A Psychosocial Approach to Understanding Substance Use Disorder (SUD) Among Adolescents

Adway S. Wadekar
Saint John's High School, Shrewsbury, Massachusetts, USA

Introduction and Motivation

- Substance use leads to a wide variety of problems for adolescents.
  - Poor peer relationships
  - Lowered performance in school
  - Motor vehicle accidents.
- Adolescent use of illicit drugs and alcohol remains high (National Survey on Drug Use and Health, 2018)
  - ~ 17% adolescents used an illicit drug in the past year
  - ~ 21% adolescents used alcohol in the past year
  - ~ 13% adolescents used marijuana in the past year
- The most severe form of substance use is **substance use disorder**.
  - Substance use disorder is defined as abuse or dependence on illicit drugs or alcohol (Diagnostic and Statistical Manual, Fourth Edition)

Introduction and Motivation (cont.)

- Prior research suggests that several risk & protective factors may be involved in the development of SUD among adolescents.
  - Peer pressure
  - Seeking peer approval
  - Perceived risk of harm
  - Availability of drug education programs
  - Presence of impulse control or risk-taking behaviors.
- Many studies consider only a cross section of adolescents' lives when attempting to understand adolescent SUD.
  - Parenting style and drug abuse (Shakya et. al)
  - Obesity and drug misuse (Denoth et. al; Sansone et. al)
  - Community attributes and substance abuse (Cleveland et. al)

Introduction and Motivation (cont.)

- No prior study has completely and cohesively considered the entirety of an adolescent's **sphere of life** in a single model.
  - We must consider many germane factors together.
  - Otherwise, it is not possible to understand how the factors influence the adolescent in developing SUD.
  - Also, by considering factors together, we can understand how they influence **each other** when contributing to risk and protection of adolescents developing SUD.
  - We can move beyond explanation and towards prediction through and integrated consideration.
- To address this, we develop a comprehensive psychosocial approach to understanding substance use disorder among adolescents.
  - We consider a combination of intrinsic and extrinsic traits
  - We use national-level survey data
  - The approach relies on machine learning techniques.

A Psychosocial Approach

- Building the data-driven model:
  - Whether or not an adolescent developed an SUD is viewed as an outcome.
    - This outcome is made more or less probable by risk and protective factors, respectively.
  - We built a data-driven model of an adolescent's sphere of life by considering a variety of factors.
- Data Source – National Survey on Drug Use and Health (NSDUH)
  - Conducted by the Substance Abuse and Mental Health Services Administration, part of the US Department of Health and Human Services.
  - Annual survey of the civilian, noninstitutionalized, United States population aged 12 years or older.
  - Large and nationally representative of the US population
- Model was built using over 100 questions relating to youth experiences, sociodemographic factors, physical factors, and psychological factors.

A Psychosocial Approach (cont.)

- For better classification and interpretation, these 100 questions were compressed.
  - Simple factors and compound factors were built.
- Simple factors:
  - Simple factors are a standard, one-to-one mapping from the factor itself to a question in the survey.
  - Examples of simple factors include (question → factor):
    - "What is your race?" → Race
    - "What is your gender?" → Gender
- Compound factors:
  - Compound factors are a many-to-one mapping from the factor itself to **multiple** questions in the survey.
  - Examples of compound factors include:
    - Religiosity
    - Peer approval
    - Support system

A Psychosocial Approach (cont.)

- An example of a compound factor: religiosity
  - Four questions were used to gauge an adolescent's religious beliefs or religiosity.
    - Number of religious services attended in the past 12 months. (25 or more = 1, < 25 = 2)
    - Religious beliefs are very important to me (Agree/Strongly Agree = 1, Disagree/Strongly Disagree = 2)
    - Religious beliefs influence my decisions (Agree/Strongly Agree = 1, Disagree/Strongly Disagree = 2)
    - Important that my friends share my beliefs (Agree/Strongly Agree = 1, Disagree/Strongly Disagree = 2)
  - These coded responses, once added together, range from 4 to 8 which is the "spectrum of religiosity."
    - As the number increases from 4 to 8, the adolescent becomes milder about religion.
    - Note: this system assumes each instantiation of religiosity to be of equal importance
- 34 risk and protective factors were created from these 100+ questions
  - 12 simple factors and 22 complex factors.

