IMPLEMENTATION OF THE HARDY-WEINBERG TEST FOR EQUILIBRIUM IN A STUDY EXAMINING THE RELATIONSHIP BETWEEN THE DOPAMINE TRANSPORTER GENE (SCL6A3) AND SMOKING CESSATION IN WOMEN

by

MaGuadalupe Leon-Verdin

B.S., University of Guanajuato, 2001

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

MaGuadalupe Leon-Verdin

It was defended on

April 13, 2007

and approved by

Thesis Advisor:
Lisa Weissfeld, PhD
Professor and Director of Academic Program
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Michele D. Levine, PhD.
Assistant Professor of Psychiatry and Psychology
Department of Psychiatry and Psychology
University of Pittsburgh Medical Center

Committee Member:
Lan Kong, PhD.
Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Lisa Weissfeld, PhD

IMPLEMENTATION OF THE HARDY-WEINBERG TEST FOR EQUILIBRIUM IN A STUDY EXAMINING THE RELATIONSHIP BETWEEN THE DOPAMINE TRANSPORTER GENE (SCL6A3) AND SMOKING CESSATION IN WOMEN

MaGuadalupe Leon-Verdin, M.S.

University of Pittsburgh, 2007

Abstract

The dopamine transporter gene is an untranslated polymorphic region, which consists of a replication of 40-base pairs. The locus of this 3' variable number tandem repeat (VNTR) polymorphism is on 5p15.3 and repeats from 3 to 13 times. In most populations, the most common alleles are 9 and 10. The distribution of the dopamine transporter genotypes also varies among races.

These genotypes have been shown to be associated with different conditions of health such as smoking status, obesity and food intake. Nine-carriers have been associated with late initiation of smoking. Homozygous for SCL6A3 – 10 are related to have higher concentration of dopamine transporter protein and to have lower postsynaptic concentration of dopamine. In this paper, the role of the SCL6A3 genotypes and allele carriers are investigated in a sample of women smokers willing to quit smoking who are concerned with postcessation weight gain. Because of the small number of women carrying alleles other than 9 or 10 allele, the sample size

was limited to three genotypes. The main purpose of this work is to test departure from Hardy-Weinberg Equilibrium.

The result shows that the proportions of the genotype were p^2, $2pq$, and q^2; therefore, this gene is in Hardy-Weinberg Equilibrium. The genotype proportions in Caucasian women were similar to those previously reported in European-Caucasian women, and the proportions in African-American women were similar to previously reported literature values among African-American. These findings could have public health relevance in smoking cessation programs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to acknowledge at Consejo Nacional de Ciencia y Tecnología (CONACYT) for its financial support in Mexico. Many thanks to Dr. Michele D. Levine for allowing me to use her data for this work and being a part of my committee. Also, thank to Dr. Lan Kong for being part of my thesis committee.

I would like to thank Dr. Lisa Weissfeld for her support and for her help to get this process ready to complete my degree.

Also, I would like to express my gratitude to Dr. John W. Wilson and Dr. Roslyn Stone for their kindness and their valuable suggestions throughout my graduate studies.

Finally, I would like to thank my family, especially my husband, Leonardo Álvarez Valtierra, for his support and encouragement in my graduate studies.

1.0 INTRODUCTION

Gaining weight is a common concern for women who participate in smoking cessation programs. One of the therapies used to aid smoking cessation is ***nicotine replacement therapy*** (Haxby *et al.* 1995; Flore *et al.* 2000). It has been shown that such therapy significantly increases smoking cessation rates, but a large percentage of smokers still relapse and resume smoking (Jorenby *et al*. 1999; Sabol *et al*. 1999; Lerman *et al*. 2006). A new approach to aid in smoking cessation is the use of a ***pharmacologic treatment,*** bupropion, which is not nicotine based. Bupropion is the only non-nicotine drug used in smoking cessation programs that is also approved by the Food and Drug Administration. These approaches control the release of neurotransmitters or blunt the addictive effects of nicotine. Bupropion (Zyban) acts as a norepinephrine and dopamine-reuptake inhibitor (Ascher *et al*. 1995) which is why it is effective for smoking cessation.

The genetic variability in the dopamine receptor gene (DRD2, DRD4) and dopamine transporter gene (SLC6A3) could be related to the response to smoking cessation. In 1999, significant evidence of association between smoking behavior and the untranslated region polymorphism (SLC6A3, chromosome 5p15.3 (Vandenbergh *et al*. 1992)) in the dopamine transporter (DAT) was shown in a case control study (Lerman *et al*. 1999). Consistent results were found when comparing former smokers, current smokers, and non smokers in a clinical trial (Sabol *et al*. 1999). On the other hand, this gene has been associated with other behaviors such as

food intake (Epstein *et al*. 2002 and 2004), Parkinson's disease (Kelada *et al*. 2006), cocaine-induced paranoia (Lerman *et al*. 1999) and obesity (Shinohara *et al*. 2004; Need *et al*. 2006).

Women are more likely to gain weight than men after they quit smoking (Perkins *et al*. 2001). Genetic factors, as well as food intake, have been shown to be associated with the increased weight in women and men after smoking cessation. For this study, we analyze the dopamine transporter genotype and focus on the untranslated region which consists of 3 to 13 repeats, but as in most populations, the alleles SCL6A3 – 9 and 10 are most common. The allele frequencies of DAT also vary across race/ethnicity; for instance, in the Chilean population the 10 allele has a frequency of 74%, in American-Indian it has a frequency of 100% (Hutz *et al*. 2000; Vieyra *et al*. 2003), in the Chinese population the frequency is 90% (Leighton *et al*. 1997) in the Korean population it is approximately 90% (Kim *et al*. 2006), and for the Jewish population the frequency is 48% (Mitchell *et al*. 2000; Galili-Weisstub *et al*. 2005).

It was hypothesized that this sample is a representative sample from the population studied and the distribution of the allele frequencies in African-Americans (AA) is not the same as the distribution in Caucasians. The role of this gene is examined for a potential association with smoking and weight gain in women smokers.

1.1 BACKGROUND AND REVIEW

1.1.1 Genetic polymorphism

Genetic variation or multiple alleles exist in most populations at each locus, which cause the differences in individuals. The genetic differences among organisms of the same species are called genetic polymorphisms. There are different kinds of polymorphisms; the simplest one is the variation in a particular defined site known as single nucleotide polymorphism (SNP). It can be either a non-synonymous or synonymous polymorphism. An indel is an insertion or deletion polymorphism, where a fragment of less than 10 base pairs is in or out of the DNA sequence. Among other important polymorphisms, the minisatellite polymorphism, called a 3' untranslated variable number of tandem repeats, depends on the number of replications in the DNA analysis. The dopamine transporter gene is an example of this kind of polymorphism, which has the same sort of repeating structure. Each allele has a different number (*n*) of copies of the repeat, the most common DAT alleles are the SCL6A3 – 9 and SCL6A3 – 10, with the repeat varying from 3 to 13 times.

1.1.2 Allele frequency

The allele frequency is the proportion of the allele of interest divided by the total number of alleles in the population. The allele frequency can be estimated from the analysis of a

representative sample from a population. In diploid organisms, such as humans, each individual contributes with two alleles that can be the same or different. Therefore, the proportion of the allele frequency is the number of alleles present in the sample divided by twice the number of individuals in the sample (N).

$$p_i = \frac{number\ of\ the\ same\ allele_i}{twice\ the\ number\ of\ individuals\ analized}$$

$$\hat{p}_i = \frac{2\,homozigote + heterozigote}{2N}$$

and the estimated sampling variance is

$$\hat{\sigma}^2_{\hat{p}_i} = \frac{(\hat{p}\ )(1 - \hat{p}\,)}{2N}$$

It is possible to calculate the 100(1 – α)% confidence interval by using the normal distribution,

$$\hat{p}_i \pm z_{\alpha/2}(\hat{\sigma}_{\hat{p}_i})$$

where $z_{\alpha/2}$ is the cut off of the $\alpha/2$ area under the normal curve. In some cases where the alleles are rare, the frequencies might be considered negligible. Therefore, the sum of the allele frequencies is equal to one.

1.1.3 Genotype frequency

The genotype frequency is the proportion of individuals with such genotype divided by the total number of individuals in the population.

$$P_j = \frac{num\ of\ individuals\ with\ same\ Genotype_j}{Sum\ of\ all\ individuals\ with\ all\ possible\ Genotypes\ at\ one\ locus}$$

This quantity can be also estimated after a representative random sample is selected from the population of interest. A second condition for an unbiased estimator is that the observations have to be independent and that the individuals are an identical organism. The number of different genotypes depends on the number of alleles at that locus, for instance if there exists only two different alleles, there will be only three possible genotypes.

Another estimate of homozygote genotype frequency (P_j) is the square of the allele frequency ($P_j = p_i^2$), and twice the product of the two allele frequencies in heterozygote genotypes ($P_j = 2p_i p_k$).

2.0 METHOD

2.1 HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg principle or Hardy-Weinberg equilibrium states that, the genotype frequencies at one locus will be fixed at one particular equilibrium value, under the following assumptions:

- random mating
- diploid organisms
- nonoverlapping generations
- sexual reproduction
- large population
- equal allele frequencies in the sexes
- no migration
- no mutation
- no selection

Although some of these assumptions might not hold true, this model still can be used. This model is not sensitive to violation in its assumptions.

The dopamine transporter allele is located in the short arm of chromosome 5 at band 15 and sub-band 3. At this locus a 40 base pair polymorphism has been found in DNA analysis. In

this sample, only the alleles 9 and 10 at this locus were considered. An estimated value of the allele frequency can be calculated from a sample for the population as,

$$\hat{p}_9 = \frac{2(\#SLC6A3_{9-9}) + (\#SLC6A3_{9-10})}{2N}$$

where, $\#SLC6A3_{9-9}$ is the number of individuals concordant for the allele 9, the $\#SLC6A3_{9-10}$ is the number of individuals discordant with one allele 9 and one allele 10, and N is the total sample size.

The complementary allele frequency $\hat{p}_{10}$ can be calculated by either

$$\hat{p}_{10} = 1 - \hat{p}_9$$

or

$$\hat{p}_{10} = \frac{2(\#SLC6A3_{10-10}) + (\#SLC6A3_{9-10})}{2N}$$

where, $\#SLC6A3_{10-10}$ is the number of individuals concordant for the allele 10.

The estimated sample variance of $\hat{p}_9$ equals the sample variance of $\hat{p}_{10}$ in this case and can be estimated by the following formula

$$\hat{Var}(\hat{p}_9) = \hat{Var}(\hat{p}_{10}) = \frac{\hat{p}_9 \hat{p}_{10}}{2N}$$

2.2 HARDY-WEINBERG TEST

According to the Hardy-Weinberg principle the alleles for the next generation are chosen randomly and independently in a large population. With two different alleles, there are three different genotypes (D, H, and R) and six types of matings (DD, 2DH, 2DR, HH, 2HR, and RR), which occur in proportion to the genotypic frequency in the population (P_D, P_H, and P_R), as shown in Tables 1 and 2.

Table 1. Mating type.

	D 9 – 9	H 9 – 10	R 10 – 10
D 9 – 9	DD 9 – 9 x 9 – 9	DH 9 – 9 x 9 – 10	DR 9 – 9 x 10 – 10
H 9 – 10	HD 9 – 10 x 9 – 9	HH 9 – 10 x 9 – 10	HR 9 – 10 x 10 – 10
R 10 – 10	RD 10 – 10 x 9 – 9	RH 10 – 10 x 9 - 10	RR 10 – 10 x 10 – 10

Where DH = HD, DR = RD, and HR = RH

Here, P_D, P_H and P_R are the genotypic frequencies and can be calculated as the probability of having a specific genotype given a sample (N),

$$P_D = P(9-9) = \frac{\#SLC6A3_{9-9}}{N}$$

The mating type frequencies are calculated as the corresponding products of the genotype frequencies. For instance, the frequency for the mating type DD is P_D^2. When the genotypes involved in the mating are different the resulting mating type frequency is twice the product of the genotype frequencies. The frequencies for the six different mating types are shown in Table 2.

Table 2. Mating type frequencies.

Genotype Frequency	P_{D}	P_{H}	P_{R}
P_{D}	P_D^2	$2P_DP_H$	$2P_DP_R$
P_{H}		P_H^2	$2P_HP_R$
P_{R}			P_R^2

The new genotype can be derived using cross-multiplication square (Punnett square) and the genotype frequencies can be calculated using Bayes' rule. The probability that an individual has a 9 – 9 genotype is given by

$$P_D^{'} = P^{'}(9-9) = \sum_{mating\ type} P(mating\ type)P[(9-9)\,|\,mating\ type]$$

$$P^{'}(9-9) = P_D^2 + \frac{1}{2}2P_DP_H + \frac{1}{4}P_H^2$$

$$P'(9-9) = \left(P_D + \frac{1}{2}P_H\right)^2$$

$$P'(9-9) = \hat{p}_9^2$$

and the 9 – 10 and 10 – 10 genotype frequencies in the following generations are given by

$$P_H' = P'(9-10) = \sum_{mating\ type} P(mating\ type) P[(9-10) \mid mating\ type] = 2\hat{p}_9\hat{p}_{10}$$

$$P_R' = P'(10-10) = \sum_{mating\ type} P(mating\ type) P[(10-10) \mid mating\ type] = \hat{p}_{10}^2$$

From the new generation frequencies, it is possible to calculate the new frequencies of $\hat{p}_9^{'}$ and $\hat{p}_{10}^{'}$. Using the last three equations, it is not difficult to show that the frequencies in the following generation are exactly the same as they were in the previous generation. Then, the locus is said to be in HW-equilibrium if it can be shown that the allele frequency remains the same, generation after generation, and the genotype frequencies must satisfy the proportion of p_9^2, $2p_9p_{10}$, and p_{10}^2.

The HW test tests the null hypothesis that the genotype proportions are p_9^2, $2p_9p_{10}$, and p_{10}^2, which are known as the Hardy-Weinberg proportions. It is also known as a "test for goodness of fit" of the observed data. In most cases under the null hypothesis, it follows a chi-square distribution, whose degrees of freedom are the number of data (in this case 3 genotypes) minus the number of parameters estimated from the data. In this case, the only parameter estimated is p_9, therefore to calculate the degrees of freedom is given by the following equation

$$df = num\ of\ genotypes - num\ of\ parameters\ calculated - 1$$

It is clear that for two alleles, there are three genotypes. And therefore, the degree of freedom associated with the goodness of fit test is one.

Other quantities necessary for the HW test are the HW-expectation values, which are calculated for each genotype as the genotype frequency times the sample size.

$$E\left(SLC6A3_{9-9}\right) = \hat{p}_9^2 N$$

$$E\left(SLC6A3_{9-10}\right) = 2\hat{p}_9\hat{p}_{10} N$$

$$E\left(SLC6A3_{10-10}\right) = \hat{p}_{10}^2 N$$

The proportions of the genotype frequencies have to be converted to the expected people in the following generation, given that the HW test is based on the observed people, instead of the genotype proportions. The test statistic is symbolized as χ^2 and is calculated by

$$\chi^2 = \sum_{All\ genotypes} \frac{(obs - exp)^2}{exp}$$

therefore, the HW test for this data is computed using the following equation

$$\chi^2 = \frac{(\#\ SLC\ 6A3_{9-9} - E(SLC\ 6A3_{9-9}))^2}{E(SLC\ 6A3_{9-9})} + \frac{(\#\ SLC\ 6A3_{9-10} - E(SLC\ 6A3_{9-10}))^2}{E(SLC\ 6A3_{9-10})} + \frac{(\#\ SLC\ 6A3_{10-10} - E(SLC\ 6A3_{10-10}))^2}{E(SLC\ 6A3_{10-10})}$$

If the probability associated with the χ^2 is small, it means that chance alone is not likely to lead to a deviation as large as obtained, and the model of equilibrium is invalid.

In this sample, besides the allele 9 and 10, two other alleles where identified in the DNA analysis, but their frequencies were considered negligible frequencies. As has been shown in other genetic analyses, African-Americans and Caucasians have different genotype distributions. A Fisher's exact test was used to test equal genotype distribution. Based on previous reports in the literature, it is expected a small number of homozygote SLC6A3 – 9 in AA. The analysis of the association between genotype and smoking was tested using a chi-square and Fisher's exact test. The association analysis was stratified by race, by medication groups, and by allele carrier. Interactions between genotypes and treatment groups were tested in a logistic regression model controlling for race as a confounder, at 6 and 12 months of follow up.

3.0 SAMPLE SELECTION

Participants of a large randomized clinical trail for smoking cessation and weight concern were invited to participate in an adjunct genetic project to investigate the relationship between the dopamine transporter gene and bupropion efficacy of the population studied. Exclusion criteria for the sample were: 1) participant not randomized, 2) unknown race/ethnicity, 3) other race than Caucasian or AA, 4) allele frequency less than 0.7%.

Participants were asked to rate their weight concerns and desire to quit smoking in a self-report scale at both a telephone interview and in person. Exclusion criteria from the trial were: 1) pregnancy, lactation, or no medically approved method of contraception; 2) major medical problem; 3) history of seizure disorder or head injury; 4) current or historical psychosis or bipolar disorder; 5) history of alcohol or substance abuse within the previous year; 6) current or historical eating disorders; 7) use of psychiatric medication in the previous months; 8) multiple drug allergies; 9) current Major Depressive Disorder.

Women who met eligibility criteria for the clinical trial attended an informational meeting during which the genotyping study was explained in detail, risks and benefits discussed, and informed consent obtained. Participants in the larger trial were required to have a physical examination, including phlebotomy, and a psychosocial interview. Thus, participants provided an additional 2 tubes of blood for genotyping purposes. Samples were sent to the Human

Genetics lab at the Graduate School of Public Health, under the direction of Robert Farrell, Ph.D., where the DNA was extracted and stored for analysis.

4.0 STATISTICAL METHODS AND ANALYSIS

4.1 LOGISTIC MODEL

In a logistic regression model the outcome variable is a binary or dichotomous variable, which can take only one of two possible values, for instance 0 or 1, yes or no. The outcome (Y) can be related to one or several other predictors known as covariates (X's), these covariates might or might not be dichotomous variables. The expected value of the outcome is conditioned on the predicted values E(Y|X's), which represents the conditional mean of the outcome given the predictors. With a *logit transformation*, the expected value has desirable properties of the general linear regression model. It is linear in its parameters, may be continuous and range from -∞ to +∞ depending on the range of the X's. Then, the logistic regression model is part of the general linear model where the link function is the logit function,

$$E(Y \mid X) = logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i}$$

where, $i = 1, \ldots, n$ indexes independent observations, p is the probability of an event in a fixed interval of time, and $\beta_0, \beta_1, \ldots,$ and β_k are the coefficient parameters that are estimated using the maximum likelihood method.

If the outcome depends on more than one variable or a variable with more than one level, the logistic model is known as multiple logistic regression model. This model is fit when the outcome is associated with at least two predictors or one predictor with two levels. Sometimes the predictors are also associated among themselves. When the predictors are not independent, that is, when the effect of one predictor depends on the level of another covariate (effect modification), it is practical to assess interaction among predictors in the logistic model, before making any assumption about the effect of the predictor. For instance, in following model,

$$logit(p_i) = \beta_0 + \beta_1 race + \beta_2 Genotype + \beta_3 TreatmentGroup + \beta_4 Genotype * TreatmentGroup$$

the term (*Genotype*TreatmentGroup*) represents the interaction or effect modifier between the predictors *Genotype* and *TreatmentGroup*. To evaluate whether the interaction is statistically significant, it is necessary to test the hypothesis that β_4 is equal to zero, which is evaluated using the Wald test or a likelihood ratio test. Logistic regression and stratified analysis of 2 X 2 tables assess interaction and they can be use to control for confounding.

One of the logistic models fitted in this data, uses the dichotomous answer for whether a woman resumes to smoke or not as the outcome variable. The predictors used in this model were categorical variables. In the logistic model, it is necessary to create dummy variables for the categorical predictors. The required number of dummy variables equals the number of categories minus one; for instance, two dummy variables are required for the genotypes and three dummy variables for the treatment groups. The interaction terms are given by the product of the dummy variables, as shown in the following equation

$$logit(p) = \beta_0 + \beta_1 race + \beta_2 H + \beta_3 R + \beta_4 G_2 + \beta_5 G_3 + \beta_6 G_4 + \beta_7 HG_2 + \beta_8 HG_3 + \beta_9 HG_4 + \beta_{10} RG_2 + \beta_{11} RG_3 + \beta_{12} RG_4$$

In order to test the significance of the interaction between these variables is necessary to compare the deviances of two models. One model with the interaction terms and another model without interaction terms; where, the null hypothesis $\beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$ is tested using the difference of the deviances known as likelihood ratio test. This quantity is compared to the chi-square distribution with 6 degrees of freedom.

All of the analyses were done in STATA 9 and SAS 9.1 with the whole sample as well as separately in each ethnicity group. A two sided significant level of 0.05 was used to determine statistical significance.

5.0 RESULTS

Two hundred twenty-eight women met the criteria of inclusion and agreed to participate in this genetic study. Sixteen women (7%) of original sample were excluded due to one of the following criteria: six participants (2.6%) were not randomized to one of the four groups. The treatment group was unknown in two (0.9%) participants. Three (1.3%) did not state their race. One woman (0.4%) was neither African-American nor Caucasian. Four participants (1.8%) had one allele different than 9 or 10, those genotype frequencies were considered negligible.

Accordingly, the eligible sample consisted of 212 women; all women provided written informed consent. In this sample the average age was 41.9 years (*SD* = 10.14, range 19 – 62), the ethnicity composition was 86.3% Caucasian (N = 183) and 13.7% African-American (N = 29). The demographic characteristics of these women are shown in Table 3. This sample includes 7.5 % (N = 16) with the SLC3A6 – 9 homozygote, 51% (N = 108) with the SLC3A6 – 10 homozygote, and the remaining 41.5% (N = 88) were heterozygote with SLC3A6 – 9/10. The allelic and genotype frequencies by race are listed in Table 4. The allele frequencies are 0.283 for the allele 9 and 0.717 for the allele 10. The 95% confidence intervals are (0.240, 0.326) for allele 9 and (0.674, 0.760) for the allele 10. The allele frequencies were stratified by race/ethnicity. These results are consistent with those reported for Europeans and AA in literature, but in this sample there were two other alleles for DAT which were not considered for this analysis.

The SLC6A3 – 9/* genotype was significantly less common in AA than in Caucasian (27.6% vs. 52.5%) with a Fisher's exact p = 0.033. The distribution of this genotype was considered in HW equilibrium for this sample (χ^2 = 0.11 and *p-value* = 0.73). The HW test in each race are χ^2 = 0.79 and *p-value* = 0.37 for withes and χ^2 = 2.19 and *p-value* = 0.1386 for AA.

Table 3. Demographic characteristics.

		All participant N = 212		**Caucasian N = 183**		**AA N = 29**	
Variable		Mean	SD	Mean	SD	Mean	SD
Age		41.99	10.14	41.58	10.51	44.07	7.08
Attendance to treatment		8.04	2.92	8.19	2.88	7.07	3.01
BMI pre-quit		27.52	5.34	27.15	5.04	29.98	6.57
Height		64.27	2.43	64.18	2.31	64.76	3.11
Number of cigarettes/day		20.57	8.50	21.09	8.37	17.31	8.74
Years of smoking		24.83	12.55	24.58	11.84	26.48	16.53
Education	Less than high school	N = 6	(2.83%)	N = 6	(3.28%)	N = 0	(0%)
	High school	N = 27	(12.74%)	N = 21	(11.48%)	N = 1	(3.45%)
	No graduated from college or technical school	N = 111	(52.36%)	N = 94	(44.34%)	N =17	(58.62%)
	Graduated from college	N = 48	(22.64%)	N = 45	(24.59%)	N = 3	(10.34%)
	Post-graduate degree	N = 20	(9.43%)	N = 17	(9.29%)	N = 3	(10.34%)

Table 4. Dopamine transporter frequencies by race/ethnicity.

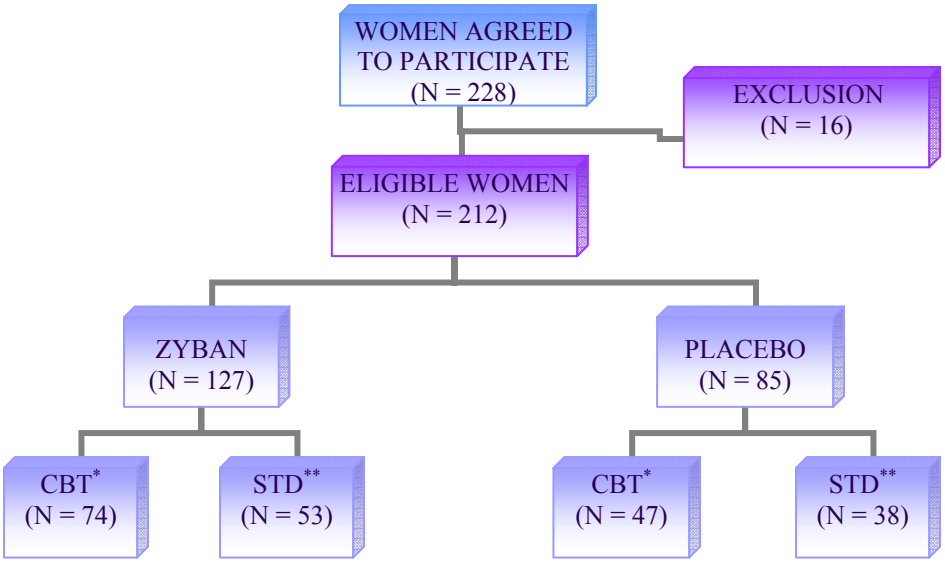
Alleles

	Caucasian			AA		
N	N	Frequency	95% CI	N	Frequency	95% CI
9	110	0.301	0.254 – 0.347	10	0.172	0.075 – 0.270
10	256	0.699	0.652 – 0.746	48	0.828	0.739 – 0.925

Genotype

	Caucasian (N = 183)		AA (N = 29)	
	N	%	N	%
9 – 9	14	7.65	2	6.9
9 – 10	82	44.81	6	20.69
10 – 10	87	47.54	21	72.41

All participants were randomly assigned to one of the four groups 1) placebo and standard (STD), 2) placebo and cognitive behavioral therapy (CBT), 3) zyban and STD, and 4) zyban and CBT as depicted in Figure 1.



* Cognitive behavioral therapy
** Standard

Figure 1. Sampling diagram.

As shown in Tables 5 and 6, there is not a statistically significant difference in the distribution of genotypes among the groups χ^2 (6) = 5.511, Fisher's exact p = 0.435, N = 212. The same result was obtained when combining all individuals with at least one gene SLC6A3 – 9 χ^2 (3) = 0.221, *p-value* = 0.974, N = 212, and combining all individual with at least one allele SLC6A3 – 10, χ^2 (3) = 4.807, Fisher's exact p = 0.147, N = 212. Stratifying by ethnicity/race the four different treatment groups were compared with the genotype groups in Caucasians. First, three possible combinations of the alleles (9 – 9, 9 – 10, and 10 – 10) were considered and compared with the treatment group for association, χ^2 (6) = 6.164, Fisher's exact p = 0.318, N = 183. Second, the 9 – carrier individuals (9 – *, 10 – 10) were compared with the treatment group χ^2 (3) = 0.204, Fisher's exact p = 0.981, N = 183. Finally combining 10 – carrier individuals (9 – 9, 10 – *) were compared with the treatment group χ^2 (3) = 5.719, Fisher's exact p = 0.073, N = 183. The sample size is too small for African-Americans to attain statistical significance in these analyses; however their results are listed in Tables 5 and 6.

Table 5. Association of genotype and treatment group by race.

		Genotype					
		9 – 9		9 – 10		10 – 10	
Sample	Group	N	%	N	%	N	%
All participants N = 212	Placebo + STD*	4	10.53	14	36.84	20	52.53
	Placebo + CBT**	6	12.77	18	38.30	23	48.94
	Zyban + STD	4	7.55	21	39.62	28	52.83
	Zyban + CBT	2	2.70	35	47.30	37	50.00
χ^2 (6)		5.511					
p-value		0.480					
Fisher's exact *p*		0.435					
Caucasian N = 183	Placebo + STD	4	12.12	14	42.42	15	45.45
	Placebo + CBT	5	12.82	16	41.03	18	46.15
	Zyban + STD	4	8.33	20	41.67	24	50.00
	Zyban + CBT	1	1.59	32	50.79	30	47.62
χ^2 (6)		6.164					
p-value		0.405					
Fisher's exact *p*		0.318					
AA N = 29	Placebo + STD	0	0	0	0	5	100
	Placebo + CBT	1	12.50	2	25.00	5	62.50
	Zyban + STD	0	0	1	20.00	4	80.00
	Zyban + CBT	1	9.09	3	27.27	7	63.64
χ^2 (6)		3.259					
p-value		0.776					
Fisher's exact *p*		0.910					

* STD = Standard

** CBT = cognitive-behavioral therapy

Table 6. Association of allele carrier and treatment group by race.

		Genotype							
		9 – *		10 – 10		9 – 9		10 – *	
Sample	Group	N	%	N	%	N	%	N	%
All participant N = 212	Placebo + STD[a]	18	47.37	20	52.63	4	10.52	34	89.47
	Placebo + CBT[b]	24	52.06	23	48.94	6	12.77	41	87.23
	Zyban + STD	25	47.17	28	52.83	4	7.55	49	92.45
	Zyban + CBT	37	50.00	37	50.00	2	2.70	72	97.30
χ^2 (3)		0.221				4.807			
p-value		0.974				0.186			
Fisher's exact *p*		0.974				0.147			
Caucasian N = 183	Placebo + STD	18	54.55	15	45.45	4	12.12	29	87.88
	Placebo + CBT	21	53.85	18	4615	5	12.82	34	87.18
	Zyban + STD	24	50.00	24	50.00	4	8.33	44	91.67
	Zyban + CBT	33	52.38	30	47.62	1	1.59	62	98.41
χ^2 (3)		0.204				5.719			
p-value		0.977				0.126			
Fisher's exact *p*		0.981				0.073			
African-American N = 29	Placebo + STD	0	0	5	100	0	0	5	100
	Placebo + CBT	3	37.5	5	62.50	1	12.50	7	87.50
	Zyban + STD	1	20.00	4	80.00	0	0	5	100
	Zyban + CBT	4	36.36	7	63.64	1	9.09	10	90.91
χ^2 (3)		2.867				1.214			
p-value		0.413				0.750			
Fisher's exact *p*		0.562				1.00			

[a] Standard
[b] Cognitive behavioral therapy

The participants were stratified according to their treatment group considering only zyban and placebo. The results are listed in Tables 7 and 8. Similarly, there is not a significant association between the genotypes and the treatment groups, considering all participants, N = 212, χ^2 (2) = 3.858, Fisher's exact p = 0.144 for the three possible genotypes. χ^2 (1) = 0.007, *p-value* = 0.933 for at least one allele 9. χ^2 (1) = 3.617, *p-value* = 0.057 for at least one allele 10. Similar results were obtained when considering only Caucasians for three possible combinations of the alleles and at least one 9 gene N = 183, χ^2 (2) = 3.984, *p-value* = 0.136 and χ^2 (1) = 0.139, *p-value* = 0.763, respectively. In addition, the genotypes 9/10 and 10/10 were grouped, χ^2 (1) = 3.952, and *p-value* = 0.047.

Table 7. Association of genotype and drug by race.

		Genotype					
		9 – 9		9 – 10		10 – 10	
Sample	Group	N	%	N	%	N	%
All participant N = 212	Placebo	10	11.76	32	37.65	43	50.59
	Zyban	6	4.72	56	44.09	65	51.19
χ^2 (2)		3.858					
p-value		0.145					
Fisher's exact *p*		0.144					
Caucasian N = 183	Placebo	9	12.50	30	41.67	33	45.83
	Zyban	5	4.50	52	46.85	54	48.65
χ^2 (2)		3.984					
p-value		0.136					
Fisher's exact *p*		0.155					
AA N = 29	Placebo	1	7.69	2	15.38	10	76.92
	zyban	1	6.25	4	25.00	11	68.75
χ^2 (2)		0.408					
p-value		0.815					
Fisher's exact *p*		0.827					

Table 8. Association of allele carrier and drug by race.

		Genotype							
		9 – *		10 – 10		9 – 9		10 – *	
Sample	Group	N	%	N	%	N	%	N	%
All participant	Placebo	42	49.41	43	50.59	10	11.76	75	88.24
N = 212	Zyban	62	48.82	65	51.18	6	4.72	121	95.28
χ^2 (1)		0.007				3.617			
p-value		0.933				0.057			
Fisher's exact *p*		1.000				0.067			
Caucasians	Placebo	39	54.17	33	45.83	9	12.50	63	87.50
N = 183	Zyban	57	51.35	54	48.65	5	4.50	106	95.50
χ^2 (1)		0.139				3.952			
p-value		0.763				0.047			
Fisher's exact *p*		0.763				0.084			
AA	Placebo	5	31.25	11	68.75	1	7.69	12	92.31
N = 29	Zyban	3	23.08	10	76.92	1	6.25	15	93.75
χ^2 (1)		0.2398				0.023			
p-value		0.624				0.879			
Fisher's exact *p*		0.697				1.00			

The number and proportion of women that relapsed at six and twelve months in each group are listed in Tables 9 and 10.

Table 9. Percentage of women that relapsed in different genotypes and drug.

All participants

Genotype		9 – 9		9 – 10		10 – 10	
Relapsed	Drug	N	%	N	%	N	%
6 months	Placebo	9	90.00	28	87.50	38	88.37
	Zyban	3	50.00	35	62.50	45	69.23
12 months	Placebo	9	90.00	30	93.75	38	88.37
	Zyban	4	66.67	38	67.86	55	84.62

Caucasians

6 months	Placebo	8	88.89	26	86.67	29	87.88
	Zyban	2	40.00	32	61.54	36	66.67
12 months	Placebo	8	88.89	28	93.33	29	87.88
	Zyban	3	60.00	34	65.38	46	85.19

AA

6 months	Placebo	1	100.00	3	75.00	9	81.82
	Zyban	1	100.00	2	100.00	9	90.00
12 months	Placebo	1	100.00	4	100.00	9	81.82
	Zyban	1	100.00	2	100.00	9	90.00

Table 10. Percentage of women that relapsed.

All participants

Genotype		9 – *		10 – 10		9 – 9		10 – *	
Relapsed	Drug	N	%	N	%	N	%	N	%
6 months	Placebo	37	88.10	38	88.37	9	90.00	66	88.00
	Zyban	38	61.29	45	69.23	3	50.00	80	66.12
12 months	Placebo	39	92.86	38	88.37	9	90.00	68	90.67
	Zyban	42	67.74	55	84.62	4	66.67	93	76.86

Caucasians

6 months	Placebo	34	87.18	29	87.88	8	88.89	55	87.30
	Zyban	34	59.66	36	66.67	2	40.00	68	64.15
12 months	Placebo	36	92.31	29	87.88	8	88.89	57	90.48
	Zyban	37	64.91	46	85.19	3	60.00	80	75.47

AA

6 months	Placebo	3	100.00	9	90.00	1	100.00	11	91.67
	Zyban	4	80.00	9	81.82	1	100.00	12	80.00
12 months	Placebo	3	100.00	9	90.00	1	100.00	11	91.67
	Zyban	5	100.00	9	81.82	1	100.00	13	86.67

There was no association between the genotypes and the women in the STD and CBT groups, N = 212, χ^2 (2) = 0.786, *p-value* = 0.675 for all participants. In Caucasians, N = 183, χ^2 (2) = 1.213, *p-value* = 0.545. There was not a statistically significant association by segregating alleles (see Table 12). Additionally, the association of the genotype with smoking cessation was attained at 6 and 12 months, χ^2 (2) = 0.709, Fisher' exact *p* = 0.691, and χ^2 (2) = 2.583, Fisher's exact *p* = 0.253, respectively considering all participants, N = 212. Consistent results were obtained by stratifying the sample by ethnicity/race and by allele type (results are summarized on Tables 13 and 14).

Table 11. Association of genotypes and cognitive-behavioral therapy group by race.

		Genotype					
		9 – 9		9 – 10		10 – 10	
Sample	Group	N	%	N	%	N	%
All participant N = 212	STD*	8	8.79	35	38.46	48	52.75
	CBT**	8	6.12	53	43.80	60	49.59
χ^2 (2)		0.786					
p-value		0.675					
Fisher's exact *p*		0.698					
Caucasian N = 183	STD	8	9.88	34	41.98	39	48.14
	CBT	6	5.88	48	47.06	48	47.06
χ^2 (2)		1.213					
p-value		0.545					
Fisher's exact *p*		0.552					
AA N = 29	STD	0	0	1	10.00	9	90.00
	CBT	2	10.52	5	26.32	12	63.16
χ^2 (2)		2.548					
p-value		0.280					
Fisher's exact *p*		0.327					

* STD = Standard
** CBT = Cognitive-Behavioral Therapy

Table 12. Association of allele carrier and therapy group by race.

		Genotype							
		9 – *		10 – 10		9 – 9		10 – *	
Sample	Group	N	%	N	%	N	%	N	%
All participant	STD*	43	47.25	48	52.75	8	8.79	83	91.21
N = 212	CBT**	61	50.41	60	49.59	8	6.61	113	93.39
χ^2 (1)		0.208				0.354			
p-value		0.649				0.552			
Fisher's exact *p*		0.679				0.605			
Caucasians	STD	42	51.86	39	48.14	8	9.88	73	90.12
N = 183	CBT	54	52.94	48	47.06	6	5.88	96	94.12
χ^2 (1)		0.0215				1.020			
p-value		0.883				0.313			
Fisher's exact *p*		1.000				0.404			
AA	STD	1	10.00	9	90.00	0	0.00	10	100.0
N = 29	CBT	7	36.84	12	63.16	2	10.53	17	89.47
χ^2 (1)		2.363				1.1306			
p-value		0.124				0.288			
Fisher's exact *p*		0.201				0.532			

* STD = Standard

** CBT = Cognitive-behavioral therapy

Table 13. Association of genotype and relapsed at six and twelve months.

		Genotype					
		At six months			At 12 months		
Sample	Relapsed	9 – 9	9 – 10	10 – 10	9 – 9	9 – 10	10 – 10
All participant	No	4	25	25	3	20	15
N = 212	Yes	12	63	83	13	68	93
χ^2 (2)		0.709			2.583		
p-value		0.702			0.275		
Fisher's exact *p*		0.691			0.253		
Caucasian	No	4	24	22	3	20	12
N = 183	Yes	10	58	65	11	62	75
χ^2 (2)		0.3488			3.117		
p-value		0.840			0.210		
Fisher's exact *p*		0.817			0.204		
AA	No	0	1	3	0	0	3
N = 29	Yes	2	5	18	2	6	18
χ2 (2)		0.366			1.275		
p-value		0.833			0.529		
Fisher's exact *p*		1.000			1.000		

Table 14. Association of allele carrier and relapsed at six and twelve months.

		Genotypes															
		At six months								At twelve months							
		9 – *		10 – 10		9 – 9		10 – *		9 – *		10 – 10		9 – 9		10 – *	
Sample	Relapse	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
All participant	No	29	53.70	25	46.30	4	7.41	50	92.59	23	60.53	15	39.47	3	7.89	35	92.11
N = 212	Yes	75	47.47	83	52.53	12	7.59	146	92.41	81	46.55	93	53.45	13	7.47	161	92.53
χ^2 (1)		0.626				0.002				2.437				0.008			
p-value		0.429				0.964				0.118				0.929			
Fisher's exact *p*		0.436				1.000				0.152				1.000			
Caucasians	No	28	56.00	22	44.00	4	8.00	46	92.00	23	65.71	12	34.29	3	8.57	32	91.43
N = 183	Yes	68	51.13	65	48.87	10	7.52	123	92.48	73	49.32	75	50.68	11	7.43	137	92.57
χ^2 (1)		0.346				0.012				3.049				0.052			
p-value		0.556				0.913				0.081				0.820			
Fisher's exact *p*		0.620				1.000				0.092				0.733			
AA	No	1	25.00	3	75.00	0	0.00	4	100.0	0	0.00	3	100.0	0	0.00	3	100.0
N = 29	Yes	7	28.00	18	72.00	2	8.00	23	92.00	8	30.77	18	69.23	2	7.69	24	92.31
χ^2 (1)		0.016				0.3437				1.275				0.248			
p-value		0.901				0.558				0.259				0.619			
Fisher's exact *p*		1.000				1.000				0.54				1.000			

Using logistic analysis, some models where fitted. Whether the women relapsed at six months was used as the outcome variable; and race, genotype, and medication group were used as predictors. Using a model with a first order interaction of genotype and medication group and compared with a model without the interaction term, the likelihood ratio test was performed. The result for likelihood ratio test at six months of follow up was not statistically significant, χ^2 (6) = 7.243, and *p-value* 0.299. Similar results were o btained at twelve months of follow up χ^2(6) = 7.399, and *p-value* 0.286,. The deviance for each model is listed on Table 15.

Table 15. Logistic models.

Model	Outcome	Variables	Deviance	χ^2	df	*p-value*
1	Relapsed at 6 months	Race Genotype Treatment Group	219.142	7.243	6	0.299
2	Relapsed at 6 months	Race Genotype Treatment Group Interaction	211.899			
3	Relapsed at 12 months	Race Genotype Treatment Group	187.087	7.399	6	0.289
4	Relapsed at 12 months	Race Genotype Treatment Group Interaction	179.688			

Although there were some calculation problems due to some zero counts, the tests for interaction between the genotype and the treatment group were not statistically significant. The odds ratios are listed in Table 16, where the baseline person is from Cognitive-behavioral therapy and bupropion group with genotype 10 – 10.

Table 16. Estimated Odds Ratios.

