmacol. 2017; 27:543–553. [PubMed: 28502529]

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

Khandaker GM, Pearson RM, Zammit S, Lewis G, Jones PB. Association of serum interleukin 6 and C-reactive protein in childhood with depression and psychosis in young adult life: a population-based longitudinal study. JAMA Psychiatry. 2014; 71:1121–1128. [PubMed: 25133871]

Khandaker GM, Stochl J, Zammit S, Goodyer I, Lewis G, Jones PB. Childhood inflammatory markers and intelligence as predictors of subsequent persistent depressive symptoms: a longitudinal cohort study. Psychol Med. 2017:1–12.

Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol. 2016; 16:626–638. [PubMed: 27546235]

Kuhn, M., Johnson, K. Applied predictive modeling. Springer; New York, NY: 2013.

Luan J, Furuta Y, Du J, Richmond A. Developmental expression of two CXC chemokines, MIP-2 and KC, and their receptors. Cytokine. 2001; 14:253–263. [PubMed: 11444905]

Maza-Quiroga R, Garcia-Marchena N, Romero-Sanchiz P, Barrios V, Pedraz M, Serrano A, Nogueira-Arjona R, Ruiz JJ, Soria M, Campos R, Chowen JA, Argente J, Torrens M, Lopez- Gallardo M, Marco EM, Rodriguez de Fonseca F, Pavon FJ, Araos P. Evaluation of plasma cytokines in patients with cocaine use disorders in abstinence identifies transforming growth factor alpha (TGFalpha) as a potential biomarker of consumption and dual diagnosis. PeerJ. 2017; 5:e3926. [PubMed: 29038767]

Miklowitz DJ, Portnoff LC, Armstrong CC, Keenan-Miller D, Breen EC, Muscatell KA, Eisenberger NI, Irwin MR. Inflammatory cytokines and nuclear factor-kappa B activation in adolescents with bipolar and major depressive disorders. Psychiatry Res. 2016; 241:315–322. [PubMed: 27227701]

O'Brien SM, Fitzgerald P, Scully P, Landers A, Scott LV, Dinan TG. Impact of gender and menstrual cycle phase on plasma cytokine concentrations. Neuroimmunomodulation. 2007; 14:84–90. [PubMed: 17713355]

Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocr Rev. 2003; 24:668–693. [PubMed: 14570750]

Pascual-Brazo J, Baekelandt V, Encinas JM. Neurogenesis as a new target for the development of antidepressant drugs. Curr Pharm Des. 2014; 20:3763–3775. [PubMed: 24180394]

Patton GC, Hibbert ME, Carlin J, Shao Q, Rosier M, Caust J, Bowes G. Menarche and the onset of depression and anxiety in Victoria, Australia. J Epidemiol Community Health. 1996; 50:661–666. [PubMed: 9039386]

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team. R package version 3.1-128. 2016. nlme: Linear and Nonlinear Mixed Effects Models.

R Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2016. A language and environment for statistical computing.

Raison CL, Capuron L, Miller AH. Cytokines sing the blues: inflammation and the pathogenesis of depression. Trends Immunol. 2006; 27:24–31. [PubMed: 16316783]

Raison CL, Miller AH. Is depression an inflammatory disorder? Curr Psychiatry Rep. 2011; 13:467–475. [PubMed: 21927805]

Rantala MJ, Luoto S, Krams I, Karlsson H. Depression subtyping based on evolutionary psychiatry: Proximate mechanisms and ultimate functions. Brain Behav Immun. 2017

Roberts CW, Walker W, Alexander J. Sex-associated hormones and immunity to protozoan parasites. Clin Microbiol Rev. 2001; 14:476–488. [PubMed: 11432809]

Semple BD, Kossmann T, Morganti-Kossmann MC. Role of chemokines in CNS health and pathology: a focus on the CCL2/CCR2 and CXCL8/CXCR2 networks. J Cereb Blood Flow Metab. 2010; 30:459–473. [PubMed: 19904283]

Sharma AN, da Costa e Silva BF, Soares JC, Carvalho AF, Quevedo J. Role of trophic factors GDNF, IGF-1 and VEGF in major depressive disorder: A comprehensive review of human studies. J Affect Disord. 2016; 197:9–20. [PubMed: 26956384]

Straub RH. The complex role of estrogens in inflammation. Endocr Rev. 2007; 28:521–574. [PubMed: 17640948]

Suchting R, Hebert ET, Ma P, Kendzor DE, Businelle MS. Using Elastic Net Penalized Cox Proportional Hazards Regression to Identify Predictors of Imminent Smoking Lapse. Nicotine Tob Res. 2017

Venables, WN., Ripley, BD. Modern applied statistics with S. 4. Springer; New York, NY: 2002.

Wang K, Lu H, Cheung EF, Neumann DL, Shum DH, Chan RC. "Female Preponderance" of Depression in Non-clinical Populations: A Meta-Analytic Study. Front Psychol. 2016; 7:1398. [PubMed: 27695433]

Young JJ, Bruno D, Pomara N. A review of the relationship between proinflammatory cytokines and major depressive disorder. J Affect Disord. 2014; 169:15–20. [PubMed: 25128861]

Zhang Z. Variable selection with stepwise and best subset approaches. Ann Transl Med. 2016; 4:136. [PubMed: 27162786]

## HIGHLIGHTS

- TGF-alpha is a potential novel biomarker in the development of adolesdent depression.
- Machine learning provided optimized modeling of adolescent depression and anxiety.
- Female sex, growth-related oncogene (GRO), and transforming growth factor alpha (TGF-alpha) predicted adolescent depression.
- Female sex predicted adolescent anxiety.