A Psychosocial Approach (cont.)

- Organization of the factors
  - In the literature, there is no consensus about the organization of factors.
  - One study proposes three group classification: familial, social, individual
  - Another study proposes four group classification: individual, family, peer, and school.
- We used a two-group classification scheme
  - Proximal factors
    - *Proximal* factors are factors that are immutable or factors that can be controlled by the adolescent themselves.
    - Examples include race and gender; obesity, religious beliefs.
  - Distal factors
    - *Distal* factors are ones which are primarily controlled by where and how the adolescent is nurtured.
    - Examples include economic status, family composition, attitudes of parents towards substance use.

A Psychosocial Approach (cont.)

- Some factors could fit into both the proximal and distal group
  - Religiosity, for example could be a product of both personal beliefs but also the beliefs of the family which are impressed upon the adolescent.
  - However, each of the 34 factors was classified according to whether it **predominantly** reflects an inherent trait of an adolescent or a trait which is modulated by factors that the adolescent can't control (e.g., environment, family)
- Distribution of the factors across proximal vs. distal groups and simple vs. compound type

	Simple	Compound
Proximal	4 (26.67%)	11 (73.33%)
Distal	8 (42.11%)	11 (57.89%)

A Psychosocial Approach (cont.)

TABLE I
PROXIMAL FACTORS & MAPPING TO NSDUH VARIABLES

ID.	Feature	Mapping/Description	# Var.
1.	Risk of Harm – Daily Use	Perceives people risk harming by binge drinking and smoking.	2
2.	Risk of Harm – Weekly	Whether an adolescent perceives that people risk harming themselves by binge drinking, or using cocaine, marijuana, LSD, Heroin.	2
3.	Risk of Harm – Monthly	Whether an adolescent perceives that people risk harming themselves by using cocaine, marijuana.	2
4.	Risk of Harm – Lifetime	Whether an adolescent perceives that people risk harming themselves by trying LSD and Heroin.	2
5.	Mental Illness, Depression	Feeling sad, empty, depressed, discouraged, or hopeless, lost interest and became bored with most things usually enjoyable.	9
6.	Special Drugs	At least once use of cough/cold meds, GSB.	2
7.	Health	Self-assessed overall health	1
8.	Obesity	Based on BMI (Obese, Overweight, Normal)	1
9.	Race	Seven levels (Non Hispanic White, Non Hispanic Black/African American, Non Hispanic Native American/Alaska Native, Non Hispanic Native Hawaiian/Pacific Islander, Non Hispanic Asian, Non Hispanic more than one race, Hispanic).	1
10.	Gender	Male/female	1
11.	Gang Affiliation, Violence	Whether an adolescent engaged in a serious fight at school/work, or fought with a group versus other group, carried a handgun, sold illegal drugs, stole/tried to steal item greater than $50, Attacked with intent to seriously harm anyone.	6
12.	Peer Approval	What an adolescent thinks close friends feel about them smoking 1 pack of cigarettes per day, trying marijuana, use marijuana monthly, drinking 1-2 alcoholic beverages/day.	4
13.	Religious Beliefs	Number of times attended religious services, whether religious beliefs are important, whether religious beliefs influence life decisions, whether it is important for friends to share the religious beliefs.	4
14.	Risk taking behavior	Whether the adolescent gets a real kick out of doing dangerous things, whether the adolescent likes to test limits by doing risky things, whether an adolescent wears a seatbelt when ride in front passenger seat.	3
15.	Youth Approval	What youth feels about peers trying marijuana, using it monthly, smoking more than 1 pack of cigarettes per day, or drinking 1-2 alcoholic drinks/day.	4

A Psychosocial Approach (cont.)