Effect	Odds ratio 6 months	95% Wald Confidence Limits		Odds Ratios 12 months	95% Wald Confidence Limits	
9 – 10	0.868	0.438	1.719	0.599	0.279	1.285
9 – 9	0.601	0.165	2.186	0.509	0.123	2.110
Zyban + STD	2.038	0.933	4.455	1.649	0.688	3.956
Placebo + CBT	6.059	2.101	17.473	4.130	1.283	13.290
Placebo + STD	4.836	1.661	14.077	3.279	1.010	10.640
Caucasian	0.413	0.130	1.315	0.539	0.148	1.966

From Table 16, it is clear that genotype make no difference in the odds ratios. The effect of genotype is no statistically significant. At 6 months and one year, the group that does better was the bupropion and cognitive behavioral treatment. Other models of interest were relapsed vs. drug, adjusted for race and genotype, including interaction between the genotype and the drug. The likelihood ratio tests is $\chi^2(2) = 0.57$ and the *p-value* = 0.7507. The odds ratios for this model are listed on the Table 17. The baseline person is a Caucasian woman with genotype 9 – 9 in the zyban group.

Table 17. Estimated odds ratios relapsed vs. zyban, adjusted by race and genotype.

Effect	Odds Ratio At 6 months	95% Confidence Interval		Odds ratio At one year	95% Confidence interval
AA	2.221	0.709	6.952	1.760	0.488 6.346
Placebo	4.055	1.881	8.741	3.023	1.291 – 7.076
9 – 10	1.257	0.343	4.607	1.068	0.263 – 4.336
10 – 10	1.498	0.412	5.440	1.820	0.441 – 7.513

6.0 DISCUSSION

This study provides evidences of different distribution of genotypes among races. The frequencies of the alleles are different between Caucasian and African American women. The frequencies of the genotypes are similar to those reported in the literature, where the allele 10 is more likely to be present in African American women than in Caucasians. In fact, there were not many 9 homozygotes on the African American participants. The allele frequencies of Caucasian women were within the 95% confidence intervals of the whole sample, nevertheless the allele frequencies of AA were not included in such ranges. In spite of the differences, the sample of women participants in the smoking cessation program is in Hardy Weinberg equilibrium.

Based on previous evidence of bupropion's efficacy and cognitive behavioral therapy in smoking cessation program, a large clinical trial was performed by UPMC and University of Pittsburgh researchers. In such study participants were asked to participate in this ancillary program, where participants agreed to undergo a DNA test for the dopamine transporter gene. There were some independent hypotheses that were stated before the study began. There were some post hoc analyses, where the alpha level was not adjusted for the multiple comparisons. The relapsed ratios were not significantly different by adding the genotype and phenotype information, neither by adding race. It was not possible to calculate the coefficient for the interaction term, due to few data and zero cells in the designed matrix.

The logistic models were computed to analyze the interaction of the SCL6A3's genotypes and the treatment groups, controlling for ethnicity/race. One model was run at the 6th month and another model at the 12th month. Interaction between allele carriers and treatment groups were also analyzed at the same times. However, there were not significant results in any of the models tested. There were some undefined parameters in the interaction terms given the small sample size and some zero counts, in order to check the interaction terms. Therefore some groups were combined to have an estimated value for the treatment effect.

The likelihood ratio test for none of the models analyzed was statistically significant. Thus, the interaction coefficients are not significantly different than zero. The only significant explanatory variable was the bupropion use. Women were more likely to respond if they were in the treatment group.

APPENDIX SAS OUTPUT

MODEL 1

The SAS System

The LOGISTIC Procedure

Model Information

Data Set	WORK.LUNA	
Response Variable	relapsed6	relapsed6
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	212
Number of Observations Used	212

Response Profile

Ordered Value	relapsed6	Total Frequency
1	0	54
2	1	158

Probability modeled is relapsed6='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	242.602	233.142
SC	245.959	256.638
-2 Log L	240.602	219.142

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.4601	6	0.0015
Score	20.1109	6	0.0026
Wald	18.1232	6	0.0059

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2314	0.5816	4.4827	0.0342
het	1	-0.1420	0.3487	0.1657	0.6839
h9	1	-0.5087	0.6585	0.5967	0.4398
g2	1	0.7122	0.3989	3.1873	0.0742
g3	1	1.8016	0.5404	11.1160	0.0009
g4	1	1.5760	0.5452	8.3571	0.0038
white	1	-0.8849	0.5914	2.2391	0.1346

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
het	0.868	0.438	1.719
h9	0.601	0.165	2.186
g2	2.038	0.933	4.455
g3	6.059	2.101	17.473
g4	4.836	1.661	14.077
white	0.413	0.130	1.315

Association of Predicted Probabilities and Observed Responses

Percent Concordant	66.0	Somers' D	0.411
Percent Discordant	24.9	Gamma	0.452
Percent Tied	9.1	Tau-a	0.157
Pairs	8532	c	0.706

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
het	1.0000	0.868	0.438	1.719
h9	1.0000	0.601	0.165	2.186
g2	1.0000	2.038	0.933	4.455
g3	1.0000	6.059	2.101	17.473
g4	1.0000	4.836	1.661	14.077
white	1.0000	0.413	0.130	1.315

** het = genotype 9-10
** h9 = genotype 9-9
** g2 = zyban + STD
** g3 = placebo + CBT
** g4 = placebo + STD
** hetg2 = interaction between het and g2
** hetg3 = interaction between het and g3
** hetg4 = interaction between het and g4
** h9g2 = interaction between h9 and g2
** h9g3 = interaction between h9 and g3
** h9g4 = interaction between h9 and g4

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
het	Score	0.1614	0.7308	0.6675
	Probability	0.1265	0.7308	0.6675
h9	Score	0.5802	0.4824	0.3858
	Probability	0.1931	0.4824	0.3858
g2	Score	3.1787	0.0880	0.0725
	Probability	0.0310	0.0880	0.0725
g3	Score	12.4750	0.0007	0.0006
	Probability	0.000218	0.0003	0.0002
g4	Score	9.0838	0.0043	0.0036
	Probability	0.00145	0.0024	0.0017
white	Score	2.2917	0.1474	0.1148
	Probability	0.0651	0.1474	0.1148

Exact Parameter Estimates

Parameter	Estimate	95% Confidence Limits		p-Value
het	-0.1384	-0.8717	0.5970	0.8166
h9	-0.4890	-1.8846	1.1012	0.6515
g2	0.6981	-0.1336	1.5659	0.1080
g3	1.7606	0.6665	3.0612	0.0005
g4	1.5408	0.4321	2.8512	0.0036
white	-0.8660	-2.3358	0.3297	0.1964

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		p-Value
het	0.871	0.418	1.817	0.8166
h9	0.613	0.152	3.008	0.6515
g2	2.010	0.875	4.787	0.1080
g3	5.816	1.947	21.352	0.0005
g4	4.668	1.540	17.308	0.0036
white	0.421	0.097	1.391	0.1964

MODEL 2

The SAS System

The LOGISTIC Procedure

Model Information

Data Set	WORK.LUNA	
Response Variable	relapsed6	relapsed6
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	212
Number of Observations Used	212

Response Profile

Ordered Value	relapsed6	Total Frequency
1	0	54
2	1	158

Probability modeled is relapsed6='1'.

Model Convergence Status

Quasi-complete separation of data points detected.

WARNING: The maximum likelihood estimate may not exist.
WARNING: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	242.602	237.899
SC	245.959	281.534
-2 Log L	240.602	211.899

The LOGISTIC Procedure

WARNING: The validity of the model fit is questionable.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	28.7036	12	0.0044
Score	27.3304	12	0.0069
Wald	21.2613	12	0.0467

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.1589	0.6099	3.6106	0.0574
het	1	-0.2521	0.4855	0.2698	0.6035
h9	1	11.8369	369.7	0.0010	0.9745
g2	1	0.8469	0.5766	2.1571	0.1419
g3	1	1.8579	0.8174	5.1660	0.0230
g4	1	1.2143	0.7167	2.8705	0.0902
white	1	-0.7988	0.5932	1.8129	0.1782
hetg2	1	0.1778	0.8433	0.0445	0.8330
hetg3	1	0.0465	1.1615	0.0016	0.9680
hetg4	1	0.4695	1.1025	0.1813	0.6702
h9g2	1	-14.1426	369.7	0.0015	0.9695
h9g3	1	-12.5509	369.7	0.0012	0.9729
h9g4	1	-0.8597	455.4	0.0000	0.9985

