Multifactorial Prediction of Depression Diagnosis and Symptom Dimensions

Mary E. McNamara\textsuperscript{a}, Jason Shumake\textsuperscript{a}, Rochelle A. Stewart\textsuperscript{a}, Jocelyn Labrada\textsuperscript{a}, Alexandra Alario\textsuperscript{b}, John J.B. Allen\textsuperscript{b}, Rohan Palmer\textsuperscript{c}, David M. Schnyer\textsuperscript{a}, John McGeady\textsuperscript{d}, Christopher G. Beevers\textsuperscript{a}

\textsuperscript{a}Department of Psychology and Institute for Mental Health Research, University of Texas at Austin, Austin, Texas, USA

\textsuperscript{b}Department of Psychology, University of Arizona, Tucson, Arizona, USA

\textsuperscript{c}Department of Psychology, Emory University, Atlanta, Georgia, USA

\textsuperscript{d}Veterans Affairs, Providence RI and Department of Psychiatry and Human Behavior, Brown University School of Medicine, Providence, Rhode Island, USA

Abstract

While depression is a leading cause of disability, prior investigations of depression have been limited by studying correlates in isolation. A data-driven approach was applied to identify out-of-sample predictors of current depression from adults ($N = 217$) sampled on a continuum of no depression to clinical levels. The current study used elastic net regularized regression and predictors from sociodemographic, self-report, polygenic scores, resting electroencephalography, pupillometry, actigraphy, and cognitive tasks to classify individuals into currently depressed (MDE), psychiatric control (PC), and no current psychopathology (NP) groups, as well as predicting symptom severity and lifetime MDE. Cross-validated models explained 20.6\% of the out-of-fold deviance for the classification of MDEs versus PC, 33.2\% of the deviance for MDE.
versus NP, but −0.6% of the deviance between PC and NP. Additionally, predictors accounted for 25.7% of the out-of-fold variance in anhedonia severity, 65.7% of the variance in depression severity, and 12.9% of the deviance in lifetime depression (yes/no). Self-referent processing, anhedonia, and psychosocial functioning emerged as important differentiators of MDE and PC groups. Findings highlight the advantages of using psychiatric control groups to isolate factors specific to depression.

Keywords
statistical learning; psychiatric control; classification

1. Introduction
Globally, depression is estimated to affect 300 million individuals and is the leading cause of disability worldwide (World Health Organization & Others, 2017). Even more concerning, prevalence rates have remained static, despite increased accessibility of treatment over the last four decades (Ormel et al., 2019). Even the best available treatments are largely unsuccessful at producing lasting outcomes, as approximately 40–50% of patients relapse within 1–2 years of receiving treatment (Dobson et al., 2008; Paykel, 2007; Vittengl et al., 2007). This disparity highlights a clear need to better understand the critical processes associated with the maintenance of depression.

Several factors may have hampered progress in this area. Traditionally, depression has been studied using one or two units of analysis (e.g., self-report, neurocognitive assessments, neuroimaging), and rarely do studies examine multiple mechanisms simultaneously. For instance, few studies have included objective measures of biological risk factors in combination with behavioral or self-report data (Kaplan, Roberts, Camacho, & Coyne, 1987). Such work often requires interdisciplinary collaboration, which until relatively recently was somewhat uncommon in psychiatric research. Additionally, psychology has emphasized monocausal theories, undoubtedly in a hopeful attempt to mirror general medicine’s success at discovering specific etiologies (e.g. pathogens engendering infectious disease; Kendler, 2019). As a result, truly integrative cross-level research in psychiatric science has been rare (Kendler, 2014), even though it is necessary to move the field forward (Fried & Robinaugh, 2020).

Work in this area has also focused primarily on developing explanatory models at the expense of models that can predict outcomes in a new sample (Yarkoni & Westfall, 2017). It is important to highlight prediction in this context refers not to prediction over time (e.g. longitudinal prediction), but expected performance in out-of-sample data the model was not trained on (e.g. predictive R²). Models built for these distinct purposes often differ in important ways—while an explanatory model may be more elegant and offer more satisfying explanations, these models may perform poorly with out-of-sample data and fail to generalize, due to overfitting of the algorithm to statistical noise specific to the dataset. On the other hand, prediction-based models may generalize with better accuracy to new data but may lack clarity about how the mechanisms give rise to the outcomes. Applied to the
problem at hand, an emphasis on prediction would allow for multiple types of data from a variety of theories to predict the presence of depression, without getting bogged down by a lack of theory regarding how these components may entangle and interact with each other.

Increasingly popular machine learning techniques have made testing multifactorial models of depression (i.e. models incorporating multiple factors including biological, genetic, psychosocial, and cognitive factors) a possibility. In machine learning, the prediction of “unseen” data (data not used to train the model) is treated as the gold standard of success, an additional reason the use of machine learning algorithms is an attractive choice (Yarkoni & Westfall, 2017). Data-driven analyses of multifactorial datasets show great promise towards improving and testing models of depression and unveiling prospective treatment targets. Preliminary multifactorial work has identified correlates of depression outcomes, but to date, similar research has not distinguished the best variables that differentiate depression (Dinga et al., 2018).

As a first step to identify the most important predictors, it is necessary to tease apart specificity to depression (Fried & Kievit, 2016). To this end, psychiatric control groups should be used in place of (or at least in addition to) healthy controls. The lack of a psychiatric control group fails to tease apart which factors are specific to depression versus the presence of negative affect, impairment, and other nonspecific factors associated with having a mental health condition (Fritzsche et al., 2010).

Moreover, comparison groups with stricter inclusion and exclusion criteria than the case group have the potential to be sampled from an entirely different population. To use language from case-control studies, the only difference between a ‘case’ and a ‘control’ should be the development of the condition; both groups should have been “exposed” to the same conditions. A better alternative is the addition of a psychiatric control group in which the sample meets for the exact same inclusion and exclusion criteria as the mental illness in question, except for the focal feature (Schwartz & Susser, 2011).

To address these issues, we applied elastic net regularized regression to identify variables that predicted depression both categorically and continuously in a secondary analysis of a large, multifactorial dataset. Elastic net was used to predict three main outcomes: categorical grouping of MDE vs. psychiatric control vs. no current psychopathology, continuous measures of symptom severity, and lifetime depression status. Predictors in the model included a large range of demographic (e.g., sex, age) and self-report (e.g., rumination, psychosocial functioning) data, resting state EEG (i.e., frontal alpha asymmetry), actigraphy (i.e., daily physical activity, sleep consistency), polygenic risk scores (i.e., derived using a broad depression phenotype from the UK Biobank (Howard et al., 2018), and cognitive-affective task (e.g., self-referent negative processing, negative attention bias) data.

2. Methods and Materials

2.1 Participants

Data were collected from 218 unrelated adults; however, 1 participant was missing 25% or more of their data and was excluded from the analyses (further detail on handling of missing
data can be found in the supplementary materials). The present study included a sample of 217 adults recruited from the Austin, Texas community. Of these, 60 were included in the MDE group, 54 in the Psychiatric Control, and 101 in the No Current Psychopathology group (we were unable to obtain diagnostic data from two individuals, so they were not used in the analyses, but were used in the models predicting anhedonia and depression continuously). Since there was a genetic component to the primary aim of the study, the sample was limited to individuals of European ancestry. Consistent with dimensional approaches to psychopathology (Gibb et al., 2004), participants were recruited along a continuum of no symptomology to clinical levels, to approximate a normally distributed sample of depression. This required over-sampling the depressed end of that continuum, so our sample was not random in that regard.

Data were collected at two timepoints separated by approximately one week. Participants came into the lab for an initial four-hour appointment, during which written consent was obtained and timepoint one data was collected. Actigraphy data was collected over the 5–7 day interim period between study visits. Apart from the actigraphy data, all other data in the present analysis (e.g. self-report, cognitive task, EEG, etc.) were obtained from the first timepoint in order to utilize the largest sample size possible. Additional information on test-retest reliability of the metrics over the one-week period are documented in the Supplementary Materials. The Institutional Review Board at the University of Texas at Austin approved all study procedures.

Participants were eligible if they met the following inclusion criteria: 1) 18 to 35 years of age; 2) of European descent; 3) able to speak and read proficiently in English, and 4) either normal or corrected to normal vision. The exclusion criteria were: 1) current use of steroidal or psychotropic medications; 2) serious medical conditions; 3) heavy tobacco use defined as 20 cigarettes per day or greater than 20 pack years (Kamholz, 2004; World Health Organization, 2011); 4) score of two or higher on the drug subscale of the Psychiatric Diagnostic Screening Questionnaire (Zimmerman & Mattia, 2001); 5) a score of two or higher on the alcohol subscale of the Psychiatric Diagnostic Screening Questionnaire; 6) a score of one or higher on the psychosis subscale of Psychiatric Diagnostic Screening Questionnaire; or 7) being in imminent danger to others or self, or any recent suicidal behavior (suicidal ideation at level 4 on the Columbia-Suicide Severity Rating Scale in the past two months, or any suicidal behavior in the past two months).

