# Predicting the Likelihood of Relapse for Select Addictive Substances Following Rehabilitation

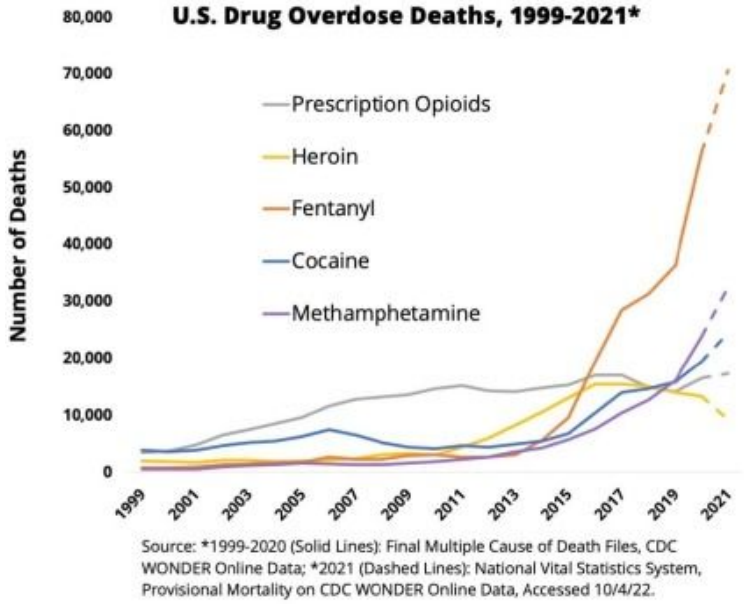
Cathy Wang, Erica King, Yeninda Tjoa

Team Sober Signals

# Table of Contents

# Proposed Research

- **Background:**
  - Substance abuse and drug-related deaths are rising in the U.S.
  - Rehabilitation plays a key role in addiction treatment.
  - Evaluate the effectiveness of rehabilitation in preventing relapse.
- **Objective:**
  - Analyze data on substance users who have undergone rehabilitation
  - Measure relationship between substance abuse and rehabilitation
  - Predict likelihood of relapse after rehabilitation



# Dataset

National Survey on Drug Use and Health 2022 and 2023:

- Measures substance use, mental illness, and treatment in ages 12 and older
- Additional analysis: subset of individuals who participated in the survey in both yea


SAMHSA
Substance Abuse and Mental Health Services Administration

NSDUH
NATIONAL SURVEY ON DRUG USE AND HEALTH

# Survey Questions

- About users: Demographics, Education, Employment, Health Insurance, Income
- Substances: Nicotine, Alcohol, Marijuana, Blunts, Heroin, Hallucinogens, Inhalants, Methamphetamine, Pain Relievers, Tranquilizers, Sedatives and Stimulants usage & dependence & frequency
- Additional features:
  - Age /date of first drug use
  - Substance use disorder, mental health, health
  - Past treatment
  - Market information
  - Risk/availability
  - Social environment, youth experiences

# Data Specifications

## 2023 Raw Data - 56705 rows x 2636 columns

| | CADRKCOCN COCAINE OR "CRACK" | CADRKHERN HEROIN | CADRKHALL HALLUCINOGENS | CADRKINHL INHALANTS |
|---|---|---|---|---|
| 1 | 91 | 91 | 91 | 91 |
| 2 | 93 | 93 | 93 | 93 |
| 3 | 91 | 91 | 91 | 91 |
| 4 | 91 | 91 | 91 | 91 |
| 5 | 93 | 93 | 93 | 93 |
| 6 | 91 | 91 | 91 | 91 |
| 7 | 99 | 99 | 99 | 99 |
| 8 | 93 | 93 | 93 | 93 |
| 9 | 91 | 91 | 91 | 91 |
| 10 | 91 | 91 | 91 | 91 |
| 11 | 91 | 91 | 91 | 91 |

## Questionnaire Specifications - 700 pages

EDUCATION

***EDUCATION***

NOTE: EDUSCHLGO takes into account information from follow-up probes if respondents originally reported that they were not enrolled in school. Based on these probes, respondents were coded as yes (i.e., enrolled) if they were on break from school and intended to return to school once their break was over. EDUSCHLGO was assigned a code of 11 if the respondent reported currently being enrolled but there was uncertainty about the respondent's school enrollment status.

The next questions are about school. By 'school', we mean elementary school, junior high or middle school, high school, or a college or university. Please include home schooling as well. Do you go to school?

If you are on a holiday or break from school, such as spring break or summer vacation, but plan to return when the break is over, please answer yes.

*(QD17, QD17A, QD17B)*

EDUSCHLGO Len : 2 NOW GOING TO SCHOOL

| | Freq | Pct |
|---|---|---|
| 1 = Yes | 17131 | 30.21 |
| 2 = No | 37569 | 66.25 |
| 11 = Yes (EDUSKPMON = 30) | 45 | 0.08 |
| 85 = BAD DATA Logically assigned | 93 | 0.16 |
| 94 = DON'T KNOW | 346 | 0.61 |
| 97 = REFUSED | 413 | 0.73 |
| 98 = BLANK (NO ANSWER) | 1108 | 1.95 |

# Technical Plan

Two approaches:

1. **Logistic Regression:** if response variable clearly defined.
    a. Clean data
    b. Compute coefficients of each feature and associated p-values
    c. Determine significant features
2. **Clustering and Logistic Regression:** if no response variable
    a. Clean data
    b. Possible techniques: K-Means, GMM, DBSCAN
    c. Determine clusters
    d. Fit logistic regression model on data with clusters as response variable

Evaluation of model performance: Cross-validation

# Potential Impact

- Helps rehabilitation programs recognize high-risk individuals.
- Provides insights for improving treatment plans.
- Supports long-term rehabilitation strategies and monitoring efforts.

# References

*National Survey on Drug Use and health*. SAMHSA.gov. (n.d.). https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health

U.S. Department of Health and Human Services. (2025, February 7). *NIDA IC Fact sheet 2024*. National Institutes of Health. https://nida.nih.gov/about-nida/legislative-activities/budget-information/fiscal-year-2024-budget-information-congressional-justification-national-institute-drug-abuse/ic-fact-sheet-2024

# Discussion

- Beyond what we mentioned, what machine learning techniques or statistical methods would be best suited for this type of prediction?
- What additional data sources could enhance our project?
- What potential biases could exist in our model, and how can we mitigate them?
- What ethical considerations should be taken into account when using data to predict relapse?
- Any other questions?

Thank you!