Published in final edited form as:

Drug Alcohol Depend. 2014 October 1; 143: 29–35. doi:10.1016/j.drugalcdep.2014.04.030.

Multimodal predictive modeling of individual treatment outcome in cocaine dependence with combined neuroimaging and behavioral predictors

Sean X. Luo¹, Diana Martinez¹, Kenneth M. Carpenter¹, Mark Slifstein¹, and Edward V Nunes¹

¹Division of Substance Abuse, Department of Psychiatry, New York State Psychiatric Institute, Columbia University. 1051 Riverside Drive #32, New York, NY 10032

Abstract

Background—Developing personalized treatments for cocaine dependence remains a significant clinical challenge. Positron emission tomography (PET) has shown that the [¹¹C]raclopride signal in the ventral striatum is associated with treatment success in a positively reinforced contingency management program. The present study investigates whether this signal can be used to predict treatment outcome at an individual level.

Methods—Predictive models were developed using PET signals from 5 regions of the striatum and follow-up data in 24 patients, and evaluated using cross-validation.

Results—The ventral striatal PET signal alone can predict individual treatment response with a substantial degree of accuracy (cross-validated correct rate = 82%). Incorporating information from other regions-of-interest (ROIs) in the striatum does not improve predictive performance, except for a small improvement with adding the posterior caudate. The addition of baseline demographic variables, including baseline severity measures, does not improve predictive performance. On the other hand, early treatment response and motivation, reflected by cumulative clinic attendance, performs as well as the PET signal (83%) by week 3 in the 24-week study. The combined model with both PET signals and cumulative clinic attendance demonstrates a significant improvement of performance, peaking at 96% during week 3 of the trial.

Conclusions—These results suggest that a multimodal model can predict treatment success in cocaine dependence at an individual level, and pose hypotheses for the underlying neural circuitry mechanisms responsible for individual variations in treatment outcome.

Address correspondence and reprint requests to Dr. Luo, New York State Psychiatric Institute, Box 83, 1051 Riverside Dr., New York, NY 10032; xsl2101@columbia.edu.

Conflict of Interest: Dr. Nunes receives medication and/or placebo for research studies from Alkermes/Cephalon, Inc., Reckitt-Benckiser, and Duramed Pharmaceuticals, and he has received or will receive from Brainsway devices under investigation and reimbursement for travel expenses for investigators' meeting in connection with a Brainsway research study; HealthSim, LLC supplied web-based behavioral intervention for a research study in data analysis. He was paid an honorarium and received reimbursement for travel expenses for attendance at a Lilly Advisory Board. All other authors have no Conflict of Interest to declare.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

dopamine; cocaine; predictive modeling; personalized medicine; PET; ventral striatum

1. Introduction

Recovery from cocaine dependence is a complex, multifactorial phenomenon. Previous studies have identified a number of clinical predictors of treatment efficacy, such as baseline severity of use and duration of use (Poling et al., 2007), though prediction of treatment response at the individual level remains difficult. Striatal dopamine neurons may encode many cognitive signals, such as reward response (Volkow et al., 2009), reinforcement (Tobler et al., 2005), and prediction error (Schultz, 2007), and computational models suggest that these signals are mechanistically related to clinical phenotypes (Montague et al., 2012). Biomarkers based on striatal dopamine transmission may therefore hold promise in predicting individual treatment response.

Positron Emission Tomography (PET) imaging with the dopamine D2/D3 receptor antagonist radioligand [^{11}C]raclopride provides two quantitative indices that could serve as biomarkers (Breier et al., 1997; Laruelle, 2000). The baseline binding potential (BP_{ND}) is proportional to the product of D2/D3 receptor density and the affinity of the tracer for the receptors, and measures dopamine receptor availability (Volkow et al., 1994). Because endogenous dopamine and [^{11}C]raclopride compete for binding to D2/D3 receptors, BP_{ND} reacts to changes in synaptic dopamine levels: BP_{ND} decreases after administration of psychostimulants, such as methylphenidate or amphetamine (Laruelle, 2000; Volkow et al., 1997), which increase endogenous synaptic dopamine. The percent decrease in BP_{ND} following a stimulant challenge (BP_{ND}) measures dopamine release capacity. Previous studies have shown that both baseline BP_{ND} and stimulant-induced BP_{ND} are decreased in cocaine-dependent subjects compared with healthy controls (Martinez et al., 2003; Volkow et al., 1997). Additionally, cocaine abusers with lower BP_{ND} in the ventral striatum are more likely to choose cocaine over money in a non-treatment laboratory model of cocaine-seeking behavior (self-administration; Martinez et al., 2007). Lower D2 receptors expression in the ventral striatum is associated with greater cue-induced craving for alcohol (Heinz et al., 2005, 2004). In the treatment setting, cocaine dependent subjects who responded to Community Reinforcement Approach (CRA) with a voucher reward (Higgins and Budney, 1993; Higgins et al., 2003; Sigmon and Higgins, 2006) have higher average measures of both BP_{ND} and BP_{ND} in the striatum compared with non-responders (Martinez et al., 2011). Similar signals of lower dopamine neurotransmission in the striatum are shown to be associated with relapse in methamphetamine abusers (Wang et al., 2012). However, even though the average intergroup differences are significant in these studies, the power of a given individual subject's PET scan to predict his or her treatment response has not been assessed.

While PET imaging can elucidate brain mechanisms of treatment response variations, it is expensive and burdensome to conduct. A simpler clinical variable may serve as a proxy or be used in conjunction. A parallel line of inquiry finds that early treatment success, as a behavioral metric, is also a good predictor of treatment outcome at the group level. For

instance, cessation of drug use during lead-in periods or abstinence at the time of randomization have been shown to be strong predictors of abstinence outcome in clinical trials (Bisaga et al., 2010, 2006, 2005; Brensilver et al., 2012; Kampman et al., 1996, 2001). Even a delay to the first clinic visit may be predictive of treatment attrition (Chawdhary et al., 2007). Early response to treatment (early abstinence or treatment attendance) might indicate motivation or sensitivity to reinforcement that is reflected in the PET signal of dopamine transmission. The potential for such a behavioral signal to predict individual outcome, both alone and in combination with the PET signal, needs to be assessed.

We, therefore, set out to quantify our ability to predict individual response to treatment using a subject's PET scan, her early cumulative clinic attendance, or both, in a study of behavioral treatment for cocaine dependence. Our hypotheses are: 1) The ventral striatal PET signal is a good predictor for treatment responses at an individual level; 1a) combining the PET signals from other regions-of-interest improves predictive performance; 1b) adding other demographic and clinical predictors, including baseline severity measures, improves predictive performance; 2) early clinic attendance alone is a good predictor of eventual treatment responses; and 3) while the PET signals and early clinic attendance may share some common variance, combining these variables in a multimodal model improves predictive performance.

2. Methods

2.1 Subjects and Procedure

Details of the procedures were previously published (Martinez et al., 2011). All participants were active cocaine users recruited from the community as an independent study, approved by the IRB of the New York State Psychiatric Institute. The screening evaluation included the Structured Clinical Interview for DSM-IV disorders (SCID-IV), an independent psychiatric evaluation, and a medical exam. Individuals were eligible to participate if they were between the ages of 21 and 45 and met criteria for a current DSM-IV diagnosis of cocaine dependence. Currently meeting criteria for other Axis I DSM-IV disorders, lifetime criteria for dependence on a psychostimulant other than cocaine, reporting the current use of psychotropic medication or the use of opiates, sedative-hypnotics, or cannabis more than twice per week, and/or the presence of severe medical (e.g., hypertension, active hepatitis) or neurological illnesses were exclusionary. Twenty-five participants completed a baseline assessment session, met inclusion criteria, and gave informed consent to participate in the study. The sample average age is 37, predominantly male and was racially mixed. For further details of the demographic characteristics, see (Martinez et al., 2011).

2.2 PET and MRI Scanning

PET scans were obtained on the ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, Tenn.) in three-dimensional mode over 60 minutes. For each subject, [^{11}C]raclopride was administered as a bolus and participants underwent two PET scans: one at baseline and one following oral methylphenidate (60 mg) administration (Martinez et al., 2003). The PET data were analyzed using Simplified Reference Tissue Modeling (Martinez et al., 2009a, 2009b), with cerebellum as a reference tissue. Binding potential (BP_{ND}) was defined as:

$$BP_{ND} = f_{ND} * \frac{B_{AVAIL}}{K_D}$$

where the subscript ND indicates the measurement was with respect to the non-displaceable concentration of [^{11}C]raclopride in brain tissue (free plus nonspecifically bound), f_{ND} was the free fraction of the non-displaceable tracer concentration, B_{AVAIL} was the concentration of $D_{2/3}$ receptors available for the tracer to bind to (pmoles per cm^3 of tissue), and K_D (pmoles per cm^3 of tissue) was the inverse of the affinity of the radiotracer for the receptor. The relative change in [^{11}C]raclopride BP_{ND} (BP_{ND}) following methylphenidate administration was calculated as

$$\Delta BP_{ND} = 100\% \times (BP_{ND \text{ baseline}} - BP_{ND \text{ methylphenidate}}) / BP_{ND \text{ baseline}}.$$

Data were analyzed in regions-of-interest (ROI), using each subject's magnetic resonance imaging (MRI) scan (GE Signa EXCITE 3T/94 cm scanner, GE Medical Systems, Milwaukee, Wis.) to determine the subdivisions of the striatum. These included the ventral (limbic) striatum (VST), the anterior and posterior caudate (DCA and PCA) and the anterior and posterior putamen (ANP and POP) (Figure 1A). Regional values were expressed as $BP_{ND\text{-ROI}}$ or $BP_{ND\text{-ROI}}$ (e.g. $BP_{ND\text{-VST}}$, $BP_{ND\text{-VST}}$).

2.3 Treatment

The cocaine-dependent subjects were enrolled in treatment consisting of contingency management with the Community Reinforcement Approach (CRA; Higgins and Budney, 1993; Higgins et al., 2003), using a standard NIDA manual (Budney and Higgins, 1998). Each patient received twice weekly, supervised therapy, and obtained voucher points for each negative urine sample. They had three clinic visits per week. The voucher points (worth \$0.25 each) were escalated starting at 10 points for the first cocaine negative sample, with each subsequent negative sample increasing the voucher value by 5 points. Extra points were given for three consecutive cocaine-free urine samples. A urine test positive for cocaine metabolites would reset the escalating schedule to the starting point (i.e., \$ 0.25). Patients did not lose any of the voucher points earned prior to a positive cocaine urinalysis (UA). All patients were followed through the end of the study, but missed appointments were recorded as positive UAs and reset the escalating schedule. In the first 12 weeks, the voucher reward was escalated and the total maximum amount that could be earned over that period was \$997.50, which reflected complete abstinence and perfect attendance. In weeks 13-24, subjects earned a lottery ticket for each cocaine free UA, an incentive system developed to maintain gains and fade out the high magnitude reinforcements put in place at the beginning of treatment. Treatment response was dichotomized into “responders” and “non-responders” by the end of 12 weeks using a threshold of voucher money earned (\$163), based on the observation that voucher earnings appeared bimodal. This threshold also defined the patients with a month of abstinence versus those demonstrating less than a month of abstinence. Of the 15 non-responders, 10 earned less than \$100 of vouchers and 5 between 100 and 163. (Martinez et al., 2009a).