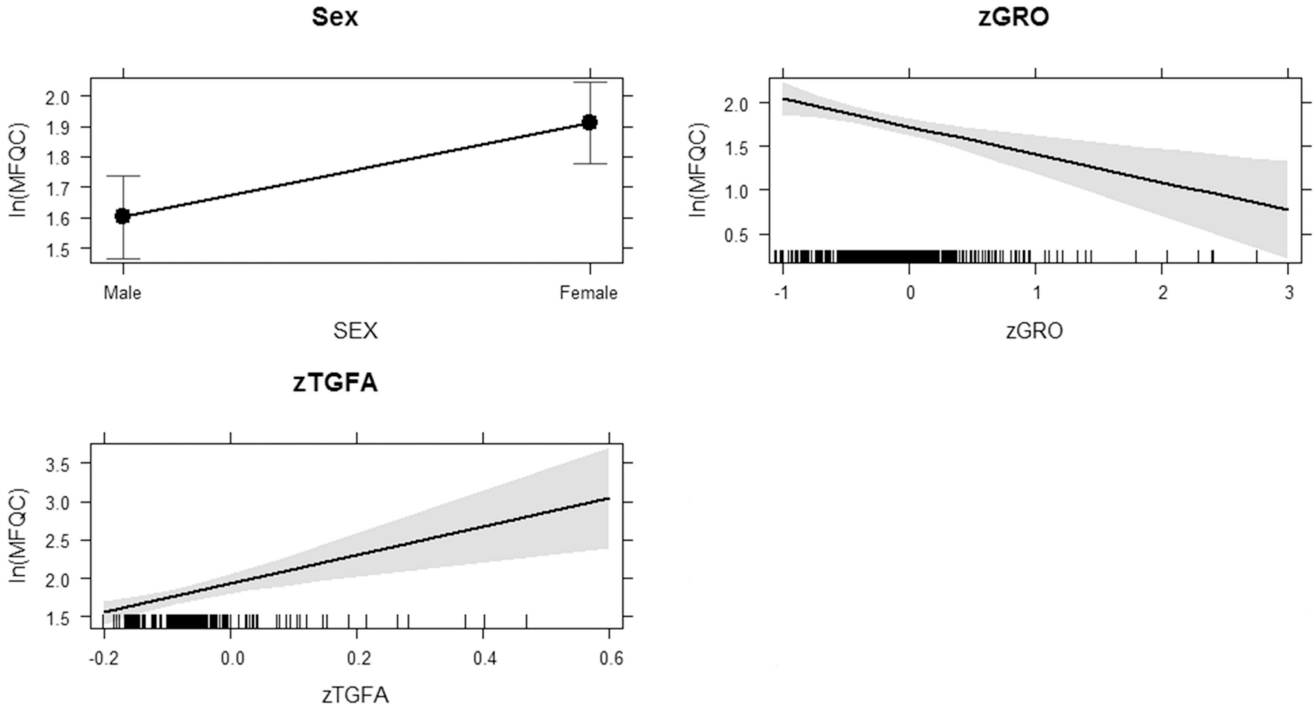


**Figure 1. Statistically Reliable Effects for Depression**

Plots show the relationships between sex, GRO, and TGFα with depression. Figure was generated using the plot() function of the effects library (Fox, 2003) in the R statistical computing environment. The y-axis for each plot shows the scale of the outcome variable (ln(MFQC)) and the x-axis describes the scale of the predictor variables: each category is described for the categorical variable (sex) and increasing quantities are described for the continuous variables (inflammatory markers).

**Table 1**

Demographic Characteristics by Time Point

| Time | N | Mean Age (SD) | Male | Female | White | Hispanic | Other Race |
|---|---|---|---|---|---|---|---|
| 1 | 254 | 13.37 (0.96) | 118 (46%) | 136 (54%) | 147 (58%) | 97 (38%) | 10 (4%) |
| 2 | 237 | 14.50 (0.95) | 111 (47%) | 126 (53%) | 133 56%) | 91 (38%) | 13 (5%) |
| 3 | 195 | 15.51 (0.97) | 97 (50%) | 98 (50%) | 109 (56%) | 76(39%) | 10 (5%) |

Baseline= time 1, year 1 follow-up= time 2, year 2 follow-up=time 3.

**Table 2**

Depression – Boosting, Reduction, & Model Comparison

| ***Boosted Model - Selected Predictors and Coefficients*** | |
|---|---|
| Predictor | Coefficient |
| SEX (Female) | 0.03 |
| zGRO | −0.14 |
| zIL127 | 0.02 |
| zSCD40 | −0.01 |
| zTGFA | 0.16 |
| zIL13:CTIME | −0.07 |
| zIL15:CTIME | −0.01 |

| ***Full Baseline Model - Fixed Effects*** | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | SE | DF | t-value | p-value |
| (Intercept) | 1.72 | 0.08 | 364 | 21.53 | < 0.001 |
| CTIME | −0.02 | 0.04 | 364 | −0.52 | 0.607 |
| SEX (Female) | 0.30 | 0.10 | 311 | 3.12 | 0.002 |
| zGRO | −0.30 | 0.10 | 364 | −3.14 | 0.002 |
| zIL127 | 0.16 | 0.10 | 364 | 1.66 | 0.099 |
| zSCD40 | −0.06 | 0.11 | 364 | −0.54 | 0.588 |
| zTGFA | 1.81 | 0.52 | 364 | 3.45 | 0.001 |
| zIL13 | 0.00 | 0.18 | 364 | −0.02 | 0.982 |
| zIL15 | −0.29 | 0.15 | 364 | −1.98 | 0.048 |
| CTIME:zIL13 | −0.14 | 0.13 | 364 | −1.07 | 0.286 |
| CTIME:zIL15 | −0.20 | 0.15 | 364 | −1.32 | 0.186 |

| ***Reduced Model - Fixed Effects*** | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Bootstrap SE | DF | Bootstrap 95% CI | |
| (Intercept) | 1.72 | 0.088 | 367 | 1.526 | 1.871 |
| CTIME | −0.01 | 0.054 | 367 | −0.108 | 0.105 |
| zIL127 | 0.15 | 0.115 | 367 | −0.100 | 0.349 |
| zTGFA | 1.87 | 0.634 | 367 | 0.595 | 3.082 |
| zGRO | −0.32 | 0.136 | 367 | −0.563 | −0.031 |

*J Affect Disord*. Author manuscript; available in PMC 2019 July 01.

