Georgia State University
Digital Archive @ GSU

Mathematics Theses | Department of Mathematics and Statistics

12-6-2006

Analyzing the Effects of Adolescent Risky Behaviors on Suicidal Ideation

Marchelle Elizabeth Sanchez
marchellesanchez@aim.com

Follow this and additional works at: http://digitalarchive.gsu.edu/math_theses

Recommended Citation
Sanchez, Marchelle Elizabeth, "Analyzing the Effects of Adolescent Risky Behaviors on Suicidal Ideation" (2006). *Mathematics Theses.* Paper 21.

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at Digital Archive @ GSU. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

ANALYZING THE EFFECTS OF ADOLESCENT RISKY BEHAVIORS ON SUICIDAL IDEATION

by

MARCHELLE E SANCHEZ

Under the Direction of Yu-Sheng Hsu

ABSTRACT

This study is an analysis of adolescent risk behaviors contributing to an increased rate of suicidal ideation for 12 to 18 year olds. The Youth Risk Behavior Surveillance System Survey (YRBSS) is an epidemiologic survey designed to monitor the prevalence of risky behaviors of adolescents in middle and high school[1]. The YRBSS is a complex sample survey with a three-stage cluster design. Multiple logistic regression is used to analyze the data, including methods of analysis to address issues in complex survey design. Results of this study indicate several different risk factors that influence the rate of suicidal ideation among adolescents, including alcohol and drug use, sexual risky behaviors, unhealthy weight loss methods, depressed mood, sex and race/ethnicity. The conclusions of this study indicate that many risk factors associated with suicidal ideation are behaviors that could be addressed with early intervention strategies to reduce the risk of suicidal ideation.

INDEX WORDS: Logistic regression, Complex sample surveys, Suicidal ideation, Suicide, Suicidal behaviors, Adolescents, Youth risk behaviors

ANALYZING THE EFFECTS OF ADOLESCENT RISKY BEHAVIORS ON SUICIDAL IDEATION

by

MARCHELLE ELIZABETH SANCHEZ

A Thesis Submitted In Partial Fulfillment of the Requirements for the Degree of

Masters of Science

In the College of Arts and Sciences

Georgia State University

2006

Copyright © by
Marchelle Elizabeth Sanchez
2006

ANALYZING THE EFFECTS OF ADOLESCENT RISKY BEHAVIORS
ON SUICIDAL IDEATION

by

MARCHELLE ELIZABETH SANCHEZ

Major Professor: Dr. Yu-Sheng Hsu
Committee: Dr. Kakoli Roy
Dr. Xu Zhang

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December, 2006

ACKNOWLEDGMENTS

I would like to acknowledge the support from my committee members, Dr. Yu-Sheng Hsu, Dr. Kakoli Roy, and Dr. Xu Zhang. Dr. Hsu, thank you for all of your help, encouragement and motivation during the completion of my thesis. Dr. Roy, I have truly appreciated your advice, mentoring and mostly your friendship throughout this process. Dr. Zhang, I would like to extend my gratitude to you for helping me with SAS programming and encouraging me to keep going, even when I was ready to give up.

I extend a special thank you to Dr. Owen Devine for the last minute inspiration and insight into my analysis. Thank you, Latonia Clay and Ying Shen, for your continued support, friendship, and a shoulder to lean on during these past few months. To my sister, Lisa Arnold, thank you for all of your comments and suggestions on editing.

Most importantly, I would like to thank my family for putting up with the long hours, missed dinners and late nights that you have endured. Alexander, you have been such a great sport, and I have loved all of your thoughts, questions and remarks about my mathematical and statistical abilities! Maddie, you have been so patient and understanding and I loved having your company during those long hours of work! Last, but certainly not least, I would like to thank my husband, Peter. There are so many things you have helped me with and allowed me to do during the pursuit of my goals. I am looking forward to giving back all that you have given me, and much more, in return for all you have done.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS…….. iv

LIST OF TABLES……. vi

LIST OF FIGURES……..... vii

LIST OF ABREVIATIONS……..……. viii

CHAPTER

1 INTRODUCTION……. 1

Background…… 1

Source of Data…… 3

Method of 6

2 METHODOLOGY…… 7

Variable 7

Statistical 8

Goodness of Fit……... 14

3 RESULTS…….. 21

Odds 21

4 DISCUSSION…… 25

REFERENCES…… 28

APPENDICES…… 30

Appendix A – SAS Code…… 30

Appendix B – SUDAAN Code……... 35

LIST OF TABLES

Table 1. Weighted Frequencies and Percents for YRBSS...……. 5

Table 2. Variable Significance for Univariate Models…….……..... 9

Table 3. Estimated Coefficients for Preliminary Model vs the Univariate Model…….. 10

Table 4. Log-Likelihood Ratios for Transformation of Age……... 11

Table 5. Significant Variables Included in Final Model for Risk Behaviors and Suicidal Ideation……... 13

Table 6. Model Fit Statistics for Preliminary and Final Logistic Regression Models…. 14

Table 7. Hosmer-Lemeshow Goodness of Fit Tests……. 16

Table 8. Wald F Test for Coefficient Significance…….……... 17

Table 9. Classification Table for Sensitivity and Specificity…….……... 18

Table 10. Estimated Odds Ratios and 95% Confidence Intervals….…….…… 22

Table 11. Odds Ratios For Interaction Between Sex & Depressed Mood……... 23

Table 12. Estimated Odds Ratios Controlling For Sex……. 24

Table 13. Trends in the Prevalence Rate of Suicidal Ideation and Attempts[9]…...…….. 26

LIST OF FIGURES

Figure 1. Ten Leading Causes of Death in 2003 for Ages 12 – 18 in the U.S.[9]……..... 2

Figure 2. 2003 Causes of Death By Suicide for Ages 12 – 18 [9]……... 2

Figure 3. Map of YRBSS States By Weight[1]……... 4

Figure 4. Sensitivity Versus 1–Specificity Plotted Against $\hat{\pi}$ …… 19

Figure 5. Receiver Operating Characteristic Curve (ROC)……..... 20

LIST OF ABBREVIATIONS

CDC	Centers for Disease Control
CI	Confidence Interval
HL	Hosmer-Lemeshow Test
OR	Odds Ratio
PSU	Primary Sampling Unit
ROC	Receiver Operating Characteristic
YRBSS	Youth Risk Behavior Surveillance System

Variable Abbreviations for YRBSS Questionnaire[1]:

ANYTB	Did you smoke cigarettes, cigars, use chewing tobacco, snuff or dip?*
CCNL	Have you used any form of cocaine, including powder, crack, or freebase?***
CCNR	Did you use any form of cocaine, including powder, crack, or freebase?*
CGUN	Did you carry a gun?*
CWPN	Did you carry a weapon such as a gun, knife, or club?*
DRPT	Did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?*
DWI	Did you drive a car or other vehicle when you had been drinking alcohol?*
ECST	Have you used ecstasy (also called MDMA)?***
ETHR1	American Indian or Alaska Native
ETHR2	Asian, Native Hawaiian or Other Pacific Islander
ETHR3	Black or African American
ETHR4	Hispanic or Latino
ETHR5	White
FAST	Did you go without eating for 24 hours or more (also called fasting) to lose weight or to keep from gaining weight?*
FGHT	Were you in a physical fight?**
FRCIG	Did you smoke more than 20 out of the last 30 days?

HAL	Have you ever taken any hallucinogenic drugs?***
HBGF	Did your boyfriend or girlfriend ever hit, slap, or physically hurt you on purpose?***
HLMT	If you rode a motorcycle during the past 12 months, did you wear a helmet?
HLTH	Do you consider your health to be fair or poor?
HRN	Have you used heroin (also called smack, junk, or China White)?***
IDA	Did you drink alcohol or use drugs before you had sexual intercourse the last time?
INJC	Have you used a needle to inject any illegal drug into your body?***
METH	Have you used methamphetamines (also called speed, crystal, crank, or ice)?***
MOOD	Did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?**
MRJA	Have you used marijuana?**
RALC	Did you have at least one drink of alcohol?*
SIM4	During the past 3 months, did you have sexual intercourse with more than four people?
SNF	Have you sniffed glue, breathed the contents of aerosol spray cans, or inhaled any paints or sprays to get high?***
SPRT	Did you play on any sports teams? (Include any teams run by your school or community groups.)**
STR	Have you taken steroid pills or shots without a doctor's prescription?***
SUID	Did you ever seriously consider attempting suicide?**
THRT	Has someone threatened or injured you with a weapon such as a gun, knife, or club on school property?**
UNSF	Did you not go to school because you felt you would be unsafe at school or on your way to or from school?*
VMT	Did you vomit or take laxatives to lose weight or to keep from gaining weight?*

**During the past 30 days?*
***During the past 12 months?*
****During your life?*

Chapter 1

INTRODUCTION

Background

Adolescence is a time in life filled with new experiences, opportunities, and freedom. Adolescence also brings with it a time of stress, concern, family tension, and other difficulties that tend to arise as one gets older. Adolescents may find themselves in difficult situations, with challenges and stressful events that they find very hard to deal with[2,3,4,5]. Many adolescents find unhealthy ways to show their distress by drinking alcohol[6], experimenting with drugs[5,7], violent behavior[5,7] and other risky behaviors[8]. In 2003, suicide was the third leading cause of death for ages 12 – 18 years[2,3,9]. Many experts on suicide and suicide prevention suggest that suicide prevention programs should focus on finding youth that are at risk of suicide and then refer the youth to treatment programs[2,7]. Suicidal ideation is a medical term for thoughts about suicide, which can include role playing, a detailed or formulated plan without acting on the plan, or even an unsuccessful attempt at suicide[10]. By determining potential risk factors, suicide prevention programs will be able to assess an individual's risk of suicidal ideation[7], and to identify and treat individuals that are at higher risk of suicidal ideation[2,11]. This study is focused on looking at risky behaviors in adolescents to determine if there is an association of certain risky behaviors and suicidal ideations in youth.

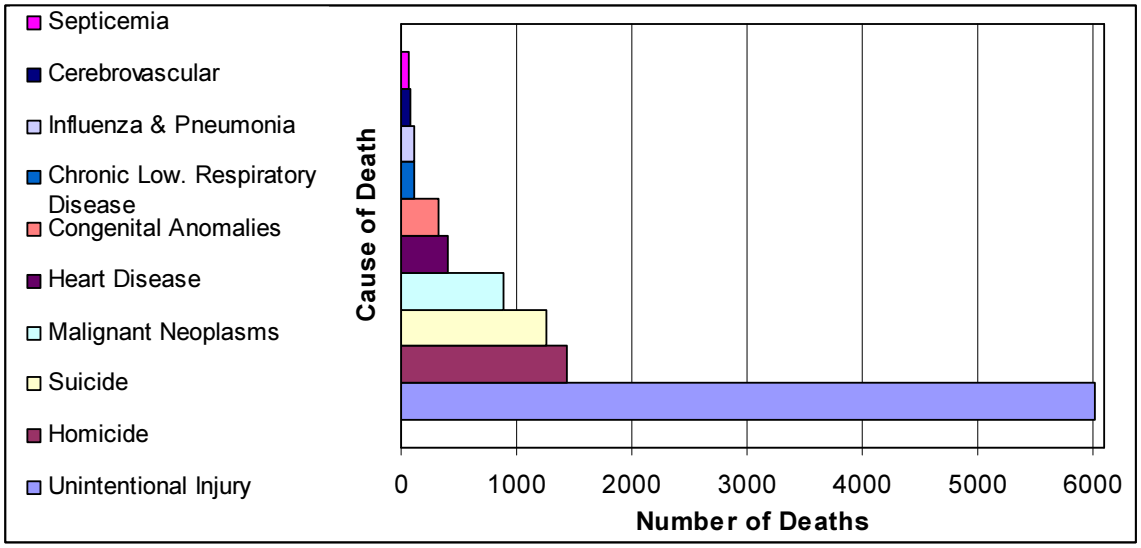

Figure 1: Ten Leading Causes of Death in 2003 for Ages 12 – 18 in the U.S.[9]

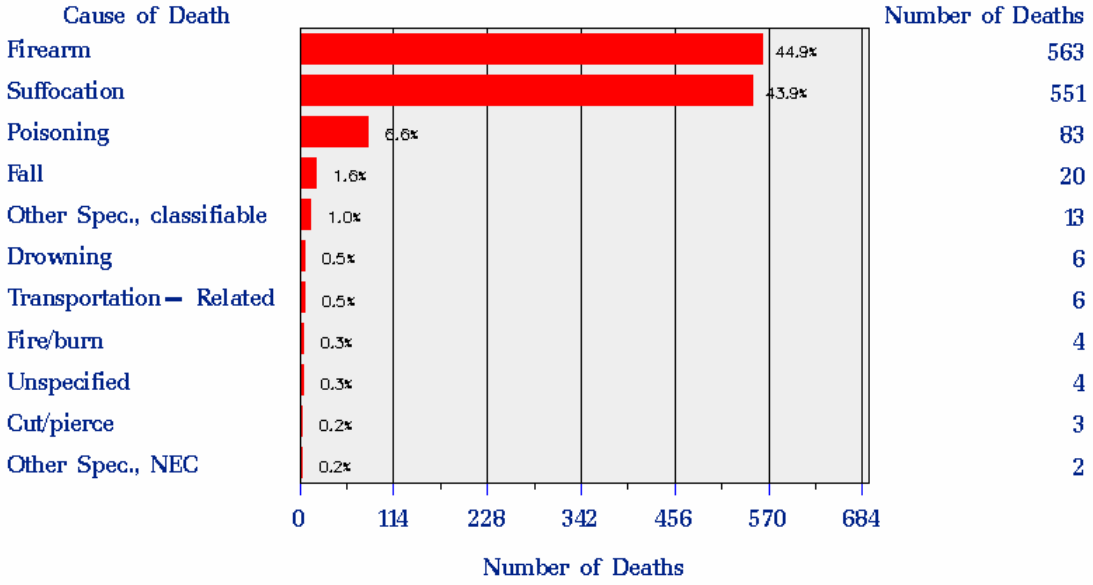

Figure 2: 2003 Causes of Death By Suicide for Ages 12 – 18 [9]

Much research has been devoted to finding correlations between adolescent risky behaviors and suicidal ideation. Various health risk behaviors have been identified that may lead to suicidal ideation and behaviors[2,5,7,8]. Behaviors such as substance use[2,5], unprotected and precocious sexual activity[4] and reckless driving[11] have been linked to an increase in suicidal behaviors, as well as self-mutilation[11] and previous suicide attempts[2,5]. For different races and ethnic groups, studies have shown that suicide is more common among whites than for African Americans[2,4]. The highest rates of suicide are among Native Americans[2,4], and Asians and Pacific Islanders tend to have the lowest rates of suicide[2,4,11].

Source of Data

The data for this study comes from the 2005 Youth Risk Behavior Surveillance System survey (YRBSS)[1], conducted by the Centers for Disease Control and Prevention (CDC). The YRBSS data is an epidemiologic surveillance system created to monitor the risk behaviors that are most prevalent among youth that adversely affect health. The focus of the questions from the YRBSS come from priority health-risk behaviors that are established during youth and that result in significant mortality, morbidity, disability and social problems that impact youth and can continue to impact their lives through adulthood as well[1,4].

The YRBSS is a complex survey design comprised of a three-stage cluster sample design to conduct a representative sample of 9th through 12th grade students. Data for the

YRBSS was collected from 44 states in the US, including Alaska and Hawaii[1]. Figure 2 shows the distribution map for the weighted data by state. The data come from many different school settings, such as private, public, and religious schools. The YRBSS oversamples the black and Hispanic population, and therefore a weighting factor is applied to each student in order to account for the oversampling and nonresponses in the survey. Final overall weights are also scaled to adjust the weighted counts of the individuals to equal the total sample population size, and so that the weighted proportions of students in each grade matched population projections for each survey year.

Figure 2: Map of YRBSS States by Weight[1]

The sample population of the YRBSS data consisted of 13,953 returned surveys in which 13,917 of the questionnaires were useable after data editing. Inclusion criteria for this study was based on all students who responded to the question '*During the past 12 months, did you ever seriously consider attempting suicide?*', which yielded a total sample size of 13,822 subjects for this study. The weighted sample population consists of 49.4% males and 50.6% females. The mean age of the sample population is 15 years and 10 months old. Weighted frequencies and percent for race and ethnicity in this study are as follows: American Indians - 141 (1.02%), Asians - 563 (4.074%), Black or African Americans -1947 (14.094%), Hispanic – 2054 (14.873%), White – 8473 (61.350%), and Other – 634 (4.59%). For the total sample population, the weighted frequency for students who had seriously considered suicide within the past 12 months was 2327 students, 16.85% of the sample population. The actual and weighted frequencies for age, sex, race and ethnicity, and suicidal ideation are included in Table 1.

Table 1: Weighted Frequencies and Percents for YRBSS

		Weighted				*Weighted*	
	Frequency	***Freq***	***%***		***Frequency***	***Freq***	***%***
Age				**Race/Ethnicity**			
≤14	1204	1457	10.58	American Indian	145	141	1.02
15	3158	3642	26.46	Asian	450	563	4.07
16	5762	5445	39.56	Black or African American	3294	1947	14.09
≥17	3649	3221	23.40	Hispanic	3204	2054	14.87
Sex				White	6115	8473	61.35
Female	7147	6803	49.44	Other	614	634	4.59
Male	6615	6956	50.56				
Suicidal Ideation							
Yes	2341	2327	83.15				
No	11481	11484	16.85				

Method of Analysis

The YRBSS questionnaire consists primarily of dichotomous variables. Variables such as race, ethnicity, age and grade are the primary variables that do not have binary responses for the YRBSS data. Many different methods are available to analyze complex sample survey data; however, for this study multiple logistic regression has been chosen as an optimal method to use for the dichotomous variables in the YRBSS data. Due to the complexity of the sampling survey, an inclusion of the weights, stratum and primary sampling units (psu's) are accounted for in the analysis of this data. While PROC LOGISTIC in SAS v8.2 is a useful tool to use for analyzing simple survey data, SAS v9.1 has developed PROC SURVEYLOGISTIC for analyzing complex survey design. Most statistical programs such as SAS, STATA, and SUDAAN do not have the capability to do full model checking and goodness of fit tests for complex sample survey data at this time[12]. SUDAAN and STATA do offer some options for model checking, such as the Wald F test and adjusted Wald F test. Therefore, in order to accurately account for the complexities of the YRBSS data, a combination of SAS v9.1 and SUDAAN software has been used for analysis of the YRBSS data.

Chapter 2

METHODOLOGY

Variable Selection

This study focuses on the risk behaviors associated with suicidal ideation. Several questions in the YRBSS questionnaire ask about suicidal thoughts, actions, attempts and injuries from suicidal ideations or behaviors. For this study, suicidal ideation (SUID) is based on the yes/no response to the question '*During the past 12 months, did you ever seriously consider attempting suicide?*'[1]. Other questions inquired more specifically into suicidal behaviors; however, due to the nature of sample surveys, associations with successful attempts of suicide are unattainable via sample surveys[3].

Risk behaviors previously shown to contribute to suicidal ideations were tested for significance in the study population. The risk behaviors identified as possible contributors to suicidal ideation were categorized in terms of demographics, drug and alcohol use, violence, sexually risky behaviors, and other health risk behaviors.

The demographic variables assessed for significance were race[2,4], age[4] and sex[2,4]. The variables of interest for violence[11] included weapon carrying[4,5], physical fights[7], and being physically threatened by another individual[2]. Variables of interest for drug usage[4] included previous and/or current use of hallucinogenic drugs, steroids, cocaine[5], heroine[11], methamphetamines, ecstasy, injection of illegal drugs by needle, and using inhalants such as aerosol sprays to get high[5]. Chain smoking was also considered for further

investigation[5,11]. Alcohol use[11] questions included for analysis dealt with binge drinking, heavy alcohol use, and driving under the influence of alcohol[4,5,6]. Sexually risky behaviors[4] such as multiple partners, use of drugs and/or alcohol prior to sexual intercourse, not using protection during sex, and also being physically hurt or assaulted by a boyfriend/girlfriend were identified as possible influences[4,5]. Other health risk factors considered for analysis included self-assessment of overall health, and unhealthy behaviors such as vomiting, fasting or using laxatives to lose weight or to keep from gaining weight[5]. A significant indicator of suicidal ideation is the state of mind of an individual[2,4,5]. As such, depressed mood is also included as a variable of significance in determining risk for suicidal ideation. Each of the dependent variables aside from age were categorized as dichotomous yes/no questions, which lead to the method of logistic regression as an optimal choice for analyzing this data.

Statistical Analysis

A univariate analysis was conducted on each of the independent variables to determine which variables were significant in predicting suicidal ideation. Variables that were significant at the $\alpha = 0.25$[12] level were then combined to test the covariates in a full model as predictors for suicidal ideation. Using the process of backwards elimination, starting from the full model, and eliminating those variables that did not contribute to the model at the specified $\alpha = 0.25$ level lead to a preliminary model of independent variables that were selected to model the data. Variables and their significance level for the

univariate models are shown in Table 2, as well as estimated coefficients, standard errors, and Wald χ^2 statistics.

Table 2. Variable Significance for Univariate Models

Demographics & Mood Factors				
Variable	**Estimate**	**StdErr**	**Wald χ^2**	**P-Value**
AGE	-0.0555	0.0358	2.4092	0.1206
SEX	0.7141	0.0494	208.7169	0.0000
ETHR1	0.4187	0.2072	4.0854	0.0433
ETHR2	0.1699	0.1285	1.7503	0.1858
ETHR3	-0.4324	0.0804	28.9028	0.0000
ETHR4	0.0882	0.0673	1.7146	0.1904
ETHR5	0.0064	0.0633	0.0101	0.9198
MOOD	2.3253	0.0637	1330.6021	0.0000

Violence, Sex and Other Risk Factors				
CGUN	0.3433	0.1231	7.7731	0.0053
CWPN	0.8280	0.1244	44.2957	0.0000
UNSF	0.7304	0.1096	44.4320	0.0000
THRT	1.1124	0.0859	167.5839	0.0000
FGHT	0.8084	0.0628	165.9798	0.0000
HBGF	1.0322	0.0756	186.2859	0.0000
SIM4	0.5332	0.0767	48.2871	0.0000
HLMT	0.3337	0.1405	5.6444	0.0175

Smoking , Alcohol & Drug Use				
Variable	**Estimate**	**StdErr**	**Wald χ^2**	**P-Value**
DWI	0.5485	0.0828	43.9301	0.0000
DRPT	0.7135	0.0729	95.9204	0.0000
RALC	0.6630	0.0760	76.1772	0.0000
MRJA	0.7500	0.0671	125.1029	0.0000
CCNL	1.2576	0.0964	170.0606	0.0000
CCNR	1.1877	0.1319	81.1124	0.0000
SNF	1.4106	0.0845	278.9000	0.0000
HRN	1.6693	0.1268	173.3050	0.0000
METH	1.4453	0.0849	289.6941	0.0000
ECST	1.2562	0.1034	147.5151	0.0000
STR	1.3484	0.1036	169.3606	0.0000
INJC	1.7060	0.1204	200.7796	0.0000
IDA	0.5541	0.1171	22.3754	0.0000
HAL	1.2151	0.1031	138.9565	0.0000
ANYTB	0.9193	0.0696	174.2529	0.0000
FRCIG	1.1601	0.0933	154.4996	0.0000

Health Factors				
HLTH	1.0336	0.1056	95.7883	0.0000
FAST	1.4803	0.0816	328.7176	0.0000
VMT	1.6335	0.1065	235.4574	0.0000
SPRT	-0.4448	0.0569	61.0496	0.0000

The variables included in the preliminary model, and which contribute significantly to the response variable of suicidal ideation (SUID) are age (AGE), sex (SEX), depressed mood (MOOD), usage of inhalants to get high (SNF), steroid use (STR), using alcohol and/or drugs prior to intercourse (IDA), vomiting or using laxatives to lose weight or to

keep from gaining weight (VMT) , and the race/ethnic groups for Black and African Americans (ETHR3), and Hispanics (ETHR4).

Table 3 lists the variables selected for the preliminary model, as well as the estimated coefficients, standard errors, Wald χ^2 statistics and p-values. Variable statistics from the univariate models are also shown for a comparison of the estimates and changes in significance of a variable from the univariate model to an inclusive model.

Table 3. Estimated Coefficients for Preliminary Model vs the Univariate Model

Variable	*PRELIMINARY MODEL*				*UNIVARIATE MODELS*			
	Estimate	**StdErr**	**Wald χ^2**	**p-value**	**Estimate**	**StdErr**	**Wald χ^2**	**p-value**
AGE	-0.1606	0.0600	7.1573	0.0075	-0.0555	0.0358	2.4092	0.1206
SEX	0.3337	0.0727	21.0665	0.0000	0.7141	0.0494	208.7169	0.0000
MOOD	1.9329	0.1019	359.6578	0.0000	2.3253	0.0637	1330.6021	0.0000
SNF	0.5864	0.1804	10.5690	0.0011	1.4106	0.0845	278.9000	0.0000
STR	0.6330	0.2123	8.8907	0.0029	1.3484	0.1036	169.3606	0.0000
IDA	0.2479	0.1334	3.4566	0.0630	0.5541	0.1171	22.3754	0.0000
VMT	0.8721	0.1534	32.3237	0.0000	1.6335	0.1065	235.4574	0.0000
ETHR3	-0.8274	0.1153	51.5215	0.0000	-0.4324	0.0804	28.9028	0.0000
ETHR4	-0.1975	0.0921	4.5947	0.0321	0.0882	0.0673	1.7146	0.1904

An assumption of logistic regression modeling is that the logit is linear in nature[12]. Many times continuous variables may not be linear, and thus must undergo a transformation. Age is the only continuous variable that is included in the preliminary model, and is investigated further to determine if a transformation is necessary for maintaining linearity of the logit. Using the log-likelihood of several different models with age as the independent variable for predicting suicidal ideation, we can compare the log-likelihood ratios to the linear logit model of age to determine the best transformation of age. Table 4 shows the values for the log-likelihood ratio tests performed for the

transformation analysis of age. Based on the results of the transformations, the linear model and the best J=2 model have the highest log-likelihoods. Using the partial likelihood ratio test, there is not a significant difference in the log-likelihoods between the J=2 model and the linear model, therefore retaining age as a linear variable will not significantly impact the linearity of the logit.

Table 4. Log-Likelihood Ratios for Transformation of Age

Transformation Models:				-2 Log L
Best J=2 model:	Age = 1 + age³ + age³*ln*(age)			12494.2
Best J=1 model:	Age = *ln*(age)			12465.1
Linear model:				12493.4
df	**Partial Likelihood Ratio Test:**			
1	G(1,p1)=	-56.644	p=	1.0000
2	G(p1,p2)=	58.256	p=	0.0000
3	G(1,p2)=	1.6120	p=	0.6567
	Age : No Transformation Needed			
	Transformation Check:			
	Linear	-2 Log L=	12493.4	
	Age transformed	-2 Log L=	12494.2	
	χ^2 = 1.612	p-value =	0.6567	

Interaction effects are an important consideration for any model, and must be thoroughly explored to determine if there is a significant interaction that should be included in the final model. Based on the variables remaining in the preliminary model, interaction terms were created as the product of two variables for each combination of the significant independent variables, and then tested for significance in the main effects model to determine if the interaction term should be included in a final model. The level of significance for interaction terms to be included as significant to a model is based on a traditional significance level of $\alpha = 0.05$[12]. The only interaction significant at the $\alpha = 0.05$

level for the YRBSS data is the interaction between SEX and MOOD. There is a significant amount of literature available in regards to the interaction between sex and depressed mood[2,4]. Studies show that females are much more likely to attempt suicide, and to suffer depression, although males are much more likely to succeed in their attempts of suicide[7].

Based on the results of the variable selection process, a final model has been developed to help determine the risk factors associated with suicidal ideation for the YRBSS data in the study. Table 5 gives the question from YRBSS that determined the risk factors associated with suicidal ideation, the significance level, estimated coefficient and standard error, as well as the Wald χ^2 statistic for each significant variable. The final logistic regression model based on the YRBSS data for suicidal ideation is:

$$\hat{\pi}(x) = \frac{\left(\begin{array}{l}\exp\{-1.8129 - 0.1652AGE + 0.5991SEX + 2.2009MOOD + 0.5970SNF + 0.6246STR \\ + 0.2542IDA + 0.8797VMT - 0.8167ETHR3 - 0.2010ETHR4 - 0.4680SEX * MOOD\}\end{array}\right)}{\left(\begin{array}{l}1 + \exp\{-1.8129 - 0.1652AGE + 0.5991SEX + 2.2009MOOD + 0.5970SNF + 0.6246STR \\ + 0.2542IDA + 0.8797VMT - 0.8167ETHR3 - 0.2010ETHR4 - 0.4680SEX * MOOD\}\end{array}\right)}$$

Table 6 gives model fit statistics for both the preliminary model, without any interactions, and also for the final model, including the sex*mood interaction term. Each of these models are significant in terms of their score tests, likelihood ratios, and the Wald statistic. More discussion in terms of the model fit for the final model will be addressed in the following chapter.

Table 5. Significant Variables Included in Final Model for Risk Behaviors and Suicidal Ideation

Variable Description	Parameter	DF	$\hat{\beta}$	Error	χ^2	P-Value
	Intercept	1	-1.8129	0.3244	31.2276	0.0000
•Age Range 12 – 18 years old	AGE	1	-0.1652	0.0596	7.6805	0.0056
•Female = 1	SEX	1	0.5991	0.1452	17.0191	0.0000
•Student felt so sad or hopeless almost every day for two weeks or more in a row that he/she stopped doing some usual activities**	MOOD	1	2.2009	0.1745	159.0890	0.0000
•Student sniffed glue, breathed the contents of aerosol spray cans, or inhaled any paints or sprays to get high one or more times***	SNF	1	0.5970	0.1824	10.7085	0.0011
•Student took steroid pills or shots without a doctor's prescription one or more times***	STR	1	0.6246	0.2211	7.9827	0.0047
•Student had sexual intercourse during the past three months, and drank alcohol or used drugs before last sexual intercourse	IDA	1	0.2542	0.1345	3.5712	0.0588
•Student vomited or took laxatives to lose weight or to keep from gaining weight*	VMT	1	0.8797	0.1509	33.9826	0.0000
•Black or African American	ETHR3	1	-0.8167	0.1163	49.2724	0.0000
•Hispanic	ETHR4	1	-0.2010	0.0914	4.8375	0.0278
•Interaction Between Sex and Mood	SEX*MOOD	1	-0.4680	0.2213	4.4709	0.0345
*During past 30 days						
**During past 12 months						
***Anytime during life						

Table 6. Model Fit Statistics for Preliminary and Final Logistic Regression Models

Preliminary Model				**Final Model**			
Criterion	**Intercept**		**Covariates**	**Criterion**	**Intercept**		**Covariates**
AIC	4511.658		3557.119	AIC	4511.658		3552.112
SC	4518.072		3621.260	SC	4518.072		3622.667
-2 Log L	4509.658		3537.119	-2 Log L	4509.658		3530.112
Test	**DF**	**χ^2**	**P-Val**	**Test**	**DF**	**χ^2**	**P-Val**
Likelihood Ratio	10	972.539	<.0001	Likelihood Ratio	10	979.546	<.0001
Score	10	976.578	<.0001	Score	10	977.514	<.0001
Wald	10	1108.664	<.0001	Wald	10	1078.858	<.0001
Association of Predicted Probabilities				**Association of Predicted Probabilities**			
%Concordant	79.8	Somers' D	0.607	%Concordant	79.9	Somers' D	0.610
% Discordant	19.1	Gamma	0.614	%Discordant	18.9	Gamma	0.617
% Tied	1.1	Tau-a	0.208	%Tied	1.2	Tau-a	0.209
Pairs	3482269	c	0.803	Pairs	3482269	c	0.805

Goodness of Fit

Once a model has been established, it is important to determine if the model is a good fit of the data, a good predictor for the response variable, and to determine if the overall terms of the model are relevant in representing the predicted outcomes. Many tests are available to assess goodness of fit for a logistic regression model derived from a simple random survey; however, with complex sample surveys the methods of obtaining many of these tests have not been implemented into any current software packages, such as SAS, SUDAAN or STATA[12]. By using the complex sample survey data to find a model with consideration taken for strata, psu's and weights, the analysis is considered design-based, whereas evaluating the data as if it came from a simple random sample by ignoring the cluster stratification would be considered model-based[12]. The current evaluation techniques for complex sample surveys is to use the design-based method in order to select

a significant model using the methods previously discussed; however, determining influential and outlier points for complex sample surveys requires further study to determine the best approach for identifying these covariate patterns. Current literature suggests to use a model-based approach for model fit using the model found by the design-based methods. The tests we will use to assess the fit of this logistic regression model will be the Hosmer-Lemeshow (HL) test, the Wald F test and the adjusted Wald F test[12]. The sensitivity and specificity can be computed as well; however, this is also accomplished by using the model-based method for obtaining these results. The issue of finding influential or outlier points with the $\Delta\chi^2, \Delta\hat{\beta},$ or ΔD diagnostics are unavailable as options in the current software programs used for analysis of complex sample survey designs, and continued research and simulation studies should be conducted to find suitable methods for obtaining these diagnostic tools[12].

The HL Goodness of Fit test is a diagnostic tool to assess the overall fit of the model . The null hypothesis being tested with the HL test is that the model is a good fit. The first step in computing the HL test is to divide the number of covariate patterns into ten approximately equal groups after sorting the covariate patterns in ascending order of $\hat{\pi}$. Once each group has been established, the average estimated probability for each group is computed, as well as the number of positive responses in each group. The HL test is then computed by the formula:

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

where n'_k is the total number of subjects in the kth group, o_k is the total number of positive responses in the group, and $\bar{\pi}_k$ is the expected probability for group k. Once the $\hat{C}$ statistic has been obtained, the distribution of the statistic can be approximated by the χ^2 distribution with g – 2 degrees of freedom. Table 7 shows the ten HL groups, and the distribution of observed versus expected values for SUID = 0 and SUID = 1. Based on the results of the HL test, χ^2 = 6.7116, and p-value = 0.5680, we conclude that the null hypothesis can not be rejected, and therefore the model is a good fit.