TABLE II
DISTAL FACTORS & MAPPING TO NSDUH VARIABLES

ID	Feature	Mapping/Description	# Var
1.	School, academics	Felt about going to school, whether the teacher let them know that they were doing a good job, grade average.	3
2.	Peer Drug Use	Associated with peers in grade who use cigarettes, marijuana, alcohol, or got drunk at least once/week	4
3.	Easy availability	Easy to obtain cocaine, crack, heroin, LSD, marijuana?	4
4.	Approach	Approached by someone selling drugs	1
5.	Parenting style involvement	Whether the parents checked and helped with homework, made them do chores, limit amount of TV and time out on school night, told them that they are proud, and talked about dangers of drugs, alcohol, and tobacco. Number of times argued and fought with parent.	8
6.	Parental attitudes towards substances	What the adolescent thinks parents feel about smoking one pack of cigarettes/day, trying marijuana, using marijuana monthly, drinking 1-2 alcoholic beverages/day.	4
7.	Mother	Presence of mother in the household	1
8.	Father	Presence of father in the household.	1
9.	Poverty	Poverty level of the family (poverty, income two or more times that of federal poverty threshold).	1
10.	County	Living in large metro, small metro, or a non-metro area.	1
11.	Insurance	Covered by private insurance, Medicare, Medicaid, Champus, VA or military, or any other health insurance.	5
12.	Family size	Number of people in household (1 through 6, or more).	1
13.	Govt Programs	Family participates in any government programs such as the Supplemental Social Security Income, Food Stamps, Cash/Non Cash Assistance.	4
14.	Youth Activities	Number of youth activities participated in school, community, church/faith-based, or other activities.	4
15.	Drug Education	Participated in any drug education in school either through a special class, or through films/lectures/discussions in or out of class such as a special assembly.	3
16.	Drug Prevention Message	Has heard any drug prevention message outside of school – through posters, pamphlets, billboards or Radio and TV.	1
17.	Drug/Self-help Programs	Participated in any groups that promote problem solving, communication skills, and self-esteem, or programs that prevent violence and drug use, or help substance abuse, counsel against pregnancy and STDs	5
18.	Support System	Who the adolescent talks to about serious problems parent, guardian, boyfriend/girlfriend, other adults, or no one.	5
19.	State Law about Marijuana	Whether the state in which the adolescent resides has legally approved marijuana.	1

A Psychosocial Approach (cont.)

- Defining the outcome variable, Substance Use Disorder
  - A broad, two-dimensional view of SUD is taken
  - We consider a range of illicit substances including:
    - Marijuana, hallucinogens, inhalants, methamphetamine, cocaine, heroin, pre-scription pain relievers, prescription sedatives, prescription stimulants, prescription tranquilizers, and alcohol.
  - We also consider both abuse/dependence (DSM-IV)
    - **Abuse**: pattern of substance use leading to neglect of roles or commitments, physical hazards, legal issues, or interpersonal problems.
    - **Dependence:** use of the substance is reoccurring to the point of reducing important social and occupational activities, causing tolerance of the substance and/or withdrawal symptoms and causing the user to devote a great deal of time to obtain and use the substance.
    - If the criteria for dependence is satisfied, then the criteria for abuse is satisfied.
- Either abuse or dependence to any substance in past year is labeled as SUD.

A Psychosocial Approach (cont.)