***Boosted Model - Selected Predictors and Coefficients***

| | | | | | |
|---|---|---|---|---|---|
| SEX (Female) | 0.31 | 0.104 | 311 | 0.154 | 0.563 |
| zIL15 | −0.29 | 0.220 | 367 | −0.725 | 0.136 |
| CTIME:zIL15 | −0.25 | 0.187 | 367 | −0.690 | 0.042 |

***Model Comparison***

| Model | Psuedo-$R^2$ | AIC | BIC | -LogLik |
|---|---|---|---|---|
| Boosted | 0.7430 | | | |
| Full (NLME) | 0.7443 | 1856.30 | 1924.26 | −913.15 |
| Reduced | 0.7467 | 1851.70 | 1906.07 | −913.85 |

**Table 3**

Anxiety – Boosting, Reduction, & Model Comparison

| ***Boosted Model - Selected Predictors and Coefficients*** | | | | | |
|---|---|---|---|---|---|
| Predictor | Coefficient | | | | |
| SEX (Female) | 0.06 | | | | |
| zSCD40 | −0.05 | | | | |
| zIL13:CTIME | −0.02 | | | | |
| zIP10:CTIME | −0.01 | | | | |
| zTNFB:CTIME | −0.05 | | | | |
| ***Full Baseline Model - Fixed Effects*** | | | | | |
| Parameter | Estimate | SE | DF | t-value | p-value |
| (Intercept) | 2.20 | 0.06 | 365 | 34.56 | < 0.001 |
| CTIME | −0.03 | 0.04 | 365 | −0.95 | 0.345 |
| SEX (Female) | 0.36 | 0.09 | 311 | 4.11 | < 0.001 |
| zSCD40 | −0.16 | 0.08 | 365 | −1.95 | 0.052 |
| zIL13 | 0.12 | 0.15 | 365 | 0.78 | 0.433 |
| zIP10 | 0.02 | 0.05 | 365 | 0.44 | 0.664 |
| zTNFB | −0.07 | 0.12 | 365 | −0.52 | 0.601 |
| CTIME:zIL13 | −0.14 | 0.14 | 365 | −1.03 | 0.306 |
| CTIME:zIP10 | −0.08 | 0.05 | 365 | −1.64 | 0.102 |
| CTIME:zTNFB | −0.11 | 0.12 | 365 | −0.96 | 0.339 |
| ***Reduced Model - Fixed Effects*** | | | | | |
| Parameter | Estimate | Bootstrap SE | DF | Bootstrap 95% CI | |
| (Intercept) | 2.21 | 0.076 | 367 | 2.071 | 2.367 |
| CTIME | −0.03 | 0.050 | 367 | −0.108 | 0.089 |
| SEX (Female) | 0.36 | 0.092 | 311 | 0.163 | 0.523 |
| zSCD40 | −0.17 | 0.103 | 367 | −0.354 | 0.050 |
| zIL13 | 0.07 | 0.221 | 367 | −0.370 | 0.498 |
| zIP10 | 0.02 | 0.071 | 367 | −0.141 | 0.138 |
| CTIME:zIL13 | −0.23 | 0.214 | 367 | −0.590 | 0.250 |
| CTIME:zIP10 | −0.08 | 0.077 | 367 | −0.237 | 0.063 |

***Boosted Model - Selected Predictors and Coefficients***

***Model Comparison***

| Model | Psuedo-$R^2$ | AIC | BIC | -LogLik |
|---|---|---|---|---|
| Boosted | 0.7920 | | | |
| Full (NLME) | 0.7971 | 1587.49 | 1650.93 | −779.75 |
| Reduced | 0.7982 | 1584.86 | 1639.23 | −780.43 |