Table 7. Hosmer-Lemeshow Goodness of Fit Tests

		SUID = 1		**SUID = 0**	
Group	Total	y_j	π_j	y_j	π_j
1	449	13	14.65	436	434.35
2	447	18	23.35	429	423.65
3	480	35	31.00	445	449.00
4	458	44	38.02	414	419.98
5	393	48	43.50	345	349.50
6	451	63	61.37	388	389.63
7	451	96	110.19	355	340.81
8	450	170	161.16	280	288.84
9	454	198	200.04	256	253.96
10	477	304	305.72	173	171.28
HL Tests:		χ^2 = 6.7116 p-value = 0.5680		y_j - Observed π_j - Expected	

The Wald F test is another measure for testing coefficient significance and for comparing the significance of the overall model with an intercept only model. The results for the Wald F test are shown in Table 8, and based on the p-values, we are able to conclude that the overall model is significant as well as the estimated coefficients of the full

model. The model with the intercept only is also significant, but since the Wald score is much lower, this indicates that the overall model is more significant.

Table 8. Wald F Test for Coefficient Significance

Contrast	*DF*	*Wald F*	*p-value*	*Adj WaldF*	*p-value*
Overall Model	11	121.73	0.0000	92.04	0.0000
Model - Intercept	10	79.34	0.0000	61.92	0.0000
Intercept	1	30.34	0.0000	30.34	0.0000
AGE	1	7.38	0.0096	7.38	0.0096
SEX	1	16.72	0.0002	16.72	0.0002
MOOD	1	173.20	0.0000	173.20	0.0000
SNF	1	10.69	0.0022	10.69	0.0022
STR	1	8.20	0.0066	8.20	0.0066
IDA	1	3.77	0.0590	3.77	0.0590
VMT	1	33.63	0.0000	33.63	0.0000
ETHR3	1	48.81	0.0000	48.81	0.0000
ETHR4	1	4.67	0.0365	4.67	0.0365
SEX*MOOD	1	4.99	0.0311	4.99	0.0311

Sensitivity and specificity are two diagnostic measurements used to assess the model in terms of reliability of the model to classify observations correctly. Sensitivity measures the number of observations that are correctly classified as true positives, or in the case of this study, those students who answered 'Yes' to the suicidal ideation question would be classified as 'at risk' based on the model. Specificity is the measure of how many subjects were classified as negatives, and responded negatively as well, so students who answered 'No' for the suicidal ideation question would be classified as 'not at risk'. Classifying observations as correct or incorrect requires setting a cutoff point for predicted probabilities to be considered either at risk or not at risk. The typical cutoff point is at $\hat{\pi} = 0.5$, and so for each $\hat{\pi} \geq 0.5$, it would be classified as 'at risk' whereas for each

$\hat{\pi} < 0.5$, it would be considered 'not at risk'. Using a different cutoff point such as $\hat{\pi} = 0.35$, for each observation where $\hat{\pi} \leq 0.35$, then $\hat{\pi} = 0.35$, and observations where $\hat{\pi} < 0.5$, then $\hat{\pi} = 0.65$. For the final model of the YRBSS data, Table 9 lists the cutoff points, sensitivity and specificity for the expected and observed values for the YRBSS data.

Table 9. Classification Table for Sensitivity and Specificity

Cutoff Point	Correctly Identified		Incorrectly Identified		Sensitivity	Specificity	False Negatives	False Positives
	SUID=1	SUID=0	SUID=1	SUID=0				
0.05	968	591	2930	21	97.90	16.80	75.20	3.40
0.10	854	1957	1564	135	86.30	55.60	64.70	6.50
0.15	770	2431	1090	219	77.90	69.00	58.60	8.30
0.20	754	2544	977	235	76.20	72.30	56.40	8.50
0.25	687	2761	760	302	69.50	78.40	52.50	9.90
0.30	657	2833	688	332	66.40	80.50	51.20	10.50
0.35	601	2938	583	388	60.80	83.40	49.20	11.70
0.40	470	3149	372	519	47.50	89.40	44.20	14.10
0.45	366	3281	240	623	37.00	93.20	39.60	16.00
0.50	290	3356	165	699	29.30	95.30	36.30	17.20
0.55	246	3399	122	743	24.90	96.50	33.20	17.90
0.60	182	3434	87	807	18.40	97.50	32.30	19.00

For a continuous range of cutpoints, the sensitivity and specificity can be calculated and plotted as Sensitivity vs. 1-Specificity, which results in a Receiver Operating Characteristic (ROC) curve[12]. The area under the ROC curve provides a measure of discrimination for the model. Figure 3 plots the sensitivity/specificity versus all cutoff points, and Figure 4 gives the ROC curve for the YRBSS data. In Figure 3, the intersection of the two curves would be considered the optimum cutoff level for the sensitivity and specificity measures. The area under the ROC curve is also calculated in SAS under the Association of Predicted Probabilities information, listed as *c*. The YRBSS suicidal

ideation model has an area under the ROC curve of $c = 0.805$, which is considered to be an excellent discrimination level.

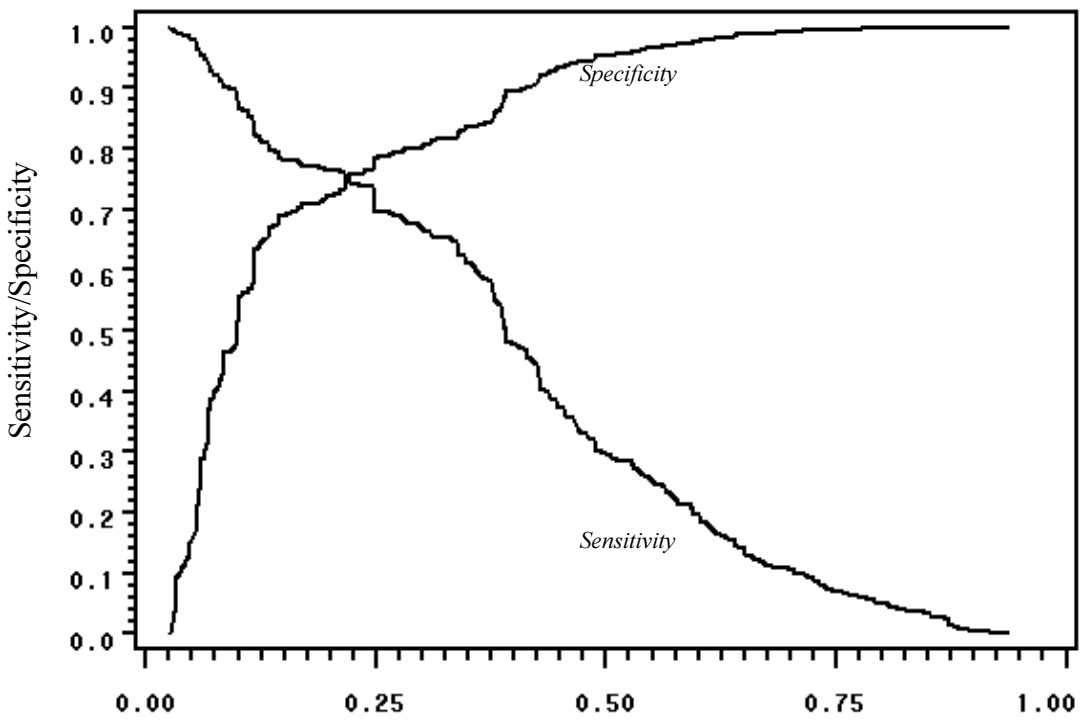

Figure 3: Sensitivity Versus 1–Specificity Plotted Against $\hat{\pi}$

In evaluating goodness of fit for the YRBSS model, model fit statistics obtained have given sufficient evidence supporting that the model is a good representation for determining the risk behaviors associated with suicidal ideation for adolescents in the YRBSS study. We now can expand the model to interpret and make predictions for suicidal ideations for adolescents based on the estimates obtained from the logistic regression model.

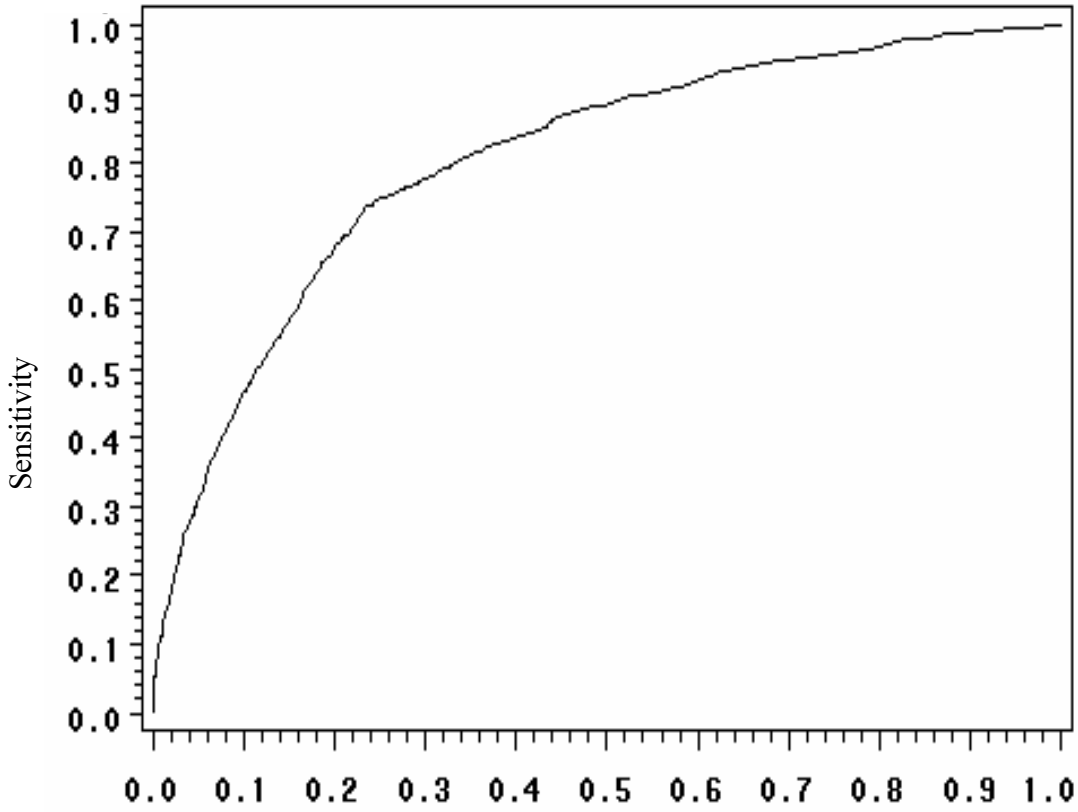

Figure 4: Receiver Operating Characteristic Curve (ROC)

Chapter 3

RESULTS

Odds Ratios

Once a logistic regression model has been found, the odds ratios for the model are looked at to determine the impact of the independent variables on the response variable, in this case, suicidal ideation. The most significant factor in assessing an adolescent's risk of suicidal ideation, based on the obtained model, is the student feeling sad or depressed for more than 2 weeks at a time in the previous 12 months. The estimated odds for an individual to consider suicide is 5.580 as compared to the general sample population, with a 95% confidence interval of 4.392 to 7.090. Other factors that increase the odds of suicidal ideation are using glue, aerosols or other inhalants to get high, with odds ratio of 1.682, using steroids without a doctor's consent with an odds ratio of 1.688, drug or alcohol use prior to having sexual intercourse with odds ratio of 1.333, and the use of laxatives or inducing vomiting to lose weight or to keep from gaining weight with an odds ratio of 2.330. In terms of race and ethnicity, Black and African Americans have an odds ratio of 0.525, indicating a lower risk of suicidal ideation for that particular racial group. Table 10 shows odds ratios and 95% confidence intervals for these risk factors of suicidal ideation.

Calculation of odds ratios for risk factors that interact with each other, such as sex and mood, must be computed in a different manner than those of risk factors not involved in interactions. For a risk factor that is not affected by interaction, the odds ratio is

calculated fairly quickly using the equation $OR = e^{\hat{\beta}}$. With interaction, the logit is calculated as:

$$g(f, x) = \beta_0 + \beta_1 f + \beta_2 x + \beta_3 f \times x,$$

where f is the level of factor 1, x is the level of factor 2, and $f \times x$ is the product of the two levels. To compare the odds ratios for one of the factors, in our case sex, the odds ratio would be calculated by taking the difference of the logit between females and males, and exponentiating the difference. Thus, to compute the odds ratio for the interaction of sex and mood for females versus males where f_1 is female and f_0 is male, the equation yields:

$$OR = \exp[0.5991(f_1 - f_0) - 0.4680x(f_1 - f_0)].$$

Table 10. Estimated Odds Ratios and 95% Confidence Intervals

Risk Factor	***OR's**	**95% CI**	
Race/Ethnicity			
Black or African American	0.525	0.388	0.710
Hispanic	0.992	0.789	1.246
Females	1.820	1.370	2.420
Feeling Sad or Depressed for More Than 2 Weeks at a Time*	5.580	4.392	7.090
Using Glue, Aerosols, or Other Inhalants to Get High*	1.682	1.145	2.472
Using Steroids Without a Doctor's Consent*	1.688	0.945	3.015
Using Drugs and/or Alcohol Before Last Sexual Intercourse*	1.333	0.893	1.990
Vomiting or Using Laxatives to Lose Weight*	2.330	1.514	3.585
Interaction between Sex and Depressed Mood	0.626	0.406	0.966

Based on the results from Table 11 showing the odds ratios for the interaction between sex and mood, it is clear to see that females have a higher rate of suicidal ideation both in terms of having a depressed mood, with an odds ratio of 1.681 for females versus 1.474 for males, and in absence of a depressed mood with females having an odds ratio of

0.297 versus males at 0.163. It is also very important to note that a depressed mood for 2 weeks or more at a time significantly increases the risk of suicidal ideation for both males and females.

Table 11. Odds Ratios For Interaction Between Sex & Depressed Mood

	Yes			No		
	OR	95% CI		OR	95% CI	
Female	1.681	1.635	1.727	0.297	0.285	0.310
Male	1.474	1.434	1.515	0.163	0.157	0.170

Many other factors affect the risks of suicidal ideation for males and females at different rates. In order to determine some of the risk factors that are more significant for a particular sex, odds ratios have been calculated for males and females separately to determine which risk factors are more significant in terms of sex. In terms of race and ethnicity, the odds ratios for Black and African Americans and Hispanics were lower for both male and female for suicidal ideation, however females had a higher odds ratio for each of theses groups. Black and African American females have an odds ratio of 0.525, whereas males have an odds ratio of 0.304. For Hispanics, the odds ratio for females is 0.992, and males again have a lower risk with odds ratios of 0.607. Use of inhalants, aerosol sprays and glue to get high has a larger effect on males than females, where males have an odds ratio of 2.101 and the odds ratio for females is 1.682. There is a difference between sexes for the other risk factors; however the differences are not as significant for these other variables. Table 12 gives the estimated odds ratios and 95% confidence limits for all of the variables in the final model by sex.

Table 12. Estimated Odds Ratios For Males and Females

Risk Factor	Females			Males		
	*OR's	95% CI		*OR's	95% CI	
Race/Ethnicity						
Black or African American	0.525	0.388	0.710	0.304	0.199	0.464
Hispanic	0.992	0.789	1.246	0.607	0.446	0.825
Age						
≤ 14	0.852	0.758	0.957	0.869	0.731	1.033
15	0.725	0.646	0.815	0.755	0.635	0.898
16	0.618	0.550	0.694	0.656	0.552	0.781
17	0.526	0.469	0.591	0.571	0.480	0.678
Using Glue, Aerosols, or Other Inhalants to Get High	1.682	1.145	2.472	2.101	1.209	3.648
Using Steroids Without a Doctor's Consent	1.688	0.945	3.015	1.930	1.151	3.234
Using Drugs and/or Alcohol Before Last Sexual Intercourse	1.333	0.893	1.990	1.203	0.818	1.770
Vomiting or Using Laxatives to Lose Weight	2.330	1.514	3.585	2.603	1.782	3.802

Chapter 4

DISCUSSION

The rates of adolescents attempting suicide have remained fairly steady since the 1990's[9]. While suicide ideation showed a decrease in the 1990's[7], the rates have not changed in recent years, as shown in Table 13. Many studies have shown the link between suicidal ideation and suicidal attempts[2,11]. In order to try and decrease the suicide rate of adolescents, prevention needs to start early[6] when risk behaviors begin[2]. If even one death can be prevented based on prevention methods used when risk behaviors are first evaluated, the efforts of this study, and others like it, will be worth the life of a young person just starting out in life.

Although this study has identified ten significant risk factors for predicting suicidal ideation, many other risk behaviors for suicidal ideation and attempts may be present, and further research into the risk factors is imperative to assure that youth are able to face the challenges of adolescence without having to think in terms of extreme measures, such as suicidal ideation, which in turn can lead to attempts and/or success of suicide[4,5].

Table 13. Trends in the Prevalence Rate of Suicidal Ideation and Attempts[9]

1991	1993	1995	1997	1999	2001	2003	2005	Changes from 1991 – 2005[1]	Change from 2003 – 2005[2]
Seriously considered attempting suicide (During the 12 months preceding the survey.)									
29.0 (±1.5)[3]	24.1 (±0.9)	24.1 (±1.1)	20.5 (±2.3)	19.3 (±1.2)	19.0 (±1.4)	16.9 (±0.7)	16.9 (±0.9)	Decreased, 1991 – 2003 No change, 2003 – 2005	No change
Made a suicide plan (During the 12 months preceding the survey.)									
18.6 (±1.5)	19.0 (±1.1)	17.7 (±1.4)	15.7 (±1.3)	14.5 (±1.4)	14.8 (±1.1)	16.5 (±3.5)	13.0 (±0.9)	Decreased, 1991 – 2005	No change
Attempted suicide (One or more times during the 12 months preceding the survey.)									
7.3 (±0.9)	8.6 (±0.8)	8.7 (±0.8)	7.7 (±0.9)	8.3 (±1.0)	8.8 (±0.8)	8.5 (±1.1)	8.4 (±0.9)	No change, 1991 – 2005	No change

[1] Based on linear and quadratic trend analyses using a logistic regression model controlling for sex, race/ethnicity, and grade.
[2] Based on t-test analyses.
[3] 95% confidence interval.