2.4 Construction of Predictive Models

In the original study (Martinez et al., 2011), the BP_{ND} signal in the ventral striatum (BP_{ND} -VST) was significantly higher in the treatment responders group compared with the non-responders group. BP_{ND} in the posterior caudate (BP_{ND} -PCA) and baseline BP_{ND} in the ventral striatum (BP_{ND} -VST) and anterior putamen (BP_{ND} -POP) also trended toward significance. A quick calculation shows that the intergroup difference in BP_{ND} -VST remains significant when it is adjusted for BP_{ND} and baseline BP_{ND} in other ROIs (ANCOVA, $p < 0.001$). There is also a significant difference between responders and non-responders across all ROIs in the striatum (MANOVA, $p < 0.001$).

We evaluated the capacity for BP_{ND} -VST as a predictor for individual treatment response using predictive models with logistic regression. The univariate model with BP_{ND} -VST as the sole predictor was

$$\text{Log(odds)}_R = \beta_1(\Delta BP_{ND} - \text{VST}) + \varepsilon$$

where R was treatment response and ε an error term. The regression coefficient β_1 was obtained from standard maximum likelihood procedures. Adding a second predictor, for example BP_{ND} -DCA, to the model led to

$$\text{Log(odds)}_R = \beta_1(\Delta BP_{ND} - \text{VST}) + \beta_2(\Delta BP_{ND} - \text{DCA}) + \varepsilon$$

Neuroimaging, demographic or behavioral predictors were added in this way. We visualized a model with two predictors by plotting the line that divided the two groups, i.e. the decision boundary. A typical example, with BP_{ND} -VST and BP_{ND} -POP as predictors, demonstrated the distinct regions of separation and overlap (Figure 1B). We calculated three quantities to measure the performances of the models: the Bayesian Information Criterion (BIC; Schwarz, 1978), a commonly used, non-cross-validated measure for balancing residual error with model complexity, and two cross-validated performance metrics, which measured the generalization error—the error in prediction during cross-validation. In cross-validation, parts of the data (the training set) were selected for fitting model parameters, and the other parts (the testing set), blinded to the model fitting procedure, were used to compare model predictions with real clinical outcomes. In our method, K-fold cross-validation, the original sample was divided into K random subsamples of the same size. A single subsample was retained as the testing set, while the rest were used as the training set, and the process was repeated K times (folds). The advantage of this method, especially for a small sample, was that all samples were used for training and validation (Kohavi, 1995), and the estimate for error was therefore less prone to bias.

Logistic regression yielded a probability measure, and we thresholded the probability to make a binary prediction of response vs. non-response. For performance evaluation, the threshold was set at 0.5, agnostic to preferences for sensitivity (correctly predicting as many true responders as possible) versus specificity (correctly predicting as few true non-

responders as responders as possible). The cross-validated correct rate was the average rate for the model to make a correct prediction. In cross-validated Receiver-Operating-Characteristic (ROC) analysis, false positive rate was plotted against the true positive rate at different levels of threshold. The area under the ROC curve (AUC, c-statistic) was the threshold independent measure of overall predictive performance. As model complexity increased, error during training decreased, but the risk of modeling noise instead of the real predictor-outcome relationship, or overfitting, increased. Such a model performed well in predicting the outcome in the training set but failed in the testing set. In general, measures of predictive performance can be considered generalizations to r^2 in a linear regression model, but r^2 cannot be used as a meaningful comparison for models with different numbers of predictor variables, and does not directly measure generalization error in cross-validation.

2.5 Statistical Analysis

For estimation of confidence intervals, we used 200 randomly re-sampled cross-validated testing sets. For nonlinear classifiers, we used logistic regression using 1st order cross-terms and Support Vector Machines (SVM) with radial basis function kernels (sigma = 1; Burges, 1998). For cross-validation, we used K-fold cross-validation with K=10. All analyses were carried out in MATLAB (Mathworks, Natick MA), using the Statistics Toolbox.

3. Results

3.1 Predictive Models with Only Neuroimaging Predictors

Table 1 (Section A) summarizes the predictive performance for all models with only the PET predictors. Smaller BICs indicate a more likely model, with a difference of greater than 10 considered substantial. Overall, the BP_{ND}-VST models show a significant improvement of performance ($p < 1e-4$, Wilcoxon rank sum test, with Bonferroni correction, alpha = 0.05) compared with models with only the baseline BP_{ND}-VST signal. Models using BP_{ND} from two or more other ROIs generally perform more poorly compared with the model using only BP_{ND}-VST, except for BP_{ND}-VST + BP_{ND}-PCA, which shows a very small improvement. We also constructed a model with both baseline BP_{ND}-VST and BP_{ND}-VST, as both signals were shown to be different between the responders and non-responders in the original study, and again showing no advantage. The model with all ROIs simultaneously is significantly worse. The cross-validated correct rate for the model with BP_{ND}-VST as the sole predictor is 82%. Predictive performance may be improved by incorporating nonlinear effects, either with the addition of interaction terms, or with a machine learning algorithm, such as Support Vector Machine (SVM), that automatically defines nonlinear decision boundaries. Overall these strategies produce worse models, suggesting that nonlinear models are poorly constrained by this dataset, and there is no evidence for interaction effects for different ROIs.

3.2 Predictive Models with Combined Neuroimaging and Demographic Predictors

We next calculate the BIC and cross-validated performance metrics for BP_{ND}-VST combined with several important demographic variables: age, gender, ethnicity (white vs. non-white), socioeconomic status (SES score), concurrent tobacco use (cigarettes per day; Roll et al., 1996), and several baseline measures of cocaine dependence severity, namely

cocaine use prior to study entry (measured by amount of money spent per week on cocaine), and total duration of cocaine use (in years). Inclusion of these demographic variables does not improve the predictive performance (Table 1, Section B). Of note, there is no evidence for correlation between baseline cocaine dependence severity and BP_{ND}-VST or between baseline cocaine dependence severity and response to treatment (Pearson's correlation, $p > 0.05$).

3.3 Predictive Models with Behavioral Predictors

We now turn to designing predictive models that use only behavioral predictors. Cumulative clinic attendance, measured by dividing the total number of follow-up appointments attended by the total possible number of follow-up appointments, is a behavioral metric that measures early treatment response and motivation. We chose cumulative clinic attendance as a predictor rather than negative urine toxicology because even a highly motivated patient who responded to treatment early may still have a few positive urines early on. Negative urine toxicology is also directly related to final outcome by study design, and therefore may be an optimistically biased predictor. In order to assess the predictive performance in time and to estimate how early is *early enough*, we show results of the analysis at the end of week 1, weeks 1 through 2, 1 through 3, and so forth, to 1 through 8, even though the later time points as less useful clinically. Cumulative clinic attendance alone reaches its maximum predictive performance (mean ROC AUC) between week 3 and week 4, with cross-validated correct rate estimated around 83% (Table 1 (Section C), also Figure 2A, B), stabilizes and drops slightly thereafter.

3.4 Predictive Models with Combined Neuroimaging and Behavioral Predictors

Neuroimaging predictors and cumulative clinic attendance share common underlying variances. Cumulative clinic attendance by the end of week 3 is significantly correlated to BP_{ND}-VST (Pearson's $r = 0.57$, $p = 0.003$) and BP_{ND}-POP ($r = 0.51$, $p = 0.01$), but not to other ROIs ($p > 0.05$). To explore whether multimodal models containing both neuroimaging and early treatment attendance have an improved performance, we build two sets of models: one with the cumulative clinic attendance and the BP_{ND}-VST signal, and the other with the cumulative clinic attendance and two neuroimaging predictors BP_{ND}-VST and BP_{ND}-PCA, shown to be the best neuroimaging predictors. Adding BP_{ND}-VST to cumulative clinic attendance does not improve predictive performance after week 2, though does show a higher predictive performance early on (Table 1, Section C also Figure 2A). Adding both BP_{ND}-VST and BP_{ND}-PCA appears to improve predictive performance as early as the end of week 1. Adjusting for multiple comparisons (Pairwise Wilcoxon rank sum test, comparing model with and without neuroimaging predictors at the end of week 3—24 tests, with Bonferroni correction, $p < 0.004$, $\alpha = 0.05$), statistically significant difference in performance is found at the end of week 2 (Table 1 Section C, also Figure 2B). The best performing combined model yields a correct rate of 96% between week 3 and week 4 (Figure 2B, marked in asterisk). More importantly, a consistent trend of performance improvement is seen throughout the rest of the trial period. To quantify the amount of performance improvement, we compare the ROC curves for the best model with behavioral predictors (cumulative clinic attendance at the end of week 3), the best model with only neuroimaging predictors (BP_{ND}-VST + BP_{ND}-PCA), and the best model with

the combined predictors (Figure 3). The percentage improvement in AUC over chance for the combined multimodality predictive model is 26.3% compared with the best model with only neuroimaging data and 29.7% compared with the best model with only behavioral data. Interestingly, the model with neuroimaging predictors only and the model with behavioral predictors only has a crossing in the ROC curves, implying that the neuroimaging predictors perform better when avoiding false-positives (less over-confidence regarding non-responders) and the behavioral predictors perform better reducing false negative results (when we are more interested in not missing anyone who will respond).

4. Discussion

Using a previously published dataset, our analysis shows that a simple linear classifier with a single neuroimaging predictor (BP_{ND} in the ventral striatum) can predict individual treatment outcome with a substantial degree of accuracy (cross-validated correct rate 82%). Adding other clinical predictor variables, including baseline severity measures, does not improve predictive performance. On the other hand, an equally good predictive model can be developed with a behavioral predictor, namely, cumulative clinic attendance over the first 3 weeks of treatment (83% correct rate). Combining the neuroimaging and behavioral predictors into a single multimodal model may significantly improve predictive performance (peaking at 96% correct rate). The best performing combined model only has 3 predictors and its performance compares favorably with the best performing neuroimaging predictive models in other psychiatric disorders (Linden, 2012). Our analysis yields a conceptually distinct but generalizable strategy to combine behavioral predictors of motivation and treatment adherence with neuroimaging predictors or other “biologically-based” predictors.

Methodologically, while investigators increasingly apply pattern recognition methods to understand psychiatric processes, such as the schizophrenia prodromal state (Koutsouleris et al., 2012, 2009), antidepressant responses (Liu et al., 2012), and progression of Alzheimer's disease (Zhang et al., 2012), the resulting models often lack explanatory power and can suffer from overfitting. Most biomarkers exploit either clinical or neuroimaging predictors, but not both. Our analysis shows that a predictive model combining a few regions-of-interest and clinical predictors, identified through a hypothesis-driven approach, can perform as well as complex, high-dimensional models, with the added advantage of better interpretability and adoptability.

Mechanistically, the ventral striatum signal (BP_{ND}-VST) contains most of the predictive information. The ventral striatum receives extensive input from the cortex and the thalamus and processes contextual and affective information (Goto and Grace, 2008; Lovinger, 2010). In particular, the correlation ($r = 0.57$) between BP_{ND}-VST and early clinic attendance and the limited improvement of predictive performance when these two predictors are combined argue that BP_{ND}-VST may be in part a neural correlate of baseline motivation for treatment. Our current results, however, cannot disambiguate whether the relative dopamine deficit, as a possible neural correlate to a lack of motivation/reward sensitivity, is a premorbid risk factor, or a modifiable state that would reverse gradually with abstinence. The posterior caudate (BP_{ND}-PCA) may contain additional predictive information, though this result should be interpreted with caution given a small sample size. The posterior

caudate may contribute to individual variations in treatment response through its hypothesized role in flexible, adaptive reversal learning of a habituated response (Clatworthy et al., 2009).

Clinically, our results have a number of implications. Our results represent a proof-of-principle and point to a potentially promising future of combining neuroimaging and clinical biomarkers. However, while the striatal PET signals offer mechanistic insights, they are currently difficult and expensive to obtain, and not feasible as a clinical assay. Achieving greater availability of substance use treatment is an important and imminent public health priority. Our results additionally suggest a patient stratification algorithm for an expensive treatment paradigm such as contingency management, and represent a first step for optimal resource allocation. We show that early clinic attendance in the first 3 weeks is able to predict outcome as well as the most informative PET signal (BP_{ND} -VST). After 3 or 4 weeks of behavioral treatment with contingency management, treatment response can be assessed, and, for those patients not already responding, one might consider either adding another treatment, such as a medication, or changing treatment strategies altogether. Such response-adaptive treatment programs can be tested in the future with adaptive randomization trials or in large systems implementation or quality improvement projects. From a treatment development perspective, our results as well as results from both human studies (Martinez et al., 2011, 2009a) and animal models (Thanos et al., 2001) suggest that refractory patients may benefit from medications that increase dopamine neurotransmission in the ventral striatum through enhancement of motivation. In fact, levodopa has shown some promise for treatment of cocaine dependence, when combined with voucher incentives contingent on abstinence (Mooney et al., 2007). Other promising strategies combine contingency management with medications such as topiramate/amphetamine (Mariani et al., 2012), modafinil/amphetamine (Schmitz et al., 2012) and bupropion (Poling et al., 2006). All of these medications may increase dopamine neurotransmission in the striatum. These results suggest that early clinic attendance may be used as a proxy for the PET signal for patient stratification for these medications.

An important limitation of the current study is the small sample size and lack of an independent validation dataset. While the cross-validation methodology provides strong evidence for internal validity, the specifics of the study protocol may introduce systematic biases in prediction that limit external validity. The small sample size, while comparing favorably to similar studies of neuroimaging biomarkers of psychiatric illnesses, is also a significant limitation. Our analysis provides an estimate of the overall error rate, but because of the small sample size, the variance for this estimate is large, and the models likely work better for some than for others. While our model selection and evaluation process is primarily driven by existing hypotheses and clinical applicability, we recognize that testing multiple models using a small dataset can be unreliable (Button et al., 2013; Isaksson et al., 2008), and that the proper way to compare and establish superiority of one predictive model versus another using cross-validation and permutation tests is still controversial (Friston, 2012; Lindquist et al., 2013). Replication in independent samples is needed. Mechanistically, our PET measures probe only the dopamine circuitry in the striatum. Other functional networks, such as the striatal-prefrontal network (Park et al., 2010), may also be informative in outcome prediction in substance use disorders. Finally, our findings apply

specifically to Community Reinforcement Approach with contingency management, and future work will examine whether similar findings will hold for other treatment modalities.

Acknowledgments

Role of Funding Source: Funding for this study was provided by NIDA grant R01 DA020855 (PI: Martinez) and in part by NIDA grant K24 DA022412 (PI: Nunes) and NIDA grant K23 DA021850 (PI: Carpenter), and a grant from the Leon Levy Foundation (PI: Luo). NIDA and the Leon Levy Foundation had no further role in study design; in collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

References

- Bisaga A, Aharonovich E, Cheng WY, Levin FR, Mariani JJ, Raby WN, Nunes EV. A placebo-controlled trial of memantine for cocaine dependence with high-value voucher incentives during a pre-randomization lead-in period. *Drug Alcohol Depend.* 2010; 111:97–104. [PubMed: 20537812]
- Bisaga A, Aharonovich E, Garawi F, Levin FR, Rubin E, Raby WN, Nunes EV. A randomized placebo-controlled trial of gabapentin for cocaine dependence. *Drug Alcohol Depend.* 2006; 81:267–274. [PubMed: 16169160]
- Bisaga A, Aharonovich E, Garawi F, Levin FR, Rubin E, Raby WN, Vosburg SK, Nunes EV. Utility of lead-in period in cocaine dependence pharmacotherapy trials. *Drug Alcohol Depend.* 2005; 77:7–11. [PubMed: 15607836]
- Breier A, Su TP, Saunders R, Carson RE, Kolachana BS, de Bartolomeis A, Weinberger DR, Weisenfeld N, Malhotra AK, Eckelman WC, Pickar D. Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. *Proc Natl Acad Sci U S A.* 1997; 94:2569–2574. [PubMed: 9122236]
- Brensilver M, Heinzerling KG, Swanson AN, Shoptaw SJ. Placebo-group responders in methamphetamine pharmacotherapy trials: the role of immediate establishment of abstinence. *Exp Clin Psychopharmacol.* 2012; 20:430–435. [PubMed: 22867036]
- Budney, AJ.; Higgins, ST. A Community Reinforcement Plus Vouchers Approach: Treating Cocaine Addiction: Therapy Manuals for Drug Addiction, Manual 2. National Institute on Drug Abuse; Rockville, Md: 1998.
- Burges CJ. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discover.* 1998; 2:121–167.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* 2013; 14:365–376. [PubMed: 23571845]
- Chawdhary A, Sayre SL, Green C, Schmitz JM, Grabowski J, Mooney ME. Moderators of delay tolerance in treatment-seeking cocaine users. *Addict Behav.* 2007; 32:370–376. [PubMed: 16713126]
- Clatworthy PL, Lewis SJ, Brichard L, Hong YT, Izquierdo D, Clark L, Cools R, Aigbirhio FI, Baron JC, Fryer TD, Robbins TW. Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. *J Neurosci.* 2009; 29:4690–4696. [PubMed: 19369539]
- Friston K. Ten ironic rules for non-statistical reviewers. *NeuroImage.* 2012; 61:1300–1310. [PubMed: 22521475]
- Goto Y, Grace AA. Limbic and cortical information processing in the nucleus accumbens. *Trends Neurosci.* 2008; 31:552–558. [PubMed: 18786735]
- Heinz A, Siessmeier T, Wrase J, Buchholz HG, Gruner G, Kumakura Y, Cumming P, Schreckenberger M, Smolka MN, Rosch F, Mann K, Bartenstein P. Correlation of alcohol craving with striatal dopamine synthesis capacity and D2/3 receptor availability: a combined [18F]DOPA and [18F]DMFP PET study in detoxified alcoholic patients. *Am J Psychiatry.* 2005; 162:1515–1520. [PubMed: 16055774]

- Heinz A, Siessmeier T, Wrase J, Hermann D, Klein S, Grusser SM, Flor H, Braus DF, Buchholz HG, Grunder G, Schreckenberger M, Smolka MN, Rosch F, Mann K, Bartenstein P. Correlation between dopamine D(2) receptors in the ventral striatum and central processing of alcohol cues and craving. *Am J Psychiatry*. 2004; 161:1783–1789. [PubMed: 15465974]
- Higgins ST, Budney AJ. Treatment of cocaine dependence through the principles of behavior analysis and behavioral pharmacology. *NIDA Res Monogr*. 1993; 137:97–121. [PubMed: 8289930]
- Higgins ST, Sigmon SC, Wong CJ, Heil SH, Badger GJ, Donham R, Dantona RL, Anthony S. Community reinforcement therapy for cocaine-dependent outpatients. *Arch Gen Psychiatry*. 2003; 60:1043–1052. [PubMed: 14557150]
- Isaksson A, Wallman M, Goransson H, Gustafsson MG. Cross-validation and bootstrapping are unreliable in small sample classification. *Pattern Recog Lett*. 2008; 29:1960–1965.
- Kampman K, Volpicelli JR, Alterman A, Cornish J, Weinrieb R, Epperson L, Sparkman T, O'Brien CP. Amantadine in the early treatment of cocaine dependence: a double-blind, placebo-controlled trial. *Drug Alcohol Depend*. 1996; 41:25–33. [PubMed: 8793307]
- Kampman KM, Alterman AI, Volpicelli JR, Maany I, Muller ES, Luce DD, Mulholland EM, Jawad AF, Parikh GA, Mulvaney FD, Weinrieb RM, O'Brien CP. Cocaine withdrawal symptoms and initial urine toxicology results predict treatment attrition in outpatient cocaine dependence treatment. *Psychol Addict Behav*. 2001; 15:52–59. [PubMed: 11255939]
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*. 1995:1137–1145.
- Koutsouleris N, Gaser C, Patschurek-Kliche K, Scheuerecker J, Bottlender R, Decker P, Schmitt G, Reiser M, Moller HJ, Meisenzahl EM. Multivariate patterns of brain-cognition associations relating to vulnerability and clinical outcome in the at-risk mental states for psychosis. *Hum Brain Mapp*. 2012; 33:2104–2124. [PubMed: 22887825]
- Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Moller HJ, Gaser C. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. 2009; 66:700–712. [PubMed: 19581561]
- Laruelle M. Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. *J Cereb Blood Flow Metab*. 2000; 20:423–451. [PubMed: 10724107]
- Linden DE. The challenges and promise of neuroimaging in psychiatry. *Neuron*. 2012; 73:8–22. [PubMed: 22243743]
- Lindquist MA, Caffo B, Crainiceanu C. Ironing out the statistical wrinkles in “ten ironic rules”. *NeuroImage*. 2013; 81:499–502. [PubMed: 23587691]
- Liu F, Guo W, Yu D, Gao Q, Gao K, Xue Z, Du H, Zhang J, Tan C, Liu Z, Zhao J, Chen H. Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. *PLoS One*. 2012; 7:e40968. [PubMed: 22815880]
- Lovinger DM. Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology*. 2010; 58:951–961. [PubMed: 20096294]
- Mariani JJ, Pavlicova M, Bisaga A, Nunes EV, Brooks DJ, Levin FR. Extended-release mixed amphetamine salts and topiramate for cocaine dependence: a randomized controlled trial. *Biol Psychiatry*. 2012; 72:950–956. [PubMed: 22795453]
- Martinez D, Carpenter KM, Liu F, Slifstein M, Broft A, Friedman AC, Kumar D, Van Heertum R, Kleber HD, Nunes E. Imaging dopamine transmission in cocaine dependence: link between neurochemistry and response to treatment. *Am J Psychiatry*. 2011; 168:634–641. [PubMed: 21406463]
- Martinez D, Greene K, Broft A, Kumar D, Liu F, Narendran R, Slifstein M, Van Heertum R, Kleber HD. Lower level of endogenous dopamine in patients with cocaine dependence: findings from PET imaging of D(2)/D(3) receptors following acute dopamine depletion. *Am J Psychiatry*. 2009a; 166:1170–1177. [PubMed: 19723785]
- Martinez D, Narendran R, Foltin RW, Slifstein M, Hwang DR, Broft A, Huang Y, Cooper TB, Fischman MW, Kleber HD, Laruelle M. Amphetamine-induced dopamine release: markedly