- Building a labeled data set
  - Based off the 34 risk and protective factors, we curated the labeled data set.
    - Outcome variable SUD
    - 34 factors split into *Proximal* and *Distal* groups.
- National Survey on Drug Use and Health
  - ~56,000 responses per survey edition
    - Of which, ~13,000 are from adolescents (aged 12-17)
  - Initially, we explored just the 2016 edition of the NSDUH
    - Many variables have missing values and eliminating the observations brought 13,000 to 9,128.
    - To mitigate this, we pooled data from the 2017 edition of the survey.
    - As shown before, the NSDUH design is quite stable and consistent.
      - Therefore, pooling is accepted practice.
    - Final data included 18,624 adolescents

A Psychosocial Approach (cont.)

- Selecting classification models
  - Use machine learning
    - Machine learning models can distinguish and handle complex interactions among many factors.
    - Barenholtz et. al. discuss the promise of machine learning in substance use research.
- Many types of models were considered
  - Support Vector Machines (SVM) not suitable because most features are categorical.
  - Naïve Bayes works well when all factors contribute **independently** to prediction
    - This is not the case – many factors are dependent upon each other.
  - Artificial Neural Networks are appropriate for unstructured, heavy data.
    - ANNs can be too robust and unnecessary.
- Tree-based classifiers
  - Because the data is lightweight and categorical, tree-based classifiers seem to be the natural choice.
  - Initially, decision trees were considered
    - However, it is well-known that decision trees are prone to overfitting.

A Psychosocial Approach (cont.)

- Ensemble classifiers were considered to avoid overfitting
  - Models that build many trees
  - Use results from those trees to aggregate into one prediction (SUD/No-SUD)
- Random Forest
  - Random Forest algorithm builds trees in unison/parallel.
  - Data point goes through each tree.
  - Final prediction is an aggregate of the individual tree prediction
    - Majority rules approach is followed.
- Gradient Boosting
  - In gradient boosting, trees are built sequentially to improve upon misclassification of previous trees.
  - Predictions aggregated along the way.
- RF vs. GB
  - RF are known for multi-class detection and bioinformatics – statistical noise.
  - GB are known for performing well with unbalanced data

A Psychosocial Approach (cont.)

- Class imbalance problem
  - Only 5% of adolescents have the problem of SUD
  - Critical to find, yet still, a “needle in a haystack.”
- Standard processing procedure
  - Split data into training and test sets.
    - Training set – 60%-80% of data
    - Test set – 20%-40% of data
    - Standard procedure is not applicable due to class imbalance problem
- Techniques to address class imbalance
  - Downsampling
  - Upsampling
- Downsampling vs. Upsampling
  - Downsampling randomly removes instances in majority class (No-SUD)
  - Upsampling randomly replicates instances in minority class (SUD)

A Psychosocial Approach (cont.)

- Benefits and drawbacks to each class-balancing method
  - Downsampling – massive loss of valuable data
  - Upsampling – simple resampling
- Synthetic Minority Oversampling Technique (SMOTE)
  - SMOTE combines both downsampling and upsampling
  - Downsamples a random number from majority class
  - Synthetically upsamples from minority class (using k-Nearest-Neighbors algorithm)
- Implementation of SMOTE
  - Partition 1 – 80%
  - Partition 2 – 20%
  - Use stratified sampling to split – preserves ratio initially
  - Apply SMOTE algorithm to Partition 1 – training set
  - Leave Partition 2 as is for testing
- By keeping Partition 2 imbalanced, we model a real-life scenario

A Psychosocial Approach (cont.)

- After applying SMOTE to data, we ran the three classifiers (repeating entire process 50 times)
  - Decision Tree, Random Forest, and Gradient Boosting
  - Also ran a multivariate logistic regression model as used commonly in medical studies
  - Main drawback of multivariate LR model is that feature importance is not possible to gauge.
- Performance metrics

		Ground Truth	
		SUD	No-SUD
Predicted	SUD	True Positive (TP)	False Positive (FP)
	No-SUD	False Negative (FN)	True Negative (TN)

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

- Area Under ROC Curve (AUC) used to estimate tradeoff between sensitivity and specificity at different decision thresholds.
  - AUC between 0.7-0.8 is fair, 0.8-0.9 is good, 0.9+ is considered excellent.