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
het	0.777	0.300	2.012
h9	>999.999	<0.001	>999.999
g2	2.332	0.753	7.221
g3	6.410	1.291	31.816
g4	3.368	0.827	13.723
white	0.450	0.141	1.439
hetg2	1.195	0.229	6.238
hetg3	1.048	0.108	10.206
hetg4	1.599	0.184	13.878
h9g2	<0.001	<0.001	>999.999
h9g3	<0.001	<0.001	>999.999
h9g4	0.423	<0.001	>999.999

The LOGISTIC Procedure

WARNING: The validity of the model fit is questionable.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	68.5	Somers' D	0.461
Percent Discordant	22.4	Gamma	0.507
Percent Tied	9.1	Tau-a	0.176
Pairs	8532	c	0.731

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
het	1.0000	0.777	0.300	2.012
h9	1.0000	>999.999	<0.001	>999.999
g2	1.0000	2.332	0.753	7.221
g3	1.0000	6.410	1.291	31.816
g4	1.0000	3.368	0.827	13.723
white	1.0000	0.450	0.141	1.439
hetg2	1.0000	1.195	0.229	6.238
hetg3	1.0000	1.048	0.108	10.206
hetg4	1.0000	1.599	0.184	13.878
h9g2	1.0000	<0.001	<0.001	>999.999
h9g3	1.0000	<0.001	<0.001	>999.999
h9g4	1.0000	0.423	<0.001	>999.999

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
het	Score	0.2705	0.6361	0.5528
	Probability	0.1667	0.6361	0.5528
h9	Score	0.9380	0.5527	0.3188
	Probability	0.4677	1.0000	0.7661
g2	Score	2.1563	0.1782	0.1408
	Probability	0.0747	0.1782	0.1408
g3	Score	5.9677	0.0165	0.0111
	Probability	0.0109	0.0165	0.0111
g4	Score	2.9856	0.1260	0.0982
	Probability	0.0557	0.1260	0.0982
white	Score	1.8051	0.2168	0.1739
	Probability	0.0857	0.2168	0.1739
hetg2	Score	0.0435	1.0000	0.8396
	Probability	0.3207	1.0000	0.8396
hetg3	Score	0.0018	1.0000	0.7825
	Probability	0.4349	1.0000	0.7825
hetg4	Score	0.1770	1.0000	0.8093
	Probability	0.3814	1.0000	0.8093
h9g2	Score	3.0486	0.1596	0.0798
	Probability	0.1596	0.1596	0.0798
h9g3	Score	1.0732	0.4824	0.2412
	Probability	0.4824	0.4824	0.2412
h9g4	Score	.#	.	.
	Probability	1.0000#	1.0000	0.5000

NOTE: # indicates that the conditional distribution is degenerate.

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Exact Parameter Estimates

Parameter	Estimate	95% Confidence Limits		p-Value
het	-0.2495	-1.3062	0.7990	0.7775
h9	0.1122*	-2.5060	Infinity	0.9354
g2	0.8281	-0.4006	2.1647	0.2259
g3	1.8208	0.1731	4.1377	0.0251
g4	1.1866	-0.3052	3.0289	0.1453
white	-0.7725	-2.2431	0.4257	0.2694
hetg2	0.1724	-1.7501	2.1256	1.0000
hetg3	0.0490	-2.8784	2.9765	1.0000
hetg4	0.4492	-2.1274	3.2880	1.0000
h9g2	-1.4715*	-Infinity	1.4519	0.3192
h9g3	-0.0706*	-Infinity	3.5929	0.9647
h9g4	.#	.	.	.

NOTE: # indicates that the conditional distribution is degenerate.
* indicates a median unbiased estimate.

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		p-Value
het	0.779	0.271	2.223	0.7775
h9	1.119*	0.082	Infinity	0.9354
g2	2.289	0.670	8.712	0.2259
g3	6.177	1.189	62.656	0.0251
g4	3.276	0.737	20.675	0.1453
white	0.462	0.106	1.531	0.2694
hetg2	1.188	0.174	8.378	1.0000
hetg3	1.050	0.056	19.620	1.0000
hetg4	1.567	0.119	26.790	1.0000
h9g2	0.230*	0	4.271	0.3192
h9g3	0.932*	0	36.341	0.9647
h9g4	. #	.	.	.

NOTE: # indicates that the conditional distribution is degenerate.
* indicates a median unbiased estimate.

MODEL 3

The SAS System

The LOGISTIC Procedure

Model Information

Data Set	WORK.LUNA	
Response Variable	relapsed12	relapsed12
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	212
Number of Observations Used	212

Response Profile

Ordered Value	relapsed12	Total Frequency
1	0	38
2	1	174

Probability modeled is relapsed12='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	201.385	201.087
SC	204.741	224.583
-2 Log L	199.385	187.087

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.2977	6	0.0556
Score	11.6085	6	0.0713
Wald	10.7273	6	0.0972

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.8094	0.6563	7.6012	0.0058
het	1	-0.5122	0.3893	1.7308	0.1883
h9	1	-0.6757	0.7258	0.8667	0.3519
g2	1	0.5005	0.4463	1.2575	0.2621
g3	1	1.4182	0.5963	5.6561	0.0174
g4	1	1.1875	0.6006	3.9094	0.0480
white	1	-0.6178	0.6602	0.8756	0.3494

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
het	0.599	0.279	1.285
h9	0.509	0.123	2.110
g2	1.649	0.688	3.956
g3	4.130	1.283	13.290
g4	3.279	1.010	10.640
white	0.539	0.148	1.966

Association of Predicted Probabilities and Observed Responses

Percent Concordant	63.6	Somers' D	0.366
Percent Discordant	27.0	Gamma	0.404
Percent Tied	9.4	Tau-a	0.108
Pairs	6612	c	0.683

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
het	1.0000	0.599	0.279	1.285
h9	1.0000	0.509	0.123	2.110
g2	1.0000	1.649	0.688	3.956
g3	1.0000	4.130	1.283	13.290
g4	1.0000	3.279	1.010	10.640
white	1.0000	0.539	0.148	1.966

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
het	Score	1.7105	0.2490	0.2165
	Probability	0.0651	0.2490	0.2165
h9	Score	0.8622	0.4001	0.3056
	Probability	0.1890	0.4001	0.3056
g2	Score	1.2470	0.2918	0.2454
	Probability	0.0927	0.2918	0.2454
g3	Score	6.1167	0.0166	0.0126
	Probability	0.00802	0.0166	0.0126
g4	Score	4.1024	0.0502	0.0379
	Probability	0.0246	0.0502	0.0379
white	Score	0.8776	0.4160	0.3375
	Probability	0.1570	0.4160	0.3375

Exact Parameter Estimates

Parameter	Estimate	95% Confidence Limits		p-Value
het	-0.5006	-1.3375	0.3198	0.2630
h9	-0.6519	-2.1699	1.1909	0.5674
g2	0.4902	-0.4455	1.4797	0.3650
g3	1.3856	0.1762	2.8647	0.0203
g4	1.1607	-0.0620	2.6485	0.0664
white	-0.6039	-2.3314	0.7140	0.5264

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		p-Value
het	0.606	0.262	1.377	0.2630
h9	0.521	0.114	3.290	0.5674
g2	1.633	0.641	4.391	0.3650
g3	3.997	1.193	17.544	0.0203
g4	3.192	0.940	14.133	0.0664
white	0.547	0.097	2.042	0.5264

MODEL 4

The SAS System

The LOGISTIC Procedure

Model Information

Data Set	WORK.LUNA	
Response Variable	relapsed12	relapsed12
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	212
Number of Observations Used	212

Response Profile

Ordered Value	relapsed12	Total Frequency
1	0	38
2	1	174

Probability modeled is relapsed12='1'.

Model Convergence Status

Quasi-complete separation of data points detected.