- blunted in cocaine dependence and predictive of the choice to self-administer cocaine. *Am J Psychiatry*. 2007; 164:622–629. [PubMed: 17403976]
- Martinez D, Slifstein M, Broft A, Mawlawi O, Hwang DR, Huang Y, Cooper T, Kegeles L, Zarahn E, Abi-Dargham A, Haber SN, Laruelle M. Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: amphetamine-induced dopamine release in the functional subdivisions of the striatum. *J Cereb Blood Flow Metab*. 2003; 23:285–300. [PubMed: 12621304]
- Martinez D, Slifstein M, Narendran R, Foltin RW, Broft A, Hwang DR, Perez A, Abi-Dargham A, Fischman MW, Kleber HD, Laruelle M. Dopamine D1 receptors in cocaine dependence measured with PET and the choice to self-administer cocaine. *Neuropsychopharmacology*. 2009b; 34:1774–1782. [PubMed: 19177067]
- Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012; 16:72–80. [PubMed: 22177032]
- Mooney ME, Schmitz JM, Moeller FG, Grabowski J. Safety, tolerability and efficacy of levodopa-carbidopa treatment for cocaine dependence: two double-blind, randomized, clinical trials. *Drug Alcohol Depend*. 2007; 88:214–223. [PubMed: 17134849]
- Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, Heinz A. Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *J Neurosci*. 2010; 30:7749–7753. [PubMed: 20519550]
- Poling J, Kosten TR, Sofuoglu M. Treatment outcome predictors for cocaine dependence. *Am J Drug Alcohol Abuse*. 2007; 33:191–206. [PubMed: 17497542]
- Poling J, Oliveto A, Petry N, Sofuoglu M, Gonsai K, Gonzalez G, Martell B, Kosten TR. Six-month trial of bupropion with contingency management for cocaine dependence in a methadone-maintained population. *Arch Gen Psychiatry*. 2006; 63:219–228. [PubMed: 16461866]
- Roll JM, Higgins ST, Budney AJ, Bickel WK, Badger GJ. A comparison of cocaine-dependent cigarette smokers and non-smokers on demographic, drug use and other characteristics. *Drug Alcohol Depend*. 1996; 40:195–201. [PubMed: 8861397]
- Schmitz JM, Rathnayaka N, Green CE, Moeller FG, Dougherty AE, Grabowski J. Combination of modafinil and d-amphetamine for the treatment of cocaine dependence: a preliminary investigation. *Front Psychiatry*. 2012; 3:77. [PubMed: 22969732]
- Schultz W. Multiple dopamine functions at different time courses. *Annu Rev Neurosci*. 2007; 30:259–288. [PubMed: 17600522]
- Schwarz GE. Estimating the dimension of a model. *Ann Stat*. 1978; 6:461–464.
- Sigmon SC, Higgins ST. Voucher-based contingent reinforcement of marijuana abstinence among individuals with serious mental illness. *J Subst Abuse Treat*. 2006; 30:291–295. [PubMed: 16716843]
- Thanos PK, Volkow ND, Freimuth P, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann R. Overexpression of dopamine D2 receptors reduces alcohol self-administration. *J Neurochem*. 2001; 78:1094–1103. [PubMed: 11553683]
- Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. *Science*. 2005; 307:1642–1645. [PubMed: 15761155]
- Volkow ND, Fowler JS, Wang GJ, Baler R, Telang F. Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*. 2009; 56(Suppl 1):3–8. [PubMed: 18617195]
- Volkow ND, Wang GJ, Fowler JS, Logan J, Gatley SJ, Hitzemann R, Chen AD, Dewey SL, Pappas N. Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature*. 1997; 386:830–833. [PubMed: 9126741]
- Volkow ND, Wang GJ, Fowler JS, Logan J, Schlyer D, Hitzemann R, Lieberman J, Angrist B, Pappas N, MacGregor R, et al. Imaging endogenous dopamine competition with [¹¹C]raclopride in the human brain. *Synapse*. 1994; 16:255–262. [PubMed: 8059335]
- Wang GJ, Smith L, Volkow ND, Telang F, Logan J, Tomasi D, Wong CT, Hoffman W, Jayne M, Alia-Klein N, Thanos P, Fowler JS. Decreased dopamine activity predicts relapse in methamphetamine abusers. *Mol Psychiatry*. 2012; 17:918–925. [PubMed: 21747399]
- Zhang D, Shen D, Alzheimer's Disease Neuroimaging I. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*. 2012; 59:895–907. [PubMed: 21992749]

Figure 1.

Predictive models with PET predictors (A) Demarcating ROIs in the striatum using anatomical landmarks, with ROIs labeled on the right (B) Visualizing the decision boundary between responders and non-responders reveals a region of ambiguity between the two groups. Colored lines indicate 75% probability of response contours for a 2-D model of $\text{BP}_{\text{ND-VST}}$ and $\text{BP}_{\text{ND-POP}}$. (C) Performance measures for various models showing that the performance is approximately the same for all linear models of classification, but change of binding potentials ($\text{BP}_{\text{ND-ROIs}}$) are better predictors than baseline binding potentials ($\text{BP}_{\text{ND-ROIs}}$). Abbreviations: ROI – region of interest; BP – change in binding potential; VST - ventral (limbic) striatum; DCA/PCA - anterior and posterior caudate; ANP/POP - anterior and posterior putamen; AUC – area under the curve; ROC – Receiver-operating-characteristics

Figure 2.