A Psychosocial Approach (cont.)

- We used the random forest model to find factor importance and importance of interactions
- Assessing factor importance
  - Compute relative importance of factors
  - Tree-based classifiers segregate observations by relying on the Gini impurity
    - Classifier will try all factors and separate first by the one that has the highest Gini impurity
  - The larger the Gini impurity, the higher the segregation power of a given factor
  - We calculated average Gini impurity for each factor across all trees in 50 forests.
- Assessing importance of interactions
  - Importance is determined by how many times two given factors appear adjacent to each other in a given forest.
  - These interactions between two factors were also averaged across the 50 forests.
- Factor importance and importance of interactions were normalized

Results and Discussion

- Prediction Performance:

TABLE III
PREDICTION PERFORMANCE

Model	Sensitivity	Specificity	AUC
DT	0.79 (± 0.013)	0.79 (± 0.008)	0.84 (± 0.006)
RF	0.81 (± 0.010)	0.83 (± 0.003)	0.90 (± 0.006)
GB	0.83 (± 0.003)	0.82 (± 0.001)	0.91 (± 0.002)
LR	0.82 (± 0.003)	0.82 (± 0.011)	0.88 (± 0.003)

- Ensemble classifiers perform exceptionally well
  - Both have excellent AUC; above 0.90.
  - Prediction stability across classifiers is excellent
    - Demonstrated through very narrow 95% confidence intervals.
  - No model decisively outperforms the other
    - GB has better sensitivity and AUC; RF has better specificity

Results and Discussion (cont.)

- In terms of prediction performance, our approach outperforms other prevalent approaches
  - Vazquez et. al use a sample of Mexican 5th and 6th graders to predict **use** of alcohol, marijuana, etc.
    - AUC reaches average of only 0.823.
  - Han et. al. predict opioid misuse among US adolescents
    - AUC reaches average of 0.811.
  - Considering use/misuse is an easier problem because cases of use/misuse are more prevalent than SUD
    - 20% vs. 5%
- We note two distinctions between our approach and the other two approaches
  - We believe our approach is more comprehensive, considering for more factors in an adolescent's life, including their social network.
  - We also specifically make a concerted attempt to mitigate the class imbalance problem using the SMOTE procedures.

Results and Discussion (cont.)

- Importance of Factors
  - RF classifier performs well so it's appropriate to draw inferences about the importance of factors
- There has been significant debate about the relative importance of factors
  - Cleveland et. al. find that risk factors are stronger predictors than protective factors
  - Prior research shows that peer and individual factors have strong association with illicit drug use.
  - Other studies show that school and peer factors are strongest predictors of adolescent drug use.
- The top factors are split fairly evenly across Proximal and Distal groups
  - We also find that risk factors are more dominant than protective factors
  - We find that peer and individual factors are more important than family, parenting factors, and school environment
  - The relative importance of the factors is also very close.
    - Of the factors that rank in the top 10, obesity and risk-taking personalities rank the highest.
  - This indicates that no one factor is dispositive in predicting SUD.
  - It is appropriate to consider interactions among pairs of factors