WARNING: The maximum likelihood estimate may not exist.
WARNING: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	201.385	205.688
SC	204.741	249.324
-2 Log L	199.385	179.688

The LOGISTIC Procedure

WARNING: The validity of the model fit is questionable.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	19.6969	12	0.0730
Score	19.5445	12	0.0762
Wald	15.4368	12	0.2184

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.9519	0.7114	7.5278	0.0061
het	1	-0.8799	0.5498	2.5617	0.1095
h9	1	11.2062	429.6	0.0007	0.9792
g2	1	0.6916	0.7433	0.8658	0.3521
g3	1	0.8859	0.8526	1.0797	0.2988
g4	1	0.2491	0.7569	0.1083	0.7421
white	1	-0.5923	0.6630	0.7982	0.3716
hetg2	1	-0.0321	0.9684	0.0011	0.9735
hetg3	1	1.4155	1.3820	1.0490	0.3057
hetg4	1	1.8362	1.3333	1.8968	0.1684
h9g2	1	-13.2574	429.6	0.0010	0.9754
h9g3	1	-11.9254	429.6	0.0008	0.9779
h9g4	1	0.0245	527.9	0.0000	1.0000

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
het	0.415	0.141	1.218
h9	>999.999	<0.001	>999.999
g2	1.997	0.465	8.572
g3	2.425	0.456	12.896
g4	1.283	0.291	5.655
white	0.553	0.151	2.028
hetg2	0.968	0.145	6.462
hetg3	4.119	0.274	61.823
hetg4	6.273	0.460	85.572
h9g2	<0.001	<0.001	>999.999
h9g3	<0.001	<0.001	>999.999
h9g4	1.025	<0.001	>999.999

The LOGISTIC Procedure

WARNING: The validity of the model fit is questionable.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	66.6	Somers' D	0.425
Percent Discordant	24.1	Gamma	0.469
Percent Tied	9.3	Tau-a	0.126
Pairs	6612	c	0.713

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
het	1.0000	0.415	0.141	1.218
h9	1.0000	>999.999	<0.001	>999.999
g2	1.0000	1.997	0.465	8.572
g3	1.0000	2.425	0.456	12.896
g4	1.0000	1.283	0.291	5.655
white	1.0000	0.553	0.151	2.028
hetg2	1.0000	0.968	0.145	6.462
hetg3	1.0000	4.119	0.274	61.823
hetg4	1.0000	6.273	0.460	85.572
h9g2	1.0000	<0.001	<0.001	>999.999
h9g3	1.0000	<0.001	<0.001	>999.999
h9g4	1.0000	1.025	<0.001	>999.999

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
het	Score	2.5882	0.1208	0.0917
	Probability	0.0582	0.1208	0.0917
h9	Score	0.3808	1.0000	0.6458
	Probability	0.7083	1.0000	0.6458
g2	Score	0.8703	0.4929	0.4015
	Probability	0.1829	0.4929	0.4015
g3	Score	1.1055	0.4606	0.3707
	Probability	0.1798	0.4606	0.3707
g4	Score	0.1066	1.0000	0.8620
	Probability	0.2760	1.0000	0.8620
white	Score	0.7825	0.4206	0.3377
	Probability	0.1657	0.5718	0.4889
hetg2	Score	0.000663	1.0000	0.8140
	Probability	0.3719	1.0000	0.8140
hetg3	Score	1.0797	0.5163	0.3689
	Probability	0.2948	0.5163	0.3689
hetg4	Score	1.9973	0.2508	0.1627
	Probability	0.1762	0.2508	0.1627
h9g2	Score	1.3756	0.4321	0.2389
	Probability	0.3864	0.4321	0.2389
h9g3	Score	0.4326	1.0000	0.6510
	Probability	0.6980	1.0000	0.6510
h9g4	Score	.#	.	.
	Probability	1.0000#	1.0000	0.5000

NOTE: # indicates that the conditional distribution is degenerate.

The SAS System

The LOGISTIC Procedure

Exact Conditional Analysis

Exact Parameter Estimates

Parameter	Estimate	95% Confidence Limits		p-Value
het	-0.8636	-2.1154	0.3133	0.1747
h9	-0.8035*	-3.4202	Infinity	1.0000
g2	0.6785	-0.9242	2.5699	0.5600
g3	0.8685	-0.9198	3.2484	0.5017
g4	0.2439	-1.3937	2.1580	1.0000
white	-0.5711	-2.2985	0.7487	0.5652
hetg2	-0.0243	-2.3712	2.1906	1.0000
hetg3	1.3292	-1.9207	5.6521	0.6709
hetg4	1.7229	-1.2053	5.9736	0.3876
h9g2	-0.4359*	-Infinity	2.5630	0.7727
h9g3_	0.8379*	-Infinity	4.5015	1.0000
h9g4	.#	.	.	.

NOTE: # indicates that the conditional distribution is degenerate.
* indicates a median unbiased estimate.

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		p-Value
het	0.422	0.121	1.368	0.1747
h9	0.448*	0.033	Infinity	1.0000
g2	1.971	0.397	13.064	0.5600
g3	2.383	0.399	25.748	0.5017
g4	1.276	0.248	8.654	1.0000
white	0.565	0.100	2.114	0.5652
hetg2	0.976	0.093	8.941	1.0000
hetg3	3.778	0.146	284.897	0.6709
hetg4	5.601	0.300	392.929	0.3876
h9g2	0.647*	0	12.975	0.7727
h9g3_	2.312*	0	90.151	1.0000
h9g4	. #	.	.	.

NOTE: # indicates that the conditional distribution is degenerate.
* indicates a median unbiased estimate.

BIBLIOGRAPHY

Ascher, J. A., J. O. Cole, et al. (1995). "Bupropion: a review of its mechanism of antidepressant activity." J Clin Psychiatry **56**(9): 395-401.

Epstein, L. H., J. L. Jaroni, et al. (2002). "Dopamine transporter genotype as a risk factor for obesity in African-American smokers." Obes Res **10**(12): 1232-40.

Epstein, L. H., S. M. Wright, et al. (2004). "Relation between food reinforcement and dopamine genotypes and its effect on food intake in smokers." Am J Clin Nutr **80**(1): 82-8.

Flore., M. (2000). "Trating tabacco use and dependence: an introdiction to the US Public Health Service Clinical Practice Guideline. ." Respir Care **45**: 1196-1199.

Galili-Weisstub, E., S. Levy, et al. (2005). "Dopamine transporter haplotype and attention-deficit hyperactivity disorder." Mol Psychiatry **10**(7): 617-8.

Haxby, D. G. (1995). "Treatment of nicotine dependence." Am J Health Syst Pharm **52**(3): 265-81; quiz 314-5.

Hutz, M. H., S. Almeida, et al. (2000). "Haplotype and allele frequencies for three genes of the dopaminergic system in South American Indians." Am J Human Biol **12**(5): 638-645.

Jorenby, D. E., S. J. Leischow, et al. (1999). "A controlled trial of sustained-release bupropion, a nicotine patch, or both for smoking cessation." N Engl J Med **340**(9): 685-91.

Kelada, S. N., H. Checkoway, et al. (2006). "5' and 3' region variability in the dopamine transporter gene (SLC6A3), pesticide exposure and Parkinson's disease risk: a hypothesis-generating study." Hum Mol Genet **15**(20): 3055-62.

Kim, J. W., B. N. Kim, et al. (2006). "The dopamine transporter gene and the impulsivity phenotype in attention deficit hyperactivity disorder: a case-control association study in a Korean sample." J Psychiatr Res **40**(8): 730-7.

Leighton, P. W., D. G. Le Couteur, et al. (1997). "The dopamine transporter gene and Parkinson's disease in a Chinese population." Neurology **49**(6): 1577-9.

Lerman, C., N. E. Caporaso, et al. (1999). "Evidence suggesting the role of specific genetic factors in cigarette smoking." Health Psychol **18**(1): 14-20.

Lerman, C., C. Jepson, et al. (2006). "Role of functional genetic variation in the dopamine D2 receptor (DRD2) in response to bupropion and nicotine replacement therapy for tobacco dependence: results of two randomized clinical trials." Neuropsychopharmacology **31**(1): 231-42.

Mitchell, R. J., S. Howlett, et al. (2000). "Distribution of the 3' VNTR polymorphism in the human dopamine transporter gene in world populations." Hum Biol **72**(2): 295-304.

Need, A. C., K. R. Ahmadi, et al. (2006). "Obesity is associated with genetic variants that alter dopamine availability." Ann Hum Genet **70**(Pt 3): 293-303.

Perkins, K. A., M. D. Marcus, et al. (2001). "Cognitive-behavioral therapy to reduce weight concerns improves smoking cessation outcome in weight-concerned women." J Consult Clin Psychol **69**(4): 604-13.

Sabol, S. Z., M. L. Nelson, et al. (1999). "A genetic association for cigarette smoking behavior." Health Psychol **18**(1): 7-13.

Shinohara, M., H. Mizushima, et al. (2004). "Eating disorders with binge-eating behaviour are associated with the s allele of the 3'-UTR VNTR polymorphism of the dopamine transporter gene." J Psychiatry Neurosci **29**(2): 134-7.

Vandenbergh, D. J., A. M. Persico, et al. (1992). "Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR." Genomics **14**(4): 1104-6.

Vieyra, G., M. Moraga, et al. (2003). "[Distribution of DRD4 and DAT1 alleles from dopaminergic system in a mixed Chilean population]." Rev Med Chil **131**(2): 135-43