# Predicting Adolescent Cannabis Use vs. Binge Drinking Using Random Forests

Philip A. Spechler[1,2], Bader Chaarani[2], Nicholas Allgaier[2], Scott Mackey[2], Robert Althoff[1,2], Hugh Garavan[1,2] & the IMAGEN Consortium

NERVE LAB
NEuroimaging Research VErmont

Vermont Center on Behavior & Health
The University of Vermont

Department of Psychological Science[1], Department of Psychiatry[2]
University of Vermont, Burlington, VT USA
Email: philip.spechler@uvm.edu

The University of Vermont
LARNER COLLEGE OF MEDICINE

## Introduction

- A goal for machine learning and psychiatric neuroimaging is to develop predictive models that differentially predict similar yet distinct phenotypes.
- Cannabis use and binge drinking are similar patterns of drug use in adolescence, but may be predicted by distinct features.
- Uncovering the predictors that differentially predict cannabis use vs. binge drinking in adolescence will inform etiological mechanisms and targets for intervention.

## Methods

- Participants were drawn from the longitudinal IMAGEN study and selected for being cannabis-naïve at the baseline assessment (age 14).
- A group of individuals who used cannabis (and some binge drinking) by age 16 were compared to a closely matched group of individuals who only had binge drinking experiences by age 16. Sex-specific samples were identified (Table 1).

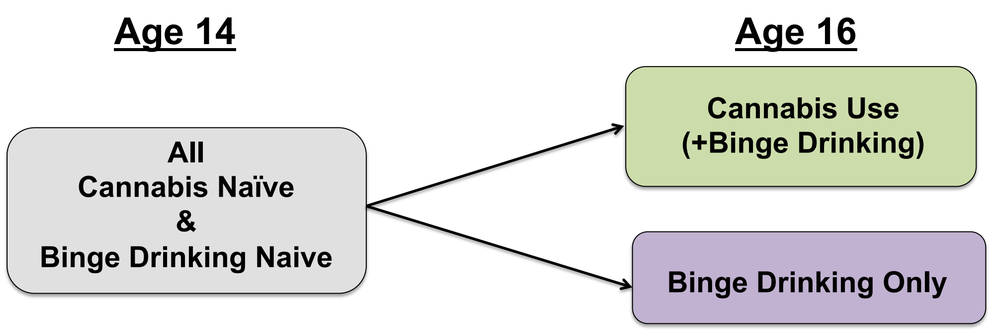



**Table 1**: Cannabis & Binge Drinking Levels For Each Group by Sex

| Use Level | Males (N=178) | | | Females (N=148) | | |
|---|---|---|---|---|---|---|
| | Cannabis Use (+Binge Drinking) Group n=89 | | Binge Drinking Only Group n=89 | Cannabis Use (+Binge Drinking) Group n=74 | | Binge Drinking Only Group n=74 |
| | Cannabis Level | Binge Drinking Level | Binge Drinking Level | Cannabis Level | Binge Drinking Level | Binge Drinking Level |
| 1-2x | 39 | 32 | 32 | 22 | 26 | 26 |
| 3-5x | 10 | 30 | 30 | 18 | 25 | 25 |
| 6-9x | 8 | 14 | 14 | 6 | 8 | 8 |
| 10-19x | 6 | 9 | 9 | 11 | 9 | 9 |
| 20-39x | 9 | 3 | 3 | 5 | 3 | 3 |
| 40+ | 17 | 1 | 1 | 12 | 3 | 3 |
| **Total** | **89** | | **89** | **74** | | **74** |

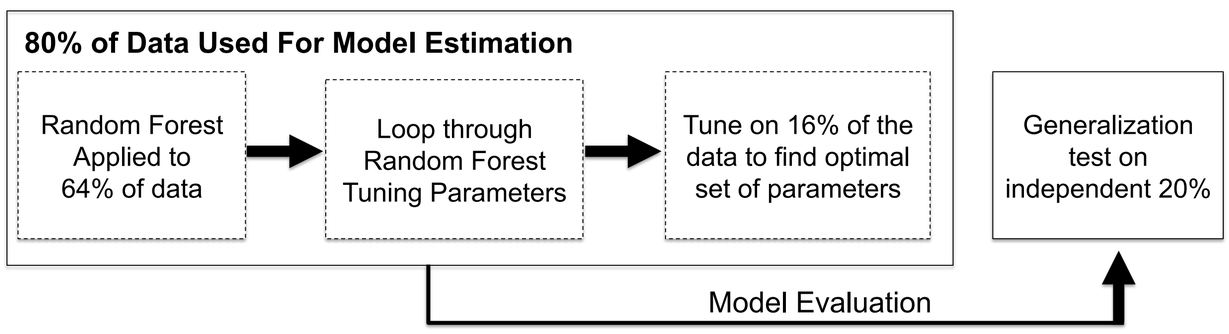
**Table 2**: Summary of Data Used as Predictors

| Domain | Measures | Data points |
|---|---|---|
| ***Psychosocial*** | • Demographics<br>• Cognitive assessments<br>• Personality assessment<br>• Life-events questionnaires<br>• Baseline cigarette use | • 80 measures |
| ***SNPs*** | • A-priori SNPs | • 108 SNPs |
| ***Structural Neuroimaging*** | • Total GMV<br>• Gray-Matter Volume ROIs | • 1 total GMV<br>• 278 GMV ROIs |
| ***Functional Neuroimaging*** | • Reward Task (2 Contrasts)<br>• Stop Signal Task (2 Contrasts)<br>• Face Processing Task (3 Contrasts) | • 278 per contrast<br>• 1946 Total ROIs |
| **Predictors per subject** | | **2413** |

## Machine Learning Methods

- Random Forests (www.scikit-learn.org) with nested 5-Fold cross validation were used to predict cannabis use vs. binge drinking for each sex.
- Prediction performance was evaluated using ROC AUC on the independent test set.

**Figure 1**: Random Forest Nested 5-Fold Cross Validation Framework




## Results

**Table 3**: Random Forest Prediction Accuracies

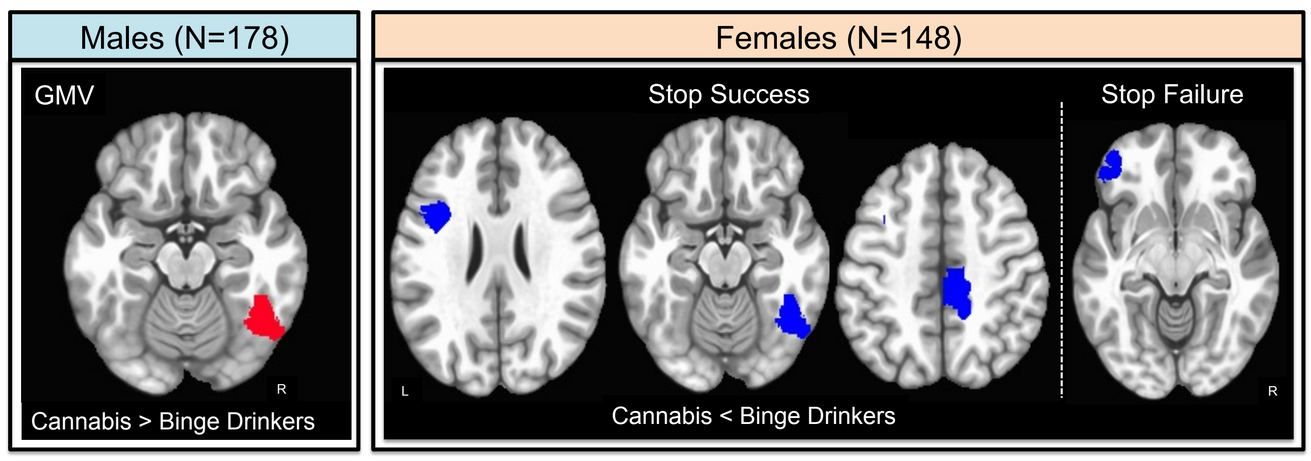
| Males (N=178) | | Females (N=148) | |
|---|---|---|---|
| AUC | P-value | AUC | P-value |
| 0.6286 | $1 \times 10^{-4}$ | 0.6593 | $1 \times 10^{-5}$ |

