Preliminary and Incomplete: Do Not Cite or Circulate

Genetic Modulation of the Effects of Tobacco Taxation on Use[1]

Jason Fletcher
Yale University
Associate Professor of Public Health

Abstract
The reduction in tobacco use as a result of taxation has been considered one of the most important public health successes in the past century. However, individuals continue to smoke at high rates and there is evidence of substantial heterogeneity in the responses to taxation. One of the key determinants of tobacco use is genetic susceptibility, yet important policies to reduce tobacco use have not successfully merged this risk factor in targeting interventions. This paper extends the standard economic framework that has evaluated tobacco taxation effects by presenting the first evidence in the literature that specific genetic polymorphisms moderate the effects of taxation on tobacco consumption. The evidence suggests that taxation only affects smoking participation decisions of individuals with a specific genotype—a polymorphism of a nicotinic receptor gene—and has no effect on others. Additionally, the results can be interpreted to be broadly consistent with the idea that policy variation can affect the expression of underlying genetic endowments and may serve to provide further support for "rational" models of addictive behavior.

[1] jason.fletcher@yale.edu.
The author thanks Phil DeCicca, Zoe Donaldson, Bruce Link, Tal Gross, Randall Reback and numerous seminar participants at McMaster University, Cornell University, Guelph University, University of Toronto, Columbia University, Tulane University, and Louisiana State University for helpful comments.
This research has been supported by a seed grant from the Robert Wood Johnson Foundation Health & Society Program at Columbia University. The author thanks the Robert Wood Johnson Foundation Health & Society Scholars program for its financial support.
The author thanks Karen Davis for help creating the data files and for answering questions about the data and Ajay Yesupriya for expeditious review of the files.
The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of the Research Data Center, the National Center for Health Statistics, or the Centers for Disease Control and Prevention.

Introduction

Tobacco use is among the most important causes of mortality and morbidity in the US and around the world. Although policy and medical interventions have targeted tobacco use for the past 50 years, a large population of individuals continues to initiate and fails to quit smoking. The difficulty of achieving additional reductions in smoking rates as well as the heterogeneity in response to each policy and medical intervention suggests that merging information and research from the biological and social sciences may be an important next step in further attempts of reducing use. From genetics and the biological sciences, there is a history of findings that suggest specific genes are directly responsible for some smoking phenotypes; however, even the cumulative impact of the genes fail to explain substantial variation in behavior. As a response to these findings, an emerging approach has sought to investigate potential gene-environment interactions in order to both explore new sources of variation and highlight the complexity of substance use outcomes. At the same time, from health policy and economics, large literatures have described the effects (and heterogeneity of effects) of the rapