Addiction Treatment as a Sociological Outcome

Christopher Carbonaro

December 10th, 2019

Introduction

Drug Use in the U.S.

- 60.2% (164.8 Million) of Americans have consumed alcohol, tobacco, or an illicit drug within the past month

Drug Use in the U.S.

- 60.2% (164.8 Million) of Americans have consumed alcohol, tobacco, or an illicit drug within the past month
- Of these users, 31.9 million used an illicit drug within the past month

Drug Use in the U.S.

- 60.2% (164.8 Million) of Americans have consumed alcohol, tobacco, or an illicit drug within the past month
- Of these users, 31.9 million used an illicit drug within the past month
- 20% of Americans are estimated to have used an illicit drug within the past year (SAMHSA, 2019, p. 3-8)

Addiction Treatment in the U.S.

Approximately "21.2 million people aged 12 or older needed substance use treatment" in 2018 (SAMHSA, p. 3).

Unfortunately, while this comprises 7.8% of the U.S. population, only 1.4 received treatment within the past year (p. 3).

This Project's Focus

Given the disparity between the number of individuals in need of treatment and the number who receive it, we are motivated to ask:

What sociological factors best predict whether an individual will undergo substance abuse treatment?

Literature Review

Current Research

This project eschews studying addiction as a neurological phenomenon (see Venniro et al., 2017).

Instead, it builds on work which looks at sociological factors which predict addiction (Boyle, Polinsky, & Hser, 2000; Taylor, Caudy, Blasko, & Taxman, 2017; Battjes, Gordon, O'Grady, Kinlock, & Carswell, 2003).

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid*.)

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:
  - Severity of substance being abused

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:
  - Severity of substance being abused
  - Gender

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:
  - Severity of substance being abused
  - Gender
  - Ethnicity

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:
  - Severity of substance being abused
  - Gender
  - Ethnicity
  - Age

Current Research (cont.)

Limitations of current research:

- Small sample sizes (Boyle et. al., 2000; Battjes et. al., 2003)
- No use of analytic tools other than multiple regression (*Ibid.*)
- Luongo et. al. (2016) conducted a Canadian study which produced results conflicting with Boyle et. al. (2000) and Battjes et. al. (2003) regarding the importance of:
  - Severity of substance being abused
  - Gender
  - Ethnicity
  - Age
  - . . . and Education

Filling the Research Gap

This project is an attempt to examine the importance of these variables (and others) on a larger scale while 1. looking for interactions between the predictors and 2. using tools other than OLS.

Methods

Source of Data

This project uses a survey conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA), namely the 2018 National Survey on Drug Use and Health (NSDUH).

The survey consists of roughly 56,000 participants.

Predictors in the Data

There are hundreds of variables for each respondant in the NSDUH; this study narrows the focus to roughly 40 predictors.

These include variables like age and gender, variables describing the individual's health insurance coverage, and variables describing their recency/frequency/type of drug use.

Logistic Regression

Our first tool for examining the data is regularized regression. We use this as a baseline against which to compare our second tool.

Here, we use the elastic net to assess the importance of each variable and avoid producing an overcomplicated model.

Logistic Regression (cont.)

$$log \frac{Pr(G = 2|X = x)}{Pr(G = 1|X = x)} = \beta_0 + \beta^T x$$

Logistic Regression (cont.)

$$\widehat{\beta}^{\text{elastic net}} = \text{argmin} \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

$$+\lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1-\alpha)|\beta_j|)$$

Boosting

One of the biggest problems with linear regression is we have a comparatively small number of positive cases (only around 5%).

This makes classifying positive cases difficult.

Boosting (cont.)

Boosting is a helpful solution here, since it will raise the relative importance of misclassified observations for each subsequent tree.

We can also use partial dependence plots to look for interactions between important terms.

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For b $= 1, 2, \ldots$, B, repeat:

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For b $= 1, 2, \ldots$, B, repeat:

a. Fit a tree $\widehat{f}^b$ with d splits to the training data

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For b $= 1, 2, \ldots$, B, repeat:

a. Fit a tree $\widehat{f}^b$ with d splits to the training data
b. Update $\widehat{f}$ by adding in a shrunken version of the new tree:

$$\widehat{f}(x) \leftarrow \widehat{f}(x) + \lambda \widehat{f}^b(x)$$

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For b = 1, 2, . . . , B, repeat:

a. Fit a tree $\widehat{f}^b$ with d splits to the training data

b. Update $\widehat{f}$ by adding in a shrunken version of the new tree:

$$\widehat{f}(x) \leftarrow \widehat{f}(x) + \lambda \widehat{f}^b(x)$$

c. Update the residuals:

$$r_i \leftarrow r_i - \lambda \widehat{f}^b(x_i)$$

Boosting (cont.)

From Witten, Hastie, & Tibshirani (2017, p. 323):

1. Set $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For b = 1, 2, . . . , B, repeat:

a. Fit a tree $\widehat{f}^b$ with d splits to the training data

b. Update $\widehat{f}$ by adding in a shrunken version of the new tree:

$$\widehat{f}(x) \leftarrow \widehat{f}(x) + \lambda \widehat{f}^b(x)$$

c. Update the residuals:

$$r_i \leftarrow r_i - \lambda \widehat{f}^b(x_i)$$

3. Output the boosted model:

$$\widehat{f}(x) = \sum_{b=1}^{B} \lambda \widehat{f}^b(x)$$

Results

OLS Results

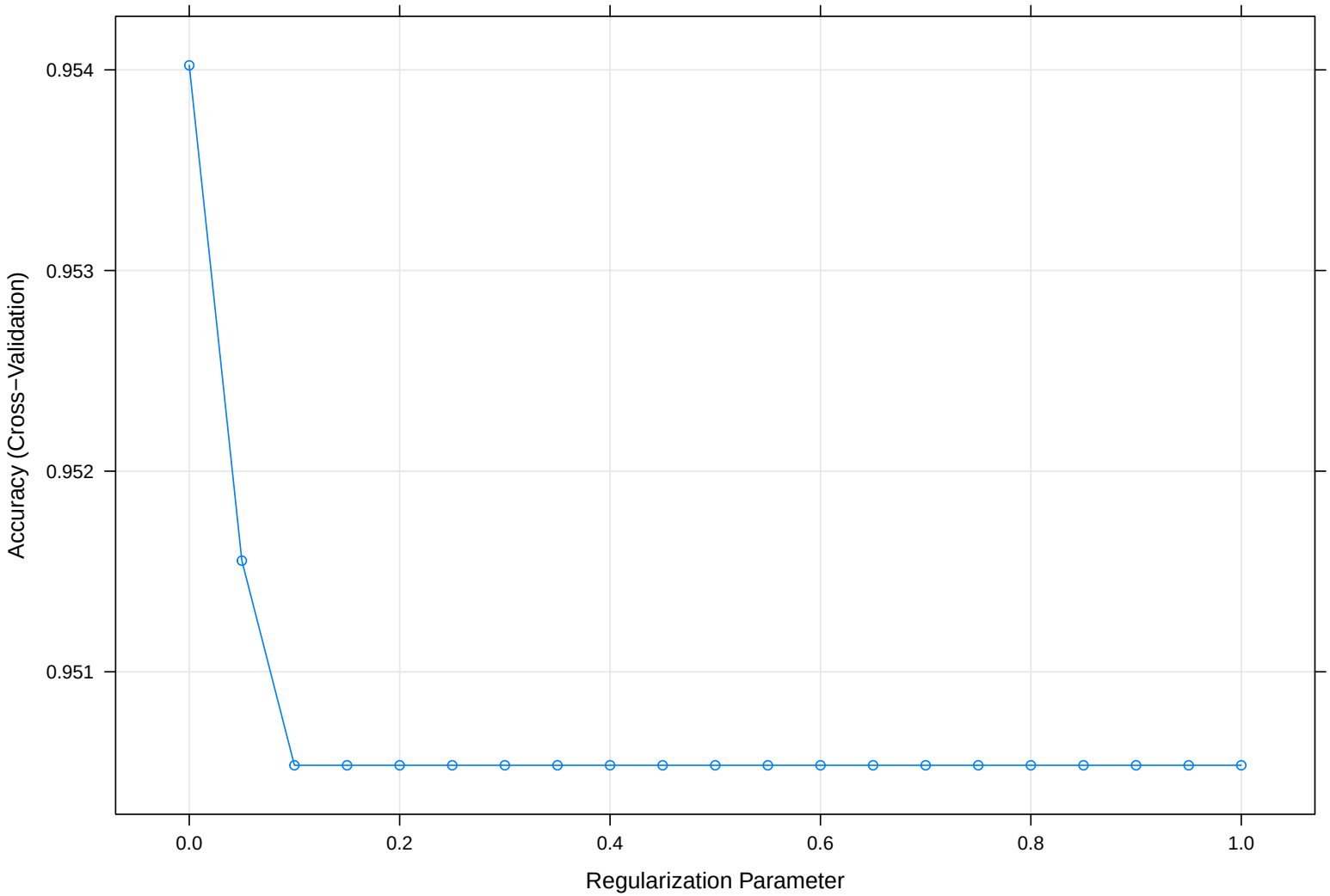

```
knitr::kable(elastic$results)
```

alpha	lambda	Accuracy	Kappa	AccuracySD	KappaSD
0.5	0.00	0.9540226	0.2759743	0.0015548	0.03551
0.5	0.05	0.9515538	0.0698760	0.0005948	0.01774
0.5	0.10	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.15	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.20	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.25	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.30	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.35	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.40	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.45	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.50	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.55	0.9505340	0.0000000	0.0000897	0.00000
0.5	0.60	0.9505340	0.0000000	0.0000897	0.00000

```
summary(elastic$results)
```

```
     alpha          lambda         Accuracy          Ka
 Min.   :0.5   Min.   :0.00   Min.   :0.9505   Min.
 1st Qu.:0.5   1st Qu.:0.25   1st Qu.:0.9505   1st Qu
 Median :0.5   Median :0.50   Median :0.9505   Median
 Mean   :0.5   Mean   :0.50   Mean   :0.9507   Mean
 3rd Qu.:0.5   3rd Qu.:0.75   3rd Qu.:0.9505   3rd Qu
 Max.   :0.5   Max.   :1.00   Max.   :0.9540   Max.
   AccuracySD          KappaSD
 Min.   :0.0000897   Min.   :0.000000
 1st Qu.:0.0000897   1st Qu.:0.000000
 Median :0.0000897   Median :0.000000
 Mean   :0.0001835   Mean   :0.002536
 3rd Qu.:0.0000897   3rd Qu.:0.000000
 Max.   :0.0015548   Max.   :0.035512
```

```
plot(elastic)
```



Boosting Results

Next Steps

Works Cited

Resources

Battjes, R. J., Gordon, M. S., Ogrady, K. E., Kinlock, T. W., & Carswell, M. A. (2003). Factors that Predict Adolescent Motivation for Substance Abuse Treatment. *Journal of Substance Abuse Treatment*, 24(3), 221–232. doi: 10.1016/s0740-5472(03)00022-9

Boyle, K., Polinsky, M. L., & Hser, Y.-I. (2000). Resistance to Drug Abuse Treatment: A Comparison of Drug Users Who Accept or Decline Treatment Referral Assessment. *Journal of Drug Issues*, 30(3), 555–574. doi: 10.1177/002204260003000304

Cohen, J., Ilumoka, A., & Salehi, I. (2015). Neural Network-Based Drug Abuse Treatment Optimization. *Procedia Computer Science*, 61, 454–459. doi: 10.1016/j.procs.2015.09.186

Resources (cont.)

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition* (2nd ed.). New York, NY: Springer New York.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.

Resources (cont.)

The Substance Abuse and Mental Health Services Administration. (2019). The Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. Retrieved from https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf

Luongo, N. M., Dong, H., Kerr, T. H., Milloy, M. J. S., Hayashi, K., & Richardson, L. A. (2017). Income Generation and Attitudes Towards Addiction Treatment Among People who use Illicit Drugs in a Canadian Setting. *Addictive Behaviors*, 64, 159–164. doi: 10.1016/j.addbeh.2016.08.041

Resources (cont.)

Venniro, M., Caprioli, D., Zhang, M., Whitaker, L. R., Zhang, S., Warren, B. L., ... Shaham, Y. (2017). The Anterior Insular Cortex→Central Amygdala Glutamatergic Pathway Is Critical to Relapse after Contingency Management. *Neuron*, 96(2). doi: 10.1016/j.neuron.2017.09.024

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. doi: 10.1177/1745691617693393