Given that several of the predictors in the parent project were cognitive in nature, the sample was restricted to those between the ages of 18–35 years in order to reduce the impact of cognitive aging on those assessments. Further, anyone with heavy tobacco, drug, or alcohol usage were also excluded, given the effect substances can have on cognitive functioning (Hall et al., 2015).

Participants were of European-ancestry and mostly non-Hispanic, mostly female, in their mid 20s, and never-married. The majority of the sample had experienced a past episode of depression (60.9%). Full participant demographics are reported in Table 1.
Despite our best efforts to ensure the highest level of similarity between groups, the MDE group did experience higher levels of impairment than the psychiatric control (Table 1). However, including impairment as a predictor in the model allows us to control for any observed differences in impairment.

2.2 Measures

2.2.1 Outcomes—Participants were separated into one of three diagnostic groups based on their current diagnoses assessed by version 7.2 of the Mini International Neuropsychiatric Interview for DSM-5 (MINI; Sheehan et al., 1998). Participants were grouped into no current psychopathology (NP), current major depressive episode (MDE), or a psychiatric control (PC) group. Consistent with the criteria for psychiatric controls to have the same inclusion and exclusion criteria as the focal group, the PC group included anyone with a current diagnosis of mania, hypomania, generalized anxiety disorder, social anxiety disorder, agoraphobia, substance or alcohol use disorders, psychosis, post-traumatic stress disorder, obsessive-compulsive disorder, but not a current major depressive episode. Individuals in the MDE group could also have comorbid psychiatric conditions. In the MDE group, 50.0% of participants also met for a comorbid anxiety disorder, and 46.3% of participants in the PC group met for an anxiety disorder. Full descriptions of comorbidities can be found in Table 2. Additionally, both the PC and NP groups included participants who had past episodes of depression. We controlled for this by entering a rank-ordered variable denoting the total number of episodes into the models as a predictor. Further detail about the diagnostic groupings can be found in the accompanying Supplementary Materials, ‘Outcome Measures.’

In our first set of analyses, we ran classification models for MDE vs PC, MDE vs NP, and PC vs NP. Next, we looked at two continuous measures of symptomatology, the Mood and Anxiety Symptom Questionnaire - Short Form (Wardenaar et al., 2010) general distress subscale (MASQ GD), a measure of depression severity, and the Snaith-Hamilton Pleasure Scale (SHAPS; Snaith et al., 1995), a measure of anhedonia. Finally, we ran a model to examine which, if any, factors differentiated those who experienced any lifetime episodes of depression, current or past, from those with no history of depression (dichotomized as yes/no). For more information on the psychometrics of each outcome, see the accompanying Supplementary Materials, ‘Psychometrics and Description of Predictor Variables.’

2.2.2 Predictors—Each model included 137–139 candidate predictors. A variable denoting the number of lifetime episodes of depression was included in 5 of the 6 models in order to control for past history of depression. Predictors were excluded for certain models if they conveyed information that was highly similar to the outcome (for example, the number of lifetime episodes was removed as a predictor from the lifetime history model only, as any number of episodes would by definition indicate whether or not someone had experienced a major depressive episode). The number of observations and predictors included in each model is documented in the Supplementary Materials, Table 2. One of the appealing features of the elastic net is that it can accommodate the inclusion of a large number of predictors without needing to worry about the ratio of predictors to observations. This is because the elastic net penalty, not the total number of predictors, determines the model degrees of
freedom (Hastie et al., 2015). We also removed five variables from the original dataset prior to beginning analyses that had near-zero variance and were unlikely to be useful predictors. A comprehensive list and descriptions of all predictors can be found in the Supplementary Materials, in Table 1.

Sociodemographic variables included age, sex, race, ethnicity, education level, years of school, marital status, income level, insurance coverage, and a rank-ordered count of lifetime depressive episodes. For the self-report questionnaires, total score and subscales (where applicable) were used. This included scores from the following measures: Ruminative Response Scale, Brooding Subscale (RRS; Treynor et al., 2003), Perseverative Thinking Questionnaire (PTQ; Ehring et al., 2011), Satisfaction With Life Scale (SWLS; Diener et al., 1985), Emotion Regulation Questionnaire (ERQ; Gross & John, 2003), Behavioral Inhibition Scale/Behavioral Activation Scale (BIS/BAS; Carver & White, 1994), International Physical Activity Questionnaire (IPAQ; Lee et al., 2011), Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989), and the Sheehan Disability Scale (SDS; Leon et al., 1997).

From the biobehavioral assessments, we used variables from digit span (largest number of digits recalled, forward and backward, and response time; Richardson, 2007), an affective pupil dilation task (index of reactivity to emotional stimuli; Steidtmann et al., 2010), resting state frontal alpha asymmetry and midfrontal mean alpha from EEG (Tomarken et al., 1990; Coan & Allen, 2004; Lopez-Duran et al., 2012), the dot probe for affective stimuli (attention bias for positive and dysphoric stimuli; Beevers et al., 2019) where reaction time, gaze, and trial-level bias variables were used (Zvielli et al., 2015), and an affective visual short-term working memory task designed to measure the ability to clear working memory of irrelevant affective material (Gonzalez & Schnyer, 2018). Self-referent negative processing was also assessed with the Self-Referential Encoding Task (SRET; Derry & Kuiper, 1981).

Finally, we also included actigraphy data (rest and activity cycles) and a polygenic risk score. Polygenic risk scores are the sum of the weighted liabilities conferred by multiple single-nucleotide polymorphisms associated with one outcome (Euesden et al., 2015). The parent project of the study calculated multiple polygenic risk scores for different outcomes, each of which was then Bonferroni-corrected. Given our outcomes of interest included classifying the presence and absence of depression categorically and continuously, we used a polygenic risk score calculated for current MDE and current BDI (Beck Depression Inventory-II). More detail regarding the assessments used in the current study can be found in Supplementary Materials, ‘Methods: Task Descriptions and Data Collection.’

2.3 Machine Learning Analyses

2.3.1 Learning Algorithms and Tuning Parameters—For classification, we selected the elastic net machine learning algorithm, which utilizes both lasso and ridge regression penalties to prevent overfitting of the model. Elastic net regularization $\alpha$ parameter finds a compromise between the lasso penalty (sum of absolute values of all coefficients) and ridge penalty (sum of squared values of all coefficients) by weighting the proportion of ridge and lasso penalties. In addition to $\alpha$, there is also an additional parameter $\lambda$, the magnitude of the shrinkage penalty to be tuned by cross-validation. The \texttt{hessel} package...
developed in-house by co-author (JS) tries over 100 possible values of \( \lambda \) (autogenerated by the package), as well as three values of \( \alpha \): 0.01 (weighting the ridge penalty more heavily and therefore including more variables), 0.99 (weighting the lasso penalty more heavily and including fewer variables), and 0.50 (equal weighting of both penalties). The values of both of these parameters were chosen to be as large as possible while remaining within one standard error of the values that resulted in the minimum cross-validation error. For classification, we compared two groups per model; MDE vs PC, then MDE vs. NP, and NP vs. PC.

2.3.2 Prediction Metrics and Cross-Validation—Cross-validation was used to estimate the predictive \( R^2 \), or the proportion of variance the model is expected to explain in unseen data (i.e. cases the model was not trained on). 10-fold cross-validation was repeated 10 times using different partitions of the data. In each repetition, 9 of the 10 “folds” or partitions of data are used to train the model and the 10th holdout data is used to test the prediction model. This is repeated 9 additional times where a model is trained on 90% of the data and tested on the remaining 10% of cases. Predictive \( R^2 \) is calculated based on the average residual error of holdout prediction across the 10 repetitions. Nested cross-validation was also conducted within each partition of training data in order to determine the optimal values of the tuning parameters. The standard error statistic reported for each model is the standard error of the prediction metric between the cross-validation folds, or the average error between the train and test partitions within each repetition. Additionally, the minimum and maximum estimates reported reflect the observed range across the 10 repetitions of cross-validation. For the classification models, the AUC statistic is reported in the supplementary materials in Table 3.

2.3.3 Calculation of Variable Importance Scores—Variable importance was operationalized as the absolute magnitude of the standardized regression coefficient for each predictor (Wei et al., 2015). While variable importance is not a well-defined concept (Grömping, 2009), using standardized regression coefficients is a valid and widely-accepted choice (for more rationale on the choice of standardized regression coefficients, see Supplementary Materials ‘Variable Importance Estimates’). The variable importance estimates derived for each model are averaged across the 10 repetitions of cross-validation. All analyses were implemented in R version 4.0.0 unless otherwise noted. Our code made extensive use of the tidyverse (Wickham, n.d.), particularly the dplyr and tidyr packages for data manipulation. The glmnet package was also used to run the elastic net models (Friedman et al., 2010). We also used the beset package (https://github.com/jashu) to run the models with cross-validation and produce variable importance plots. If the number of predictors retained by the model exceeded ten, we plotted only the top ten predictors.

2.4 Missing Data

Some participants in our study were unable to complete all the assessments. Since the algorithms require there is no missing data, we imputed missing data for the predictor variables using the missForest package (Stekhoven & Bühlmann, 2012). missForest uses all other variables for each observation to predict the value of the missing variable. We used a
cutoff of 25% for both variables and observations, so if any variable was missing for 25% or more of participants, it was dropped, and if any participant was missing 25% or more of their data, they were dropped. A detailed description of the treatment of missing data can be found in the supplementary materials section ‘Missing Data.’ Outcome variables were not included in the imputation so as to prevent any contamination between outcome and predictor values, which would optimistically bias prediction. Variables with missing data and percentage of cases missing can be found in the Supplementary Materials in Table 4. All of the code (and the associated output) used to generate the results reported in this article can be found in a PDF report titled, “Prediction-of-Depression-Main-and-Supplemental-Analyses.pdf.” All of the datasets generated and/or analyzed during the current study, as well as the code used to analyze them, are available in the Texas Data Repository, https://doi.org/10.18738/T8/DEFPNZ.

3. Results

3.1 Diagnostic Grouping

Using 10-fold cross-validation repeated 10 times with all predictors, out-of-sample predicted $R^2 = .206$ when classifying MDE versus PC groups (SE= 0.058, Min= 0.199, Max= 0.216). The variable importance plot is shown in Figure 1a. Predictors that increased the likelihood of MDE membership were: greater anhedonia (SHAPS total score) and more work and school-related disability (SDS) and global disability (SDS global score). The number of positive words endorsed (SRET) decreased the likelihood the participants belonged to the MDE group.

For classifying MDE and NP, out-of-sample predicted $R^2 = .332$ (SE= 0.057, Min= 0.287, Max= 0.367). The variable importance plot is shown in Figure 1b. Among the most important predictors positively associated with MDE status were: poor self-reported sleep quality, greater perceived dysfunction in daily activities as a result of sleep disturbance, and more global functional impairment related to sleep (PSQI composite 1, composite 7, and global scores); greater work and school-related disability and global disability (SDS work/school and global); higher levels of anhedonia (SHAPS); the number of negative words endorsed (SRET), and the number of negative, self-referential words recalled (SRET), and greater brooding (RRS brooding subscale). Additionally, the inter-trial variability of the starting point of the SRET and the rank-order-transformed number of episodes variable were associated with increased likelihood of MDE membership (e.g. higher values on the variables indicated increase likelihood of belonging to the MDE group).

For PC and NP groups, out-of-sample predictive $R^2$ was $-0.006$ (SE = 0.015, Min = $-0.022$, Max = 0.007), indicating the model added uncertainty to the prediction rather than reducing it.

3.2 Anhedonia

Out-of-sample predictive $R^2 = .257$ (SE= 0.037, Min = 0.231, Max= 0.272) for total anhedonia score in the full sample. As seen in Figure 2a, the variable importance score for the number of positive words endorsed as self-descriptive on the SRET was one of the most
important predictors of anhedonia (fewer positive words endorsed as self-descriptive was associated with higher levels of anhedonia), as was the rank ordered number of depressive episodes (more episodes of depression predicted higher levels of anhedonia). Other variables with inverse relationships to anhedonia were greater satisfaction with life (SWLS total); more motivation to pursue rewards (BIS/BAS reward responsiveness subscale), and total score for the behavioral approach system (BAS total). Additional important predictors of anhedonia were: greater perceived dysfunction in daytime activities as a result of sleep disturbance, higher global functional impairment related to sleep, and more sleep disturbances (PSQI composite score 7, global, and disturbances scores); a greater tendency to engage in brooding (RRS brooding subscale); greater global impairment (SDS global), and a greater number of negative, self-referential words recalled (SRET).

3.3 MASQ general distress

For depression severity, out-of-sample predictive $R^2 = .652$ (SE= 0.035, Min= 0.633, Max= 0.660) for the MASQ general distress subscale in the full sample. The variable importance plot is shown in Figure 2b. The most important predictors of greater depression severity were anhedonia (SHAPS); brooding (RRS brooding subscale); number of negative words endorsed (SRET); global disability and impairment (SDS); unproductiveness of negative repetitive thought and difficulty disengaging from negative repetitive thought subscales, and total score on the perseverative thinking questionnaire (PTQ); and. Higher levels of satisfaction with life (SWLS total); greater tendency to use reappraisal strategies (ERQ reappraisal score), and greater number of positive words endorsed (SRET) were associated with lower depression severity.

3.4 Lifetime history of depression

Out-of-sample predictive $R^2 = .129$ (SE= 0.033, Min= 0.109, Max= 0.150) for classification of lifetime episode of depression (yes/no). The strongest predictors of the likelihood of having had a depressive episode was anhedonia (SHAPS). Other important predictors included a greater number of negative words endorsed (SRET), and greater global functional impairment related to sleep (PSQI global score); endorsement of suicidal ideation; greater difficulty disengaging from negative repetitive thinking and total rumination score (PTQ; Figure 3). Greater satisfaction with life (SWLS) was associated with an increased likelihood of never having had an episode of depression.

Sample Size Justification—Learning curve analyses were performed to confirm the sample sizes were adequate for the elastic net models and can be found in the Supplementary Materials.

4. Discussion

The present study used a large multi-dimensional dataset to determine predictors of depression status, both categorically and dimensionally, as well as predictors of lifetime risk for an MDE. Additionally, this study is among the first to include a psychiatric control while trying to identify the best, specific correlates of depression using a data-driven approach. Potential predictor variables included demographics, biobehavioral assessments (e.g. pupil
dilation, eye-tracking), cognitive tasks (e.g. SRET), resting state EEG, a polygenic risk score for the broad depression phenotype derived from the UK Biobank, actigraphy, and self-report questionnaire data. When classifying the groups, the cross-validated deviance explained was 20.6% for MDE vs PC and 33.2% for MDE vs NP, and −0.006% for NP vs PC. When examining depression continuously, cross-validated models predicted 25.7% of the variance in anhedonia and 65.2% of general distress. Finally, the expected predictive deviance for lifetime history of depression was 12.9%. The cross-validated predicted $R^2$ indicates the variance (or deviance, for classification models) the model explained in “unseen” data that the model was not trained on, giving a good estimate of out-of-sample generalizability.

Several candidate predictors were identified as important across the categorical and continuous prediction models. Measurements of negative self-referent processing via the SRET, rumination, self-reported sleep quality, functional impairment, and satisfaction with life emerged as important predictors across analyses. Self-referential schemas have long been studied as important correlates of depression (LeMoult et al., 2017) and interventions targeting self-referential schemas are associated with a decrease in depression symptoms (Dainer-Best, Shumake, et al., 2018). Emerging evidence suggests targeting rumination may enhance cognitive-based therapies (Watkins, 2015). Interestingly, self-reported measures of sleep quality and disturbance were considered to be important predictors. The actigraphy variables measuring daily activity and circadian rhythms made marginal to no contribution to the models.

The use of a psychiatric control in comparison to the MDE group revealed important insights as well. Two of the best differentiators of people with depressive episodes from individuals with other psychiatric diagnoses were the number of positive words endorsed on the SRET and total anhedonia score. This is in line with mounting evidence of the importance of the role of low positive emotionality in depression (Khazanov & Ruscio, 2016; Vazquez, 2017). The importance of the number of positive words endorsed as self-descriptive in differentiating depressed individuals from psychiatric controls is also noteworthy. Previous research has highlighted the number of negative words endorsed as a strong, reliable predictor of depression (Dainer-Best, Lee, et al., 2018), which is consistent with our prediction models of MDE vs NP, general distress, and lifetime depression, where the number of negative words endorsed was a strong predictor. However, the lack of positive words endorsed highlights a bias that may be specific to MDE, perhaps being driven by low positive affect.

In fact, one theme that emerged across analyses was the importance of low positive emotionality and biased information processing away from positive stimuli in predicting depression. Typically, individuals show a bias towards positive information and away from negative information, but in depression, this pattern is reversed such that depressed individuals take in more negative information and filter out positive information (Peckham et al., 2010); (Duque & Vázquez, 2015); (Disner et al., 2011). The number of positive words endorsed on the SRET was one of the most important variables across the models predicting MDE vs PC, anhedonia, and general distress.
It is worthwhile to consider which predictors arose as valuable markers of depression, and which did not. Kendler set forth five criteria by which to gauge whether a variable meets the criteria of a marker for a psychiatric disorder: 1) strength of association; 2) specificity of association; 3) non-contingent on the presence of other factors; 4) causal proximity; and 5) an appropriately plausible level of explanation (Kendler, 2005). Particularly when considering the MDE vs PC model because of the specificity it afforded, number of positive words endorsed on the SRET, impairment and disability, and anhedonia all arose as strong predictors of MDE. These variables repeatedly were selected as top predictors in the other models as well, along with number of negative words endorsed (SRET), rumination, perceived sleep impairment, and satisfaction with life. Interactions among variables were not examined, consistent with the idea that these predictors are not contingent on the presence of other factors. Indeed, given the sample size, it is very unlikely that reliable interactions among predictors would have been identified in cross-validated models. This is partly why we selected elastic net regression rather than other more complex machine learning models that search for interactions among predictor variables.

Notably, all of the top markers were psychosocial in nature. Consistent with Dinga et al’s (2018) findings, the biological variables (e.g., frontal alpha asymmetry, polygenic risk score for the broad depression phenotype in the UK biobank, pupil dilation) added very little classification value to the model. As a direct test, we computed additional models without the biological variables, documented in the Supplementary Materials, Table 5, and found that the models with and without the biological variables were not significantly different in predicted $R^2$. Further, none of them correlated very strongly with the outcome measures when bivariate correlations were examined.

There are several limitations to the present study. First, all our participants were of European ancestry because of the genetic component of the parent study. Future research would need to evaluate whether the predictive models generalize to populations of non-European ancestry. Second, this sample was restricted to ages 18–35 years old to reduce any effects of cognitive aging on the heterogeneity of the sample, so we cannot speculate whether these findings extend to a wider range of ages. In our NP group, participants did not meet for any current psychopathology but were allowed to have lifetime diagnoses of various mental illnesses. Additionally, individuals in both the NP and PC groups were included even if they had experienced past episodes of depression, and we controlled for lifetime history of depression. However, we may have suppressed the predictive utility of dispositional factors because lifetime history of depression was present across groups. Considering the younger age of our sample and the range of median age of onset for depression can span from 25 to 45 (Kessler et al., 2007), it is likely that individuals in the other groups may have predisposing factors to depression but have not yet experienced their first episode. An additional limitation might be the criteria of the main study that participants could not be taking psychotropic or steroidal medications, which may limit generalizability. Finally, even in the best model (general distress), approximately 34% of the variance remained.

---

unexplained, meaning there are very likely other unmeasured variables that could also further contribute to prediction.

While these classification models reveal important correlates of depression, future research should examine the classification value of these predictions. Longitudinal studies should look at outcomes including the future onset of depression and relapse. Finally, while we took a supervised approach to identify important predictors of depression, future research should apply unsupervised approaches to multidimensional datasets to explore possible subtypes of depression. With these caveats in mind, the current study suggests that multiple mechanisms measured across units of analyses are associated with depression, with psychosocial functioning, negative self-referent processing, and perceived sleep quality among the most important predictors. Further, dysfunction with regard to positive affect (e.g., anhedonia, low endorsement of positive adjectives as self-referent) may be specifically impaired in depression and warrants additional investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment:

Funding for this study was provided by National Institute of Health (awards R56MH108650, R21MH110758, R33MH109600, R01DA042742). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References


Fried EI, & Kievit RA (2016). The volumes of subcortical regions in depressed and healthy individuals are strikingly similar: a reinterpretation of the results by Schmaal et al. [Review of The volumes of subcortical regions in depressed and healthy individuals are strikingly similar: a reinterpretation of the results by Schmaal et al]. Molecular Psychiatry, 21(6), 724–725. [PubMed: 26666203]


Highlights:

- Psychiatric control groups can help isolate factors specific to the diagnosis of interest
- Low positive emotionality is a strong differentiator of depression
- Additionally, perceived sleep quality and impairment are also important predictors
Figure 1:
Variable Importance Scores for MDE vs PC (1a) and MDE vs NP (1b)
Note: No variable importance plots are generated for the NP vs PC model, as the model added uncertainty to prediction instead of reducing it.
Figure 2:
Variable Importance Scores for the Snaith-Hamilton Pleasure Scale (2a) and the Mood and Anxiety Sensitivity Questionnaire, General Distress Subscale (2b)
Figure 3:
Variable Importance Plot for Lifetime History of Depression (3)
Table 1:
Demographics and Clinical Characteristics of the Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDE (n= 60)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>23.6 (4.4)</td>
</tr>
<tr>
<td>Female gender (%)</td>
<td>41 (71.7%)</td>
</tr>
<tr>
<td>Hispanic ethnicity (%)</td>
<td>17 (28.3%)</td>
</tr>
<tr>
<td>Caucasian (%)</td>
<td>58 (96.7%)</td>
</tr>
<tr>
<td>Single (%)</td>
<td>51 (85.0%)</td>
</tr>
<tr>
<td>Years in school (SD)</td>
<td>14.5 (2.4)</td>
</tr>
<tr>
<td>Private Health Insurance (%)</td>
<td>29 (48.3%)</td>
</tr>
<tr>
<td>Household Income (%)</td>
<td></td>
</tr>
<tr>
<td>$0 – $24,999</td>
<td>32 (53.3%)</td>
</tr>
<tr>
<td>$25,000 – $74,999</td>
<td>12 (20.0%)</td>
</tr>
<tr>
<td>$75,000 – $100,000 +</td>
<td>16 (26.7%)</td>
</tr>
<tr>
<td>Current MDD</td>
<td>60 (100%)</td>
</tr>
<tr>
<td>Past MDD</td>
<td>51 (85.0%)</td>
</tr>
<tr>
<td>Number of Depressive Episodes (median, SD) **</td>
<td>4.0 (10.0)</td>
</tr>
<tr>
<td>Any Current Anxiety Disorder</td>
<td>30 (50.0%)</td>
</tr>
<tr>
<td>Any Current Psychiatric Disorder</td>
<td>60 (100%)</td>
</tr>
<tr>
<td>BDI-II (SD)</td>
<td>29.7 (10.3)</td>
</tr>
</tbody>
</table>

* Two individuals did not have diagnostic data and were not used in the categorical analyses, but were included for the other continuous outcomes. Therefore, total percentages for variables using diagnostic data are out of a sample of n = 215

** This variable was rank-transformed before being included in the main analyses; these numbers reflect the raw data

† Anxiety disorders assessed included panic disorder, agoraphobia, social anxiety disorder, and generalized anxiety disorder
Table 2: Diagnostic Comorbidity by Group

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Group</th>
<th>MDE (n = 60)</th>
<th>PC (n = 54)</th>
<th>NP (n = 101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Depressive Episode (current)</td>
<td>60 (100.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Major Depressive Episode (past)</td>
<td>51 (85.0%)</td>
<td>35 (64.8%)</td>
<td>45 (44.6%)</td>
<td></td>
</tr>
<tr>
<td>Mania (current)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Hypomania (current)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Hypomanic symptoms (current)</td>
<td>0 (0.0%)</td>
<td>1 (1.9%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Mania (past)</td>
<td>4 (7.0%)</td>
<td>4 (7.4%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Hypomania (past)</td>
<td>1 (1.7%)</td>
<td>0 (0.0%)</td>
<td>1 (1.0%)</td>
<td></td>
</tr>
<tr>
<td>Hypomanic symptoms (past)</td>
<td>0 (0.0%)</td>
<td>1 (1.9%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Panic disorder (current)</td>
<td>10 (17.5%)</td>
<td>8 (17.0%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Panic Disorder (past)</td>
<td>23 (38.3%)</td>
<td>14 (25.9%)</td>
<td>9 (9.4%)</td>
<td></td>
</tr>
<tr>
<td>Agoraphobia (current)</td>
<td>11 (19.3%)</td>
<td>6 (12.8%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Generalized Anxiety Disorder (current)</td>
<td>18 (23.0%)</td>
<td>9 (16.7%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Social Anxiety Disorder</td>
<td>17 (29.8%)</td>
<td>11 (20.4%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Obsessive Compulsive Disorder (current)</td>
<td>9 (15.8%)</td>
<td>7 (13.9%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Post-Traumatic Stress Disorder</td>
<td>9 (15.8%)</td>
<td>3 (5.6%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Alcohol Use Disorder (current)</td>
<td>13 (22.8%)</td>
<td>21 (38.9%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Substance Use Disorder</td>
<td>13 (21.7%)</td>
<td>22 (40.7%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Psychosis (current)</td>
<td>1 (1.7%)</td>
<td>1 (2.1%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Psychosis (past)</td>
<td>1 (1.7%)</td>
<td>1 (2.1%)</td>
<td>1 (1.0%)</td>
<td></td>
</tr>
<tr>
<td>Anorexia Nervosa (current)</td>
<td>0 (0.0%)</td>
<td>1 (2.1%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Bulimia Nervosa (current)</td>
<td>2 (3.5%)</td>
<td>2 (4.3%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
<tr>
<td>Binge-Eating Disorder (current)</td>
<td>3 (5.2%)</td>
<td>2 (4.3%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
</tbody>
</table>