Paper ###-2020

Analyzing the Factors Impacting Suicidal Behavior in American Youth

Ashlesha Sampathi, Mounica Mandapati, Nitesh Maruthi, Venkat Ram Reddy

Oklahoma State University

ABSTRACT

Suicide is the 2nd leading cause of death in youth aged 12-19. Estimates from the data indicate that nearly 7.4% of the American youth attempted suicide in 2017 alone, while 17.2% of teens considered attempting suicide. The last decade has seen an increase in the number of suicide attempts in teenagers. Given the seriousness of this issue, it is important to identify the factors that are leading to this increase in suicide attempts. This paper closely examines the 2017 Youth Risk Behavior Survey (YRBS), in order to understand characteristics of teenagers who attempted suicide. Latent Class Analysis using SAS® Enterprise Guide® found that among teens who attempted suicide, three distinct classes were evident which were characterized mainly by sexual assault, bullying, and depression respectively. Moreover, using PROC SURVEYLOGISTIC, it was seen that the odds of attempting suicide were four times higher for teens who were sexually assaulted and three times higher for teens who were bullied or abused drugs. Furthermore, it was seen that teenagers who were sexually assaulted had a high co-occurrence of other risky behaviors, and ultimately a higher percentage of multiple suicide attempts. The outcome of the paper highlights the importance of early intervention in preventing teenagers from slipping down a "rabbit hole" of risky behaviors that ultimately lead them to take their own lives.

INTRODUCTION

"People who attempt suicide aren't always convinced it's the only option. It's more often that they have exhausted their emotional reserves to continue pursuing those options. It is, in many ways, the ultimate state of burnout. In order to attempt suicide, a person must be in the neurological state where they can override their own survival instincts."[1]

Suicide is a major public health concern not just in the United States, but all over the world. It is a complex issue that involves social, individual, and behavioral factors over time that drive an individual to take the ultimate step. In order to solve any problem, it is imperative to understand the causes. Although specific causes influencing the decision to commit suicide may vary greatly among teenagers who attempt it, the factors could be generalized into certain broad categories. Looking for patterns of co-occurrence and quantifying the impact of the individual factors can inform us on how to address the issue.

In the U.S, no complete count of suicide attempt data is available. The CDC gathers data from hospitals on non-fatal injuries from self-harm as well as survey data. The Youth Risk Behavior Survey (YRBS) was used by the team for the analysis of suicide attempts. The team found a general rise in the percentage of teens who claimed to have suffered depression, suicidal thoughts, and made plans to commit suicide. A trend analysis showed a significant increase over the last decade in these thoughts/behaviors. However, the data did not show a significant increase in the percentage of teenagers attempting suicide as shown in Table 4. It should be acknowledged that the actual number of teens who have attempted suicide could be higher, due to the stigma preventing an honest response on the survey

DATA

Data used in this analysis was provided by the Center for Disease Control (CDC) and Prevention and gathered using the Youth Risk Behavior Survey (YRBS) [2]. The YRBS essentially was developed to monitor health behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and adults in the United States. The YRBS includes national, state, territorial, tribal government, and local school-based surveys of representative samples of 9th through 12th grade students. The team used the consolidated national dataset which contains data from YRBS surveys conducted from 1991-2017 biennially nationwide for the United States.

For the 2017 YRBS survey, a total of 192 schools were sampled, of which, 144 schools participated totaling to 18,324 students surveyed. 14,956 of the 18,324 sampled students submitted questionnaires, which composed of 99 questions; 14,765 questionnaires were usable after data editing. The final dataset consists of 14,765 rows and 242 variables, which included original survey variables, calculated variables, and dichotomized variables.

PROBLEM STATEMENT

This paper attempts to examine co-occurrences of behavioral characteristics among teenagers who attempted suicide and to quantify their impact on suicide attempts using the latest YRBS survey data.

DATA PREPARATION

One of the questions in the survey asked the respondents about the number of times they had attempted suicide in the past. A new dichotomized variable was created from the responses to this question, that indicated if a teenager had attempted suicide or not and this variable was used as the target variable for analysis. Students who attempted suicide one or more times during the 12 months before the survey were treated under one categorical level named 'Attempted suicide' and all the others under the level 'Did not attempt suicide'.

Three other new dichotomized features were created by capturing information from multiple variables into a single variable for the ease of analysis. One such feature was 'Abuse of Narcotics' which indicates the usage of non-recreational drugs. The second feature was called 'Bullied' which indicates if the teenager was bullied at school and/or online. The third feature created was 'sexual minorities' which indicates if a person is LGBTQ+.

Since it would be impractical to use all the variables in the dataset for analysis, only a few variables related to the problem statement were selected using previous literature [3].

DATA CLEANING

There are missing values in the data due to non-responses. The missing values are imputed with SURVEYIMPUTE procedure, which uses hot-deck imputation. In this imputation, the observed data was partitioned into imputation cells such that observations in the same cell are similar in some sense. The imputation is performed independently within each cell. Observations that contain no missing values are used as donors. For every recipient, one or more donors are selected randomly from the same imputation cell. The observed values of the donors are used as the imputed values for the missing items of the recipient. Since the observation weights are unequal, a weighted selection of donors is used instead of an unweighted selection.

EXPLORATORY DATA ANALYSIS

Trend graphs were plotted for suicidal ideations variables in the data. A rise in the trend of these variables was seen from 2007-2017. A linear trend analysis using logistic regression model controlling for sex, race, grade was built to test if the rise in trend was significant. The variable of interest was taken as a target variable, and the time variable was treated as continuous and was created by coding each year with orthogonal coefficients calculated using PROC IML in SAS®. According to the results of this model, it was found that this increase is significant for the past decade.
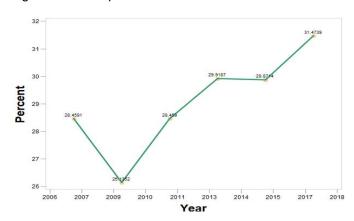
Figure 1 % of depressed teens

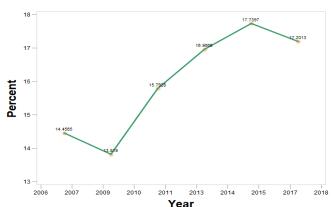Figure 2 % of teens who seriously considered suicide

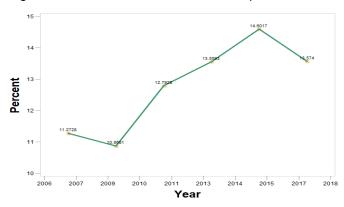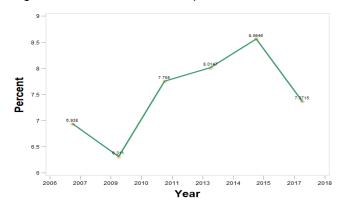

Figure 3 % of teens who made a suicide plan

Figure 4 % of teens who attempted suicide

Upon looking at the demographics of teenagers who committed suicide in 2017, it was found that more females attempted suicides than males. More teens in 9th and 10th grade attempted suicides than those in 11th and 12th grade. Among all the races surveyed, more African Americans attempted suicide than any other race. The difference between these groups was found to be significant according to a t-test analysis. This indicates that females, younger teens, and African Americans are more vulnerable to attempt suicide.

ANALYSIS

LATENT CLASS ANALYSIS

Latent class analysis was done to understand the similar characteristics among teens who attempted suicide. Surveys of only those teens who attempted suicide at least once were used for this analysis. A three latent class model was built using PROC LCA on selected

variables which included depression, sexual assault, bullied, and involvement in physical fights. With an aim to explain the behaviors, the model with lower AIC was retained.

It was seen that there was a high probability of depression in all three classes. Moreover, two classes were characterized by a high probability of teens who had been sexually assaulted and bullied respectively. No other distinguishing characteristics were found in the depressed class which indicates unknown causes of depression which may require a more complex analysis or data. In all, 20% of the teens belonged to the sexually assaulted class, and 40% each belonged to both the bullied and depressed classes.

Furthermore, on profiling, it was seen that the sexually assaulted class had the highest percentage of teens attempting suicide multiple times - more than four times on average.

LOGISTIC REGRESSION MODEL

From the literature review, 18 variables were selected to understand the relationship with suicidal risk behavior. To understand the causal relationships between the variables, the Bayesian network model was built in SAS® Enterprise Miner™ using High-Performance Bayesian Network Classifier Node. In this model, the Markov Blanket Bayesian network was selected as it helps display significant relationships not only between independent and target variables, but also between independent variables themselves. The significant relationships are shown in the appendix. Seven individual independent variables and four interaction terms were found to be significant among all the 18 independent variables, which were fed into the algorithm. Seven independent variables included drug abuse, sexual minorities, bullied, sexual assault, depression, involvement in physical fights, and perception of obesity. Four interactions were seen between variables: depression and sexual assault, depression and perception of obesity, depression and bullied, drug abuse and sexual assault.

To quantify the cause-effect relationship of the seven independent variables and the four interaction terms with the target, a logistic regression model was built using survey logistic procedure in order to account for the sampling design. From initial analysis, including all the seven independent and four interactions, it was seen that the most important variable was depression and it was found that a depressed teen is eight times more likely to attempt suicide than a non-depressed teen. Since depression is only a symptom and not a root cause of suicidal ideation, the depression variable was excluded from the analysis [4]. A logistic regression with only six variables and four interaction terms was built. The final odds ratio estimates are shown in Table **5** Odds Ratio Estimates in the appendix.

The most important variable was seen to be sexual assault, and it was found from our analysis that teens who were sexually assaulted were 4.4 times more likely to attempt suicide than those who were not. The other important variables were bullying, drug consumption, sexual minority, and perception of obesity. Odds of suicide attempts in teens who were bullied at school was 3.2 times more than non-bullied teens. Teens who abuse non-recreational drugs were found to be 2.8 times more likely to attempt suicide than those who did not. Teens who belonged to a sexual minority class (LGBTQ+) are 2.6 times more likely to attempt suicide than teens who did not. Teens who got involved in physical fights were 2.1 times more likely to attempt suicide than those who did not. Teens who felt that they were obese were 1.4 times more likely to attempt suicide than those who did not.

GENERALIZATION

Similar methodology and analysis can be extended to other age groups and countries, which could yield interesting results. Although factors influencing suicide in adults would be markedly different such as health issues, financial and relationship stability, mental illness etc., the impact of these on suicide attempts can be quantified using the methods used in this paper.

FUTURE SCOPE

Since YRBS data is cross sectional, and includes data across different sections of society, it can be expected that the impact of individual factors would vary based on these sections. For example, in the United States alone, LGBTQ+ rights vary across states, with the west being more tolerant than the south. Quantifying relative social acceptance of LGBTQ+ teens and minorities in general would provide an estimate of its impact on suicidal attempts.

Additionally, socio-economic data can be merged with the YRBS data at a more granular level, such as a county/city to examine the impact of economic wellbeing on teenage suicidal behavior.

Furthermore, surveys in the future could include questions about specific sources of social media consumption, which could show common patterns among teens who attempted suicide. For example, in early 2016, there were widespread rumors of a dark game called the "Blue whale challenge" that targeted vulnerable youth to attempt suicide that was primarily spread through a Russian social networking website called VK.com.

CONCLUSIONS

The results of the study found a high degree of co-occurrence of risky behavior in teens that suggest a "Rabbit hole" situation where at-risk teens go down a negative spiral of getting into fights, depression, and abuse drugs among others. This highlights a need for early intervention.

Creating an atmosphere where teens voluntarily seek help takes time to achieve. We propose the creation of a data driven application that uses Radio Frequency Identification (RFID) and Internet of Things (IoT) technology to track students at schools, with an aim to identify teens that fall out of normal behavior patterns such as falling grades, missing classes, and skipping lunch among others. This will enable identification of high-risk teens for early intervention. Teens who are depressed show these signs and the models showed that the odds of them attempting suicide were eight times more than teens who were not depressed. Their situation may have worsened by the time it is apparent to school authorities, hence early intervention is critical.

The US government has several programs that are aimed at reducing the suicide rate among teenagers, although the effectiveness of such programs has been questioned [5]. Measures that are more tailored for young adults, such as having contemporary youth icons like musicians and TV actors to spread awareness and positive messages in lieu of mass-market campaigns could prove to be more effective.

Teens who were sexually assaulted were seen to be at the highest risk of going into the downward spiral. Thus, we recommend treating these survivors with the utmost care. Creating a safe space at schools where they can seek help without fear of social stigma or repercussion in conjunction with the aforementioned application can prevent these teens from going down the rabbit hole by means of early intervention and care.

Most anti-bullying initiatives look at the victim of bullying but fail to address the cause - the bullies. Several studies have shown that bullies themselves suffer from issues due to domestic violence, behavioral issues etc. Treating bullies with therapy instead of merely handing out punishments could be effective in reducing their aggression.

REFERENCES

[1] Finch, S. (2019, March 8). 5 Things Suicide Loss Survivors Should Know — from Someone Who's Attempted. Retrieved from Healthline:
https://www.healthline.com/health/mental-health/losing-someone-to-suicide#4

[2] CDC – Adolescent and School Health - Youth Risk Behavior Surveillance System (YRBSS) Overview. https://www.cdc.gov/healthyyouth/data/yrbs/overview.htm . December 5, 2019.

[3] Kuehn, K. S., Wagner, A., & Velloza, J. (2019). Estimating the magnitude of the relation between bullying, e-bullying, and suicidal behaviors among United States youth, 2015. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 40(3), 157–165. https://doi.org/10.1027/0227-5910/a000544

[4] U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health. (2018). Suicide in America: Frequently Asked Questions (NIH Publication No. TR 18-6389). https://infocenter.nimh.nih.gov/pubstatic/TR%2018-6389c/TR%2018-6389.pdf

[5] Jaffe, DJ. "Preventing suicide in all the wrong ways". USC - CENTER FOR HEALTH JOURNALISM MEMBER POSTS. September 09, 2014. https://www.centerforhealthjournalism.org/2014/09/09/preventing-suicide-all-wrong-ways .

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Ashlesha Sampathi

Email: Ashlesha.sampathi@okstate.edu

Mounica Mandapati

Email: Mounica.mandapati@okstate.edu

Nitesh Maruthi

Email: Nitesh.Maruthi@okstate.edu

Venkat Ram Reddy

Email: vreddy@okstate.edu

APPENDIX

1. Trend Analysis Significance test:

Analysis of Maximum Likelihood Estimates

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		-0.8829	0.0170	-51.88	<.0001
sex	1	0.4209	0.0134	31.30	<.0001
race4	1	-0.1402	0.0227	-6.17	<.0001
race4	2	-0.1354	0.0248	-5.45	<.0001
race4	3	0.1996	0.0210	9.52	<.0001
grade	1	-0.0302	0.0174	-1.73	0.0844
grade	2	0.00919	0.0179	0.51	0.6092
grade	3	0.0369	0.0185	1.99	0.0475
T3L_L		0.1392	0.0445	3.13	0.0020

NOTE: The degrees of freedom for the t tests is 257.

Analysis of Maximum Likelihood Estimates

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		-1.6802	0.0171	-98.28	<.0001
sex	1	0.3485	0.0153	22.84	<.0001
race4	1	-0.0574	0.0231	-2.49	0.0134
race4	2	-0.2008	0.0288	-6.97	<.0001
race4	3	0.0351	0.0242	1.45	0.1484
grade	1	0.0444	0.0200	2.22	0.0271
grade	2	0.0446	0.0204	2.19	0.0297
grade	3	0.0178	0.0239	0.74	0.4572
T3L_L		0.2183	0.0460	4.75	<.0001

NOTE: The degrees of freedom for the t tests is 257.

Table 1:*Felt sad or hopeless*   Table 2: *Seriously considered suicide*

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		-1.9216	0.0210	-91.32	<.0001
sex	1	0.2808	0.0158	17.80	<.0001
race4	1	-0.0981	0.0254	-3.86	0.0001
race4	2	-0.1890	0.0335	-5.64	<.0001
race4	3	0.0602	0.0244	2.46	0.0144
grade	1	0.00583	0.0223	0.26	0.7934
grade	2	0.0924	0.0258	3.59	0.0004
grade	3	-0.00224	0.0272	-0.08	0.9346
T3L_L		0.2333	0.0524	4.46	<.0001

NOTE: The degrees of freedom for the t tests is 257.

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		-2.4734	0.0293	-84.47	<.0001
sex	1	0.3489	0.0228	15.29	<.0001
race4	1	-0.3554	0.0341	-10.43	<.0001
race4	2	-0.00616	0.0360	-0.17	0.8643
race4	3	0.1550	0.0318	4.88	<.0001
grade	1	0.1849	0.0273	6.78	<.0001
grade	2	0.1487	0.0335	4.43	<.0001
grade	3	-0.0770	0.0360	-2.14	0.0333
T3L_L		0.1245	0.0664	1.87	0.0621
T3Q		-0.1264	0.0627	-2.02	0.0448

NOTE: The degrees of freedom for the t tests is 255.

Table 3: Planned Suicide   Table 4: Attempted Suicide

T3L_L is linear continuous time variable created using proc IML in SAS®. T3Q is Quadratic continuous time variable. A logistic regression model was built using variable of interest as target controlling for sex, race, grade to test the significance in linear trends. Since p-value of the continuous time variable is <0.05, linear trend is significant.

2. Latent Class Analysis Result Snippet

Below is the result for the Latent Class Analysis run using PROC LCA. Class 1 is characterized by sexual assault, 2 by bullying. It can be seen that occurrence of depression is high in all three classes

Output 1 LCA Results

```
Class:                          1          2          3
                            0.1921     0.4064     0.4015


Rho estimates (item response probabilities):
  Response category  1 (Attempted Suicide):
Class  :                        1          2          3
   Sexual Assault  :        1.0000     0.1547     0.1500
   Depression      :        1.0000     0.8576     0.7028
   Bullying        :        0.6706     0.7853     0.0000
   Physical Fights :        0.5744     0.4693     0.3681
```
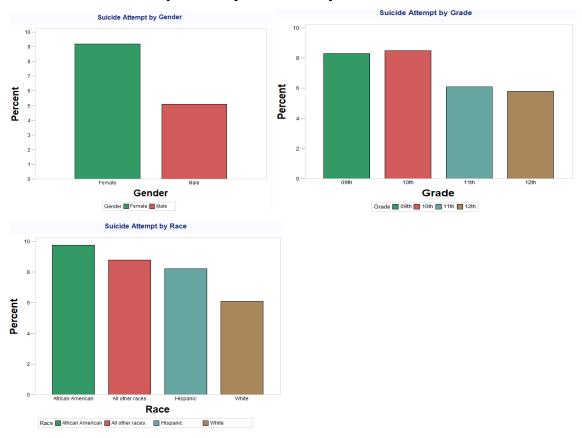
3. Logistic Regression Results

Table 5 Odds Ratio Estimates

Odds Ratio Estimates		Point Estimate	95% Confidence Limits	
Effect		Point Estimate		
sex	Male vs Female	0.724	0.548	0.956
race4	Black or African American vs All other races	1.057	0.749	1.492
race4	Hispanic/Latino vs All other races	0.894	0.638	1.252
race4	White vs All other races	0.633	0.514	0.778
grade	10th vs 12th	1.622	1.205	2.184
grade	11th vs 12th	1.005	0.707	1.428
grade	9th vs 12th	1.519	1.160	1.990
qn17	1 vs 2	2.182	1.620	2.938
qn19	1 vs 2	4.419	3.631	5.379
qn68	1 vs 2	1.436	1.112	1.856
narcoticdrugs 1 vs 2		2.847	1.760	4.604
LGBQ	1 vs 2	2.612	1.989	3.431
bully	1 vs 2	3.226	2.670	3.898

NOTE: The degrees of freedom in computing the confidence limits is 34.

TECHNICAL ADDENDUM

1. Additional Exploratory Data Analysis:

Suicide Attempt by Gender

Suicide Attempt by Grade

Suicide Attempt by Race

Two sample t-test to check if the average no. of suicide attempts in males and females are equal

Variable: qn28 (Attempted suicide)

sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		5532	1.9022	0.2971	0.00399	1.0000	2.0000
Male		5073	1.9452	0.2276	0.00320	1.0000	2.0000
Diff (1-2)	Pooled		-0.0430	0.2661	0.00517		
Diff (1-2)	Satterthwaite		-0.0430		0.00512		

sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		1.9022	1.8944	1.9100	0.2971	0.2916	0.3027
Male		1.9452	1.9389	1.9515	0.2276	0.2233	0.2321
Diff (1-2)	Pooled	-0.0430	-0.0531	-0.0329	0.2661	0.2626	0.2697
Diff (1-2)	Satterthwaite	-0.0430	-0.0530	-0.0330			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	10603	-8.31	<.0001
Satterthwaite	Unequal	10284	-8.41	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	5531	5072	1.70	<.0001

8

Anova Test to check if the differences in average no. of suicide attempts are significant between races

Welch's ANOVA for qn28			
Source	DF	F Value	Pr > F
race4	3.0000	9.08	<.0001
Error	3719.5		

Level of race4		qn28	
	N	Mean	Std Dev
All other races	1312	1.91082317	0.28510749
Black or African American	1605	1.89781931	0.30298018
Hispanic/Latino	2796	1.91917024	0.27262225
White	4773	1.93609889	0.24460232

2. PROC LCA Full results:

More than 15 different combinations on classes and variable were used, and the most significant result was used, selected on the basis of low AIC and interpretability

```
PROC LCA DATA=cons.CONSSA outpost=cons.res;
     NCLASS 3;
     id ID;
     ITEMS qn19 qn25 qn23 qn17;
     CATEGORIES 2 2 2 2;
     maxiter 999999;
     SEED 861551;
RUN;
```

```
Friday, November 22, 2019 01:36:00 PM   1

                                          '5-class model, 7 items'


            Data Summary, Model Information, and Fit Statistics (EM
Algorithm)




Number of subjects in dataset:        837

Number of subjects in analysis:       837


Number of measurement items:           4

Response categories per item:         2 2 2 2

Number of groups in the data:          1

Number of latent classes:              3

Rho starting values were randomly generated (seed = 861551).


No parameter restrictions were specified (freely estimated).


The model converged in 10430 iterations.


Maximum number of iterations: 999999
```

```
Convergence method: maximum absolute deviation (MAD)

Convergence criterion:  0.000001000


=============================================
Fit statistics:
=============================================


Log-likelihood:      -1970.68

G-squared:              1.03

AIC:                   29.03

BIC:                   95.25

CAIC:                 109.25

Adjusted BIC:          50.79

Entropy:                0.63

Degrees of freedom:       1


Test for MCAR

      Log-likelihood:     -1970.16

      G-squared:             53.64

      Degrees of freedom:      47



Friday, November 22, 2019 01:36:00 PM    2

                                         '5-class model, 7 items'


                        Parameter Estimates


(Standard errors could not be computed; please see the log file for details.
)

Gamma estimates (class membership probabilities):

Class:                       1          2          3
                          0.1921     0.4064     0.4015


Rho estimates (item response probabilities):
  Response category  1:
Class:                       1          2          3
  QN19      :        1.0000     0.1547     0.1500

  QN25      :        1.0000     0.8576     0.7028

  QN23      :        0.6706     0.7853     0.0000
```

```
   QN17         :         0.5744      0.4693      0.3681


   Response category  2:
Class:                      1           2           3
   QN19         :         0.0000      0.8453      0.8500

   QN25         :         0.0000      0.1424      0.2972

   QN23         :         0.3294      0.2147      1.0000

   QN17         :         0.4256      0.5307      0.6319
```
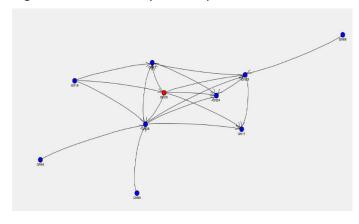
3. High-Performance BN Classifier (PC algorithm) details:

Figure 5 Settings used

Network Model	Bayesian Network
Automatic Model Selection	No
Prescreen Variables	No
Variable Selection	Yes
Independence Test Statisti	G-Square
Significance Level	0.05
Missing Interval Variable	None
Missing Class Variable	None
Number of Bins	10
Maximum Parents	5
Network Structure	Markov Blanket
Parenting Method	Set of Parents

Figure 6 Directed Acyclic Graph

4. PROC SURVEYLOGISTIC results:

Table 6 Interaction terms were not significant

Analysis of Maximum Likelihood Estimates

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		1.4522	0.3381	4.29	0.0001
sex	Male	-0.0563	0.0783	-0.72	0.4769
race4	Black or African American	0.1814	0.1228	1.48	0.1493
race4	Hispanic/Latino	-0.00566	0.1312	-0.04	0.9658
race4	White	-0.2500	0.0835	-2.99	0.0052
grade	10th	0.2262	0.0799	2.83	0.0078
grade	11th	-0.1824	0.1088	-1.68	0.1030
grade	9th	0.2109	0.0906	2.33	0.0263
qn19	1	0.6016	0.2646	2.27	0.0296
qn17	1	0.2990	0.0718	4.17	0.0002
qn68	1	0.1756	0.0663	2.65	0.0123
qn23	1	0.1996	0.1751	1.14	0.2625
qn25		-2.0416	0.2361	-8.65	<.0001
qn24	1	0.2482	0.0958	2.59	0.0141
narcoticdrugs	1	0.4234	0.2627	1.61	0.1166
qn23*qn17	1 1	0.00101	0.0692	0.01	0.9884
qn25*qn23	1	0.0611	0.1379	0.44	0.6607
qn25*qn19	1	0.0106	0.2221	0.05	0.9621
qn25*narcoticdrugs	1	-0.0179	0.1939	-0.09	0.9269

NOTE: The degrees of freedom for the t tests is 33.

Table 7 Excluding interaction terms (Final model)

Odds Ratio Estimates

Effect		Point Estimate	95% Confidence Limits	
sex	Male vs Female	0.724	0.548	0.956
race4	Black or African American vs All other races	1.057	0.749	1.492
race4	Hispanic/Latino vs All other races	0.894	0.638	1.252
race4	White vs All other races	0.633	0.514	0.778
grade	10th vs 12th	1.622	1.205	2.184
grade	11th vs 12th	1.005	0.707	1.428
grade	9th vs 12th	1.519	1.160	1.990
qn17	1 vs 2	2.182	1.620	2.938
qn19	1 vs 2	4.419	3.631	5.379
qn68	1 vs 2	1.436	1.112	1.856
narcoticdrugs 1 vs 2		2.847	1.760	4.604
LGBQ	1 vs 2	2.612	1.989	3.431
bully	1 vs 2	3.226	2.670	3.898

NOTE: The degrees of freedom in computing the confidence limits is 34.

5. Code for testing significance in trend analysis

```
proc means data = yrbs.sadc_2017_district;
run;

PROC IML;
```

```
X={2007 2009 2011 2013 2015 2017 };
XP=ORPOL(X,1);
PRINT XP;
RUN;
QUIT;


PROC IML;
X={2007 2009 2011 2013 2015 2017 };
XP=ORPOL(X,2);
PRINT XP;
RUN;
QUIT;


DATA VARSET;
SET yrbs.sadc_2017_district; /* Testing for linearity to the left of the
joinpoint. */
 IF (YEAR=2007 OR YEAR=2009 OR YEAR=2011 OR YEAR=2013 OR YEAR=2015 OR
YEAR=2017);
 IF YEAR=2007 THEN DO ;
 T3L_L=-0.597614;
 T3Q = 0.5455447;
 END;
ELSE IF YEAR=2009 THEN DO;
T3L_L=-0.358569;
T3Q= -0.109109;
END;
ELSE IF YEAR=2011 THEN DO;
T3L_L=-0.119523;
T3Q = -0.436436;
END;
ELSE IF YEAR=2013 THEN DO;
T3L_L= 0.119523;
T3Q=-0.436436;
END;
ELSE IF YEAR=2015 THEN DO;
T3L_L= 0.3585686;
T3Q = -0.109109;
END;
ELSE IF YEAR=2017 THEN DO;
T3L_L= 0.5976143;
T3Q = 0.5455447;
END;
run;

DATA VARSET;
SET yrbs.sadc_2017_district; /* Testing for linearity to the left of the
joinpoint. */
 IF (YEAR=2007 OR YEAR=2009 OR YEAR=2011 OR YEAR=2013 OR YEAR=2015 OR
YEAR=2017);
 IF YEAR=2007 THEN
 T3L_L=-0.597614;

ELSE IF YEAR=2009 THEN
T3L_L=-0.358569;

ELSE IF YEAR=2011 THEN
```

```
T3L_L=-0.119523;

ELSE IF YEAR=2013 THEN
T3L_L= 0.119523;

ELSE IF YEAR=2015 THEN
T3L_L= 0.3585686;

ELSE IF YEAR=2017 THEN
T3L_L= 0.5976143;


run;

/* Table 1: Depressed trend */
proc surveylogistic data = varset;
CLASS SEX RACE4 GRADE YEAR SURVYEAR;
model qn25 = SEX RACE4 GRADE T3L_L;
strata stratum ;
cluster psu ;
weight weight ;
run;

/* Table 2: Considered Suicide */
proc surveylogistic data = varset;
CLASS SEX RACE4 GRADE YEAR SURVYEAR;
model qn26 = SEX RACE4 GRADE T3L_L;
strata stratum ;
cluster psu ;
weight weight ;
run;

/* Table 3: Planned Suicide */
proc surveylogistic data = varset;
CLASS SEX RACE4 GRADE YEAR SURVYEAR;
model qn27 = SEX RACE4 GRADE T3L_L;
strata stratum ;
cluster psu ;
weight weight ;
run;

/* Table 4: Attempted Suicide */
proc surveylogistic data = varset;
CLASS SEX RACE4 GRADE YEAR SURVYEAR;
model qn28 = SEX RACE4 GRADE T3L_L T3Q;
strata stratum ;
cluster psu ;
weight weight ;
run;
```

6. Code for imputation and survey logistic

```
data x;
set yrbs.sadc_2017_district;
if year = 2017;
run;

data x_drug;
```

```
   set x;
   if qn57=1 or qnhallucdrug =1 then narcoticdrugs = 1;
   else narcoticdrugs = 2;
   run;

    data x_drug;
    set x_drug;
    if sexid in(2,3,4) then LGBQ = '1';
    else LGBQ = '2';
    run;
    data x_drug;
    set x_drug;
    if qn23 = 1 or qn24 =1 then bully = '1';
    else bully = '2';
    run;

   proc surveyimpute data=x_drug method=hotdeck(selection=WEIGHTED) seed=3242
ndonors=1;
   Var qn17 qn19 qn89 qn23 qn24 qn25 SEX RACE4 GRADE narcoticdrugs LGBQ
bully;
   Class qn17 qn19 qn89 qn23 qn24 qn25 SEX RACE4 GRADE narcoticdrugs LGBQ
bully;
   weight weight;
   Cluster PSU;
   Strata stratum;
   output out= x_drugImp;
   run;

/* Table 7: Code for Survey Logistic Results*/
   proc surveylogistic data = x_drugImp;
   CLASS SEX(ref ='Female') RACE4(ref = 'All other races') GRADE(ref =
'12th') qn19(ref = '2') qn68(ref = '2')
   qn23(ref = '2') qn17(ref='2') qn24(ref = '2') narcoticdrugs(ref = '2')
LGBQ(ref='2') bully(ref='2');
   model qn28(event ='1') = SEX RACE4 GRADE qn17 qn19 qn68 narcoticdrugs LGBQ
bully /rsquare;
   strata stratum ;
   cluster psu ;
   weight weight ;
   run;
```