**Table 4**: Identified Predictors For Each Sex

| Domain | | Males | Males & Females | Females |
|---|---|---|---|---|
| ***Psychosocial*** | Drug Use (ESPAD) | | Lifetime Cigarettes* | |
| | Life Events (LEQ) | | Positive Feelings Toward Deviant Behaviors* | |
| ***Neuroimaging*** | Gray Matter Volume | R. Inferior Temporal Gyrus* | | |
| | Response Inhibition: Stop Success | | | L. Middle Frontal Gyrus† |
| | | | | R. Inferior Temporal Gyrus† |
| | | | | R. Paracentral Lobule† |
| | Response Inhibition: Stop Failure | | | L. Inferior Frontal Gyrus† |

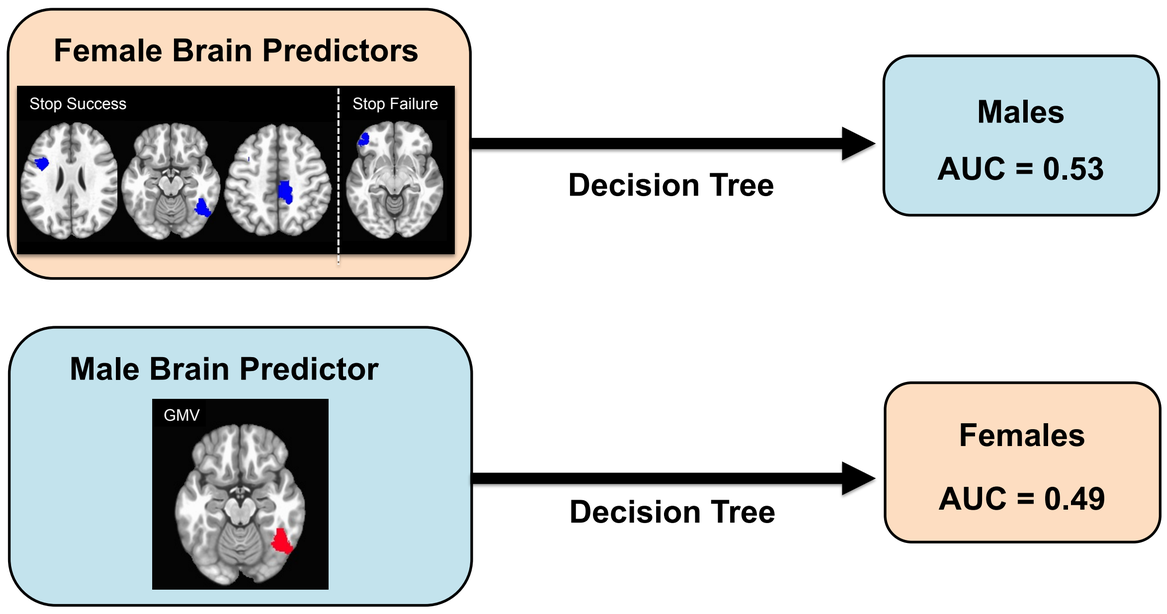
*Cannabis Users > Binge Drinkers
†Cannabis Users < Binge Drinkers

**Figure 2**: Brain Predictors For Each Sex




## Sex Specificity Results

- The identified brain predictors for one sex were used in a post-hoc decision tree predicting future cannabis use vs. binge drinking for the opposite sex




## Conclusions

- Random forests can reliably predict future cannabis use vs. binge drinking in adolescent males and females using data collected prior to exposure.
- A sparse set of predictors were identified despite starting with a very large set of multi-domain predictors.
- Two identified psychosocial predictors were common across the sexes whereas the identified neuroimaging predictors were sex-specific.
- Lower activations in brain regions supporting inhibitory control constitutes a female-specific risk profile for cannabis use in adolescence.
- More gray matter volume might indicate a neurodevelopmental delay characterizes males who go on to use cannabis in adolescence.

ACKNOWLEDGEMENTS: This work supported from the following sources: An FDA grant P50DA036114 awarded to the Vermont Center on Behavior and Health, and, the National Institutes of Health Center of Biomedical Research Excellence award P20GM103644 from the National Institute of General Medical Sciences. HG was supported, in part, by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centers of Excellence.