Limitations

The limitations of this study include issues such as limited availability of influential and outlier diagnostics, reliability of self-reporting from a survey, non-representation of subjects successful at suicide, and other questions not addressed in the survey questionnaire. Further study should be done to investigate methods that are able to more reliably obtain diagnostic tests for outlier and influential subjects. Another limitation arising from the complex sample survey data is that for model fit analysis, observations with any missing data are not included for model fit tests. Methods such as imputation can be implemented to help account for the missing values of an observation for complex sample surveys, and increase reliability of model fit assessment. Limitations are present with any sample survey in that although careful design of the sample population is implemented, certain populations may still be

underrepresented by the sample population, especially in considering risk factors for suicidal attempts when information would not be available for those succeeding in suicidal attempts[3]. Students at the age of 16 are no longer required by law to attend school, and the risks associated with suicidal ideation could be very different for this population, but due to the sample design of the YRBSS the sample population is unable to account for dropouts since the survey is conducted through school attendance.

Other studies have shown links between suicidal ideation and risk factors such as previous suicide attempts[2], homosexuality and sexual identity issues[2,4], and also biological and genetic markers[4,7]. The YRBSS survey has not implemented questions dealing with sexual identity issues, and survey data can not account for biological risk factors or genetic markers either. Therefore, additional studies focusing on these factors could help in determining the rates of risk associated with these particular variables.

Continued research in the risk factors associated with suicidal ideation identified in this study, as well as risk factors mentioned previously should be investigated further to address the limitations of this study. The more information on risk behaviors that influence suicidal ideation can help local, state and other agencies to create and fund additional prevention and treatment plans targeting the risk behaviors and to help prevent an adolescent from engaging in risky behaviors that could lead to suicidal ideations, and potentially prevent future suicidal behaviors in adolescents.

REFERENCES

1. Centers for Disease Control and Prevention. (2005) Youth Risk Behavior Surveillance Survey. Available at: www.cdc.gov/yrbss. Accessed on 5/24/2006.

2. Gould, M., Greenberg, T., Velting, DM., Shaffer, D., (2003) Youth suicide risk and preventative interventions: a review of the past 10 years. *Journal of American Academy of Child Adolescent Psychiatry* 42:4, 386:404.

3. Carter, GL., Issakidis, C., Clover, K., (2003) Correlates of youth suicide attempters in Australian community and clinical samples. *Australian and New Zealand Journal of Psychiatry* 37, 286:293.

4. Koplin, B., Agathen, J., (2002) Suicidality in children and adolescents: a review. *Current Opinion in Pediatrics* 14, 713:717.

5. King, RA., Schwab-Stone, M., Flisher, AJ., Greenwald, S., Kramer, RA., Goodman, SH., Lahey, BB., Shaffer, D., Gould, MS., (2001) Psychosocial risk behavior correlates of youth suicide attempts an suicidal ideation. *Journal of American Academy of Child Adolescent Psychiatry* 40:7, 837:846.

6. Foster, T., (2001) Dying for a drink. *BMJ* 323, 817:818 .

7. Gray, D., Achilles, J., Keller, T., Tate, D., Haggard, L., Rolfs, R., Cazier, C., Workman, J., McMahon, WM., (2002) Utah youth suicide study, phase I: government agency contact before death. *Journal of American Academy of Child Adolescent Psychiatry* 41:4, 427:433.

8. Beautrais, AL., (2003) Suicide and serious suicide attempts in youth: a multiple-group comparison study. *Am J Psychiatry* 160:6, 1093:1099.

9. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS) (2005) cited 2006, November 9. Available from URL: www.cdc.gov/ncipc/wisqars.

10. Beck, AT., Steer, RA., Kovacs, M., Garrison, B., (1985) Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry* 142 (5): 559–563.

11. Cerel, J., Roberts, TA., Nilsen, WJ., (2005) Peer suicidal behavior and adolescent risk behavior. *The Journal of Nervous and Mental Disease* 193:4, 237:243.

12. Hosmer, DW., Lemeshow, S., (2000) *Applied Logistic Regression* 2nd edition.

APPENDIX

Appendix A – SAS Code For YRBSS Analysis

```
options nodate nonumber nocenter skip=0;
libname adol 'C:\YRBSS\SAS Code\';
libname library 'C:\YRBSS\SAS Code';
data adol.rcyrbs05;
set adol.yrbs05SAS;
  If q24 =. then delete;
  IF q2=2 THEN sex=0; else sex=q2;
  IF q1='1' THEN age=1; IF q1='2' THEN age=2; IF q1='3' THEN age=3; IF q1='4' THEN age=4;
  IF q1='5' THEN age=5; IF q1='6' THEN age=6; IF q1='7' THEN age=7; else age=q1;
IF age>=1 or age<=4 THEN age1=1; ELSE age1=0;IF age>=5 or age<=7 THEN age2=1;else age2=0;
  IF q4='1' THEN race1=1;else  race1=0; IF q4='2' or q4='5' THEN race2=1; else  race2=0;
  IF q4='3' THEN race3=1;else  race3=0; IF q4='4' or q4='7' THEN race4=1; else  race4=0;
  IF q4='6' THEN race5=1;     else  race5=0;
        ARRAY qnr(97) qnr1-qnr97;
        ARRAY qn(97) qn1-qn97;
                DO i = 1 TO 97;
        If qn(i)=2 THEN qnr(i)=0;ELSE qnr(i)=qn(i); END;
  If qnfrcig=2 then qnrfrcig=0; else qnrfrcig=qnfrcig;
  If qnanytob=2 then qnranytob=0; else qnranytob=qnanytob;
  drop q1-q97 qn1-qn97;    run;
proc surveyfreq data=adol.rcyrbs05;
    cluster psu;
    strata stratum;
    tables age sex qnr23 race1-race5 qnr24;
    weight weight; run;
/*Create Interaction Variables*/
data adol.rcyrbs05; set adol.rcyrbs05;
  q70r4=qnr70*race4;  q70r3=qnr70*race3;  ageq50=age*qnr50;   ageq23=age*qnr23;
  ageq54=age*qnr54;   ageq70=age*qnr70;   ageq61=age*qnr61;   sxq54=sex*qnr54;
  sxq50=sex*qnr50; sxq23=sex*qnr23; sxq61=sex*qnr61; sxq70=sex*qnr70;
  q23q61=qnr23*qnr61; q23q54=qnr23*qnr54; q23q50=qnr23*qnr50; q23q70=qnr23*qnr70;
  q23r3=qnr23*race3;  q23r4=qnr23*race4;  q50q70=qnr50*qnr70; q50q61=qnr50*qnr61;
  q50q54=qnr50*qnr54; q50r3=qnr50*race3;  q50r4=qnr50*race4;  q54r3=qnr54*race3;
  q54q70=qnr54*qnr70; q54q61=qnr54*qnr61; q54r4=qnr54*race4;  q61r4=qnr61*race4;
  q61r3=qnr61*race3;  q61q70=qnr61*qnr70; age1q23=age1*qnr23; age2q23=age2*qnr23;
  age1q54=age1*qnr54; age2q54=age2*qnr54; age1q70=age1*qnr70; age2q70=age2*qnr70;
  age1q61=age1*qnr61; age2q61=age2*qnr61;
run;
/*Variable significance testing for Suicide Ideation*/
title 'Variable significance testing for Suicidal Ideation'; ods listing close;
%macro repeat(covlist);
  ods output ParameterEstimates=logitout;
  proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24 (descending)=&covlist;
        weight weight;
        strata stratum;
  data sum;
  set sum logitout;
%mend;
data sum;
if 1=1 then delete; run;
```

```
%repeat(race1)%repeat(race2)%repeat(race3)%repeat(race4)%repeat(qnr70)%repeat(age)
%repeat(race5)%repeat(sex)%repeat(qnr21)%repeat(qnr23)%repeat(qnr90)%repeat(qnr40)
%repeat(qnr42)%repeat(qnr44)%repeat(qnr48)%repeat(qnr49)%repeat(qnrfrcig)%repeat(qnr18)
%repeat(qnr50)%repeat(qnr51)%repeat(qnr52)%repeat(qnr53)%repeat(qnranytob)%repeat(qnr16)
%repeat(qnr54)%repeat(qnr55)%repeat(qnr61)%repeat(qnr68)%repeat(qnr7)%repeat(qnr15)
%repeat(qnr11)%repeat(qnr13)%repeat(qnr84)%repeat(qnr88)%repeat(qnr59)%repeat(qnr14)
ods listing;
proc print data=sum; run;
/*Backwards elimination for Suicidal Ideation - FULL MODEL */
ods listing close;
ods output Surveylogistic.ParameterEstimates=full;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr7 qnr11 qnr13-qnr16 qnr18 qnr21 qnr23 qnr40
qnr42
        qnr44 qnr48-qnr55 qnr59 qnr61 qnr68 qnr70 qnr84 qnr88 qnr90 qnranytob qnrfrcig
race1-race4;
        weight weight;
        strata stratum; run;
ods listing;
proc print data=full; run;
/*Backwards elimination for Suicidal Ideations*/
ods listing close;
ods output Surveylogistic.ParameterEstimates=bkwd1;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr48-qnr55 qnr61 qnr70 qnr90 race1 race3
race4;
        weight weight;
        strata stratum; run;
ods listing;
proc print data=bkwd1; run;
ods listing close;
ods output Surveylogistic.ParameterEstimates=bkwd2;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr49 qnr50 qnr54 qnr61 qnr70 race1 race3
race4;
        weight weight;
        strata stratum; run;
ods listing;
proc print data=bkwd2; run;
ods listing close;
ods output Surveylogistic.ParameterEstimates=bkwd3;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race1 race3 race4;
        weight weight;
        strata stratum; un;
ods listing;
proc print data=bkwd3; run;
ods listing close;
ods output Surveylogistic.ParameterEstimates=bkwd4;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4;
        weight weight;
        strata stratum; run;
ods listing;
proc print data=bkwd4; run;
```

```
/*Check Age for transformation*/
data trans;
  set adol.rcyrbs05;
  p1=-2;p2=-1;p3=-0.5;p4=0;p5=0.5;p6=1;p7=2;p8=3;
%macro fractional;
  %do n=1 %to 8;
  %do m=&n %to 8;
        data trans&n&m;
        set trans;
        if p&n=0 then u4=log(age);else u4=age**p&n;
        if p&m=p&n then u5=log(age)*u4; else if p&m=0 then u5=log(age);
        else u5=age**p&m;
proc surveylogistic data=trans&n&m;
        cluster psu; model qnr24 (descending) =u4 u5;
        weight weight;
        strata stratum; run;
%end; %end;
%mend fractional;
%fractional;
data j1model;
  set trans;
  age_2=1/(age**2);
  age_1=1/age;
  age_5=1/sqrt(age);
  ageln=log(age);
  agesq=sqrt(age);
  age2=age**2;
  age3=age**3; run;
ods listing close;
%macro repeat(covlist);
  ods output FitStatistics=logitout;
  proc surveylogistic data=j1model;
        class &covlist;
        cluster psu;
        model qnr24 (descending)=&covlist;
        weight weight;
        strata stratum;
data sum;
  set sum logitout;
%mend;
data sum;
  if 1=1 then delete; run;
%repeat(age_2)%repeat(age_1)%repeat(age_5)%repeat(ageln)%repeat(agesq)%repeat(age)
%repeat(age2)%repeat(age3)
ods listing;
proc print data=sum; run;
ods select FitStatistics;
ods listing close;
ods output FitStatistics=fits;
proc surveylogistic data=trans;
        cluster psu;
        model qnr24 (descending) = age;
        weight weight;
        strata stratum; run;
ods listing;
proc print data=fits; run;
/*No Transformation Needed for Age*/
/*Parameter estimates for Interaction terms ~ QN24*/
%macro repeat(covlist);
  ods output ParameterEstimates=logitout
```

```
  proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)= &covlist age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3
race4;
        weight weight;
        strata stratum;
data sum;
  set sum logitout;
%mend;
data sum;
  if 1=1 then delete; run;
%repeat(ageq50)%repeat(ageq23)%repeat(ageq54)%repeat(ageq70)%repeat(ageq61)%repeat(sxq54)
%repeat(sxq50)%repeat(sxq23)%repeat(sxq61)%repeat(sxq70)%repeat(q23q61)%repeat(q23q54)
%repeat(q23q50)%repeat(q23q70)%repeat(q23r3)%repeat(q23r4)%repeat(q50q70)%repeat(q50q61)
%repeat(q50q54)%repeat(q50r3)%repeat(q50r4)%repeat(q54r3)%repeat(q54q70)%repeat(q54q61)
%repeat(q54r4)%repeat(q61r4)%repeat(q61r3)%repeat(q61q70)%repeat(q70r4)%repeat(q70r3)
ods listing;
proc print data=sum; run;
/*Parameter estimates for Interaction terms (Using Adjusted Age Groups) ~ QN24*/
%macro repeat(covlist);
ods output ParameterEstimates=logitout;
  proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)= &covlist age1 age2 sex qnr23 qnr50 qnr54 qnr61 qnr70
race3 race4;
        weight weight;
        strata stratum;
data sum;
  set sum logitout;
%mend;
data sum;
  if 1=1 then delete; run;
%repeat(age1q23) %repeat(age2q23) %repeat(age1q54) %repeat(age2q54)
%repeat(age1q70) %repeat(age2q70) %repeat(age1q61) %repeat(age2q61)
ods listing;
proc print data=sum; run;
/*Parameter estimates for Interaction term - SXQ23 ~ QN24*/
proc surveylogistic data=adol.rcyrbs05
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr50 qnr54 qnr61 qnr70
        race3 race4 ageq70 sxq23;
        weight weight;
        strata stratum; run;
proc surveylogistic data=adol.rcyrbs05;
        cluster psu;
        model qnr24(descending)=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
        weight weight;
        strata stratum; run;
proc surveyfreq data=adol.rcyrbs05;
  cluster psu;
  strata stratum;
  tables  (sex age age1 age2 qnr23 qnr50 qnr54 qnr61 qnr70
  race1 race2 race3 race4 race5 other) * qnr24 / EXPECTED;
  weight weight; run;
ods listing close;
proc surveylogistic data=adol.rcyrbs05;
ods output OddsRatios=ageodds;
  cluster psu;
  model qnr24(descending)=age1 sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  weight weight;
```

```
  strata stratum; run;
ods listing;
proc print data=ageodds; run;
/*Goodenss of Fit Data  PROC LOGISTIC USED FOR MODEL CHECK GRAPHS DUE TO ABSENCE OF
COMPARABLE DIAGNOSTICS AVAILABLE FOR COMPLEX SAMPLE SURVEYS.*/
  title ' ';
proc logistic data=adol.rcyrbs05 descending;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23/ ctable
  outroc=fit details pprob = (.05 to .6 by .05) ;
  weight weight;run;
proc gplot data=fit;
plot _sensit_*_1mspec_;
run; quit;
proc logistic data=adol.rcyrbs05 desc;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23
  / outroc=roc1;
weight weight;run;
data roc2;
  set roc1;
  spec = 1-_1mspec_;run;
symbol1 i=join v=none ;
proc gplot data=roc2;
  plot _sensit_*_PROB_=1 spec*_PROB_=1 / overlay haxis=0 to 1 by .25 vaxis=0 to 1 by .1 ;
run;
quit;
title 'Final Model Ready for Model Checking';
proc surveylogistic data=adol.rcyrbs05;
  cluster psu;
  model qnr24(descending)=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  weight weight;
  strata stratum; run;
```

Appendix B – SUDAAN Code for YRBSS Analysis

```
PROC SORT DATA=adol.rcyrbs05;
       By stratum psu;
proc rlogist data=adol.rcyrbs05 FILETYPE=SAS design = wr;
  NEST stratum psu;
  WEIGHT weight;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  OUTPUT / FILENAME=mdlchk1 REPLACE FILETYPE=SAS BETAS=ALL; run;
proc print data=mdlchk1;
run;
proc rlogist data=adol.rcyrbs05 FILETYPE=SAS design = wr;
  NEST stratum psu;
  WEIGHT weight;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  OUTPUT / FILENAME=mdlchk2 REPLACE FILETYPE=SAS TESTS=ALL;
run;
proc print data=mdlchk2;
run;
proc rlogist data=adol.rcyrbs05 FILETYPE=SAS design = wr;
  NEST stratum psu;
  WEIGHT weight;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  OUTPUT / FILENAME=mdlchk3 REPLACE FILETYPE=SAS PREDICTED=ALL;
run;
proc print data=mdlchk3;
```

```
run;
proc rlogist data=adol.rcyrbs05 FILETYPE=SAS design = wr;
  NEST stratum psu;
  WEIGHT weight;
  model qnr24=age sex qnr23 qnr50 qnr54 qnr61 qnr70 race3 race4 sxq23;
  OUTPUT HLCHISQ HLCHIP / FILENAME=mdlchk4 REPLACE FILETYPE=SAS HLTEST=ALL;
run;
```