Combining neuroimaging and behavioral predictors. (A) Comparing models using only clinic attendance rate versus a model with the additional Δ BPND-VST signal showing no significant gain in predictive performance. (B) The combined model with Δ BPND-VST + Δ BPND-PCA + average clinic attendance shows a significant improvement of predictive performance, reaching statistical significance, adjusted for multiple comparisons, at the end of week 2, peaking between weeks 3 and 4 (asterisk). Abbreviations: Δ BPND – change in binding potential; VST - ventral (limbic) striatum; PCA – posterior caudate; AUC – area under the curve; ROC – Receiver-operating-characteristics

Figure 3.

ROC Analyses of the best models. The best predictive models using only behavioral or only neuroimaging predictors achieve a substantial performance over chance, and combining the two modalities improves overall predictive performance. There appears to be some distinct areas of coverage, with crossing of ROCs suggesting that behavioral predictors perform better if higher sensitivity is desired, and neuroimaging predictors perform better if higher specificity is desired. ROC – Receiver-operating-characteristics

Summary of performance of models predicting clinical response (sustained abstinence) among cocaine dependent patients treated with Community Reinforcement Approach (CRA) plus voucher incentives contingent on abstinence. (A) Performance for models with only PET signals, namely baseline [¹¹C]-raclopride binding potential (BP_{ND}), and change in BP after methylphenidate administration (BP_{ND}). Model performance is measured by BIC (Bayesian Information Criterion) and cross-validated ROCs (Receiver Operating Characteristic curves) and cross-validated percent correct rate for binary prediction (responder vs. non-responder). Smaller BIC indicates a more likely model, and the difference in BIC value gives the relative likelihood of different values, with a difference > 10 considered strong. None of the models achieves a BIC less than the model with only BP_{ND}-VST, except BP_{ND}-VST + BP_{ND}-PCA, only 4 units lower. (B) Same analysis for BP_{ND}-VST plus an additional clinical demographic variable, showing that adding these predictors does not improve model performance (C) Performance of models based on only the average clinic attendance rate at the end of each of the weeks and combined models with both behavioral and neuroimaging predictors, showing a significant improvement in predictive performance.

Table 1

MODEL	BIC	Mean ROC AUC	95% CI		Correct Rate
Section A					
BP _{ND} -VST only	32.0	0.63	0.09	0.09	54%
BP _{ND} -VST + BP _{ND} -ANP	34.3	0.55	0.10	0.09	56%
BP _{ND} -VST + BP _{ND} -DCA	34.6	0.55	0.11	0.08	56%
BP _{ND} -VST + BP _{ND} -PCA	35.0	0.58	0.10	0.08	51%
BP _{ND} -VST + BP _{ND} -POP	35.1	0.57	0.10	0.11	52%
Baseline BP for All ROIs	43.6	0.45	0.09	0.08	47%
BP _{ND} -VST only	24.2	0.85	0.06	0.04	82%
BP _{ND} -VST + BP _{ND} -ANP	26.4	0.82	0.06	0.05	78%
BP _{ND} -VST + BP _{ND} -DCA	26.1	0.83	0.06	0.06	75%
BP _{ND} -VST + BP _{ND} -PCA	20.8	0.87	0.06	0.05	81%
BP _{ND} -VST + BP _{ND} -POP	26.4	0.80	0.06	0.06	68%
BP for All ROIs	24.4	0.74	0.09	0.06	81%
BP _{ND} -VST + BP _{ND} -VST	24.5	0.80	0.07	0.07	77%
Section B					
BP _{ND} -VST + AGE	27.4	0.80	0.07	0.06	76%
BP _{ND} -VST + GENDER	26.5	0.83	0.09	0.06	76%
BP _{ND} -VST + ETHNICITY	26.4	0.81	0.08	0.07	76%

MODEL	BIC	Mean ROC AUC	95% CI	Correct Rate
BP _{ND} -VST + CURRENT USE	25.7	0.83	0.05 0.06	75%
BP _{ND} -VST + DURATION OF USE	26.2	0.82	0.06 0.06	75%
BP _{ND} -VST + SES	25.4	0.79	0.06 0.06	77%
BP _{ND} -VST + TOBACCO	26.1	0.82	0.06 0.06	74%
Section C				
AVG AT THE END OF WEEK 1	31.2	0.73	0.07 0.07	79%
AVG AT THE END OF WEEK 2	23.7	0.81	0.08 0.08	78%
AVG AT THE END OF WEEK 3	20.4	0.88	0.06 0.05	83%
AVG AT THE END OF WEEK 4	19.2	0.90	0.04 0.05	83%
BP _{ND} -VST + WEEK 1	32.9	0.79	0.09 0.07	68%
BP _{ND} -VST + WEEK 2	25.4	0.84	0.08 0.06	76%
BP _{ND} -VST + WEEK 3	23.3	0.85	0.07 0.06	81%
BP _{ND} -VST + WEEK 4	20.6	0.85	0.06 0.05	85%
BP _{ND} -VST + BP _{ND} -PCA + WEEK 1	28.6	0.80	0.08 0.07	80%
BP _{ND} -VST + BP _{ND} -PCA + WEEK 2	18.1	0.97	0.09 0.06	94%
BP _{ND} -VST + BP _{ND} -PCA + WEEK 3	9.5	0.98	0.06 0.05	96%
BP _{ND} -VST + BP _{ND} -PCA + WEEK 4	9.5	0.93	0.04 0.03	91%

Abbreviations: VST – ventral striatum; ANP – anterior putamen, DCA – anterior caudate, PCA – posterior caudate; POP – posterior caudate; BP_{ND} – baseline binding potential; BP_{ND} – change of binding potential; SES – socioeconomic status