TABLE IV
NORMALIZED RELATIVE IMPORTANCE SCORES OF THE FACTORS

Factor	Score	P/D
Seeking approval of peers for the use of substances	1.00	P
Easy availability of substances	0.98	D
Adolescent approving peer use of substances	0.93	P
Approached by someone selling substances	0.87	D
Peers using substances	0.85	P
Adolescent engaging in violent behavior	0.75	P
Risk taking personality of the adolescent	0.69	P
Obesity	0.57	P
Permissive attitudes of the parents	0.55	D
Parenting style and involvement	0.47	P
Perception of risk – weekly use of substances	0.35	P
Race	0.32	P
School environment of the adolescent	0.31	D
Depression and mental illness	0.30	P
Religious beliefs	0.29	P
Size of the family	0.28	P
Overall health of the adolescent	0.27	P
Participation in youth programs, Perception of risk – daily use of substances	0.24	D, P
Use of special drugs	0.23	P
Perception of risk – lifetime use of substances, Metro/Non-Metro residence	0.22	P, D, D
Perception of risk – monthly use of substances	0.19	P
Poverty level of the family	0.18	D
State law whether marijuana is legal, Presence of mother in the household	0.11	D
Participation in drug education programs, Participation of the family in government programs, Seeing drug prevention message outside of school.	0.10	D
Gender	0.09	P
Availability of a support system, Presence of father in the household	0.08	D, D
Presence of father in the household	0.08	D
Participation in youth activities	0.07	D
Family covered by government or private medical insurance	0.03	D

Results and Discussion (cont.)

- Interactions between pairs of factors
  - Each factor can interact with any other factor including itself.
  - This leads to over 1,000 interactions
- We ranked the top 25 interactions
  - Being approached with drugs and obesity
  - Easy availability of drugs and obesity
  - Adolescent's own approval of use and obesity
  - Being approached with drugs and risk-taking behavior
- A common theme emerges
  - Obesity is thought to be a manifestation of risk-taking/impulsive personalities (Sansone et. al.)
  - Risk-taking attitudes combined with easy availability of substances poses the greatest risk for an adolescent to develop SUD

TABLE V
NORMALIZED RELATIVE IMPORTANCE SCORES OF THE FACTORS

Factor A	Factor B	Rel. Score
Approached by some selling substances	Obesity	1.00
Easy availability of substances	Obesity	0.95
Seeking approval for use of substances	Obesity	0.94
Approached by someone selling substances	Risk-taking attitude	0.94
Approached by someone selling substances	Parenting style and involvement	0.94
Engaging in violent behavior	Obesity	0.93
Approval for the use of substances by peers	Obesity	0.92
Approached by someone selling substances	Family size	0.90
Easy availability of substances	Risk-taking attitude	0.90
Approached by someone selling substances	Religious beliefs	0.89
Peer approval of the use of substances	Parenting style	0.89
Easy availability of substances	Parenting style	0.89
Peer approval of the use of substances	Risk-taking attitudes	0.88
Approached by someone selling substances	Risk of harm by weekly use	0.88
Permissive parental attitudes	Obesity	0.87
Approached by someone selling substances	Peer use of drugs	0.87
Engaging in violent behavior	Parenting style	0.87
Approached by someone selling substances	Race	0.87
Engaging in violent behavior	Risk-taking attitude	0.87
Approval of use of substances by peers	Risk-taking attitude	0.87

Results and Discussion (cont.)

- Intervention strategies
  - Intervention strategies must be developed to mitigate adolescent substance use
  - Prior research shows that adolescent substance use presents a risk for use of substances, potentially harsher ones, in adult life.
- Interactions between factors should be considered when developing strategies
  - Heavy premium should be placed on limiting access to substances
  - This could be challenging in the United States
    - Access to marijuana is more prevalent than ever given the changing landscape on the outlook towards legalization of marijuana
  - Prevention and education programs should raise awareness not just of the ill-effects of substance use
    - They should inform and counsel adolescents on controlling impulsive and risk-taking tendencies

Conclusions and Future Work

- We present an approach to explore the relationship between various factors contributing to SUD
- We develop models to predict at-risk adolescents with exceptional accuracy
- Neither proximal nor distal factors are dominant
  - Development of SUD among adolescents has a complex etiology consisting of both proximal and distal factors
- We find that environmental factors combined with innate impulsive/risk-taking tendencies pose the greatest threat for adolescent development of SUD
- Broadly, we demonstrate that public domain data can improve our understanding of SUD among various cohorts
  - Because certain key factors are missing in the NSDUH data, integrating alternative sources of data is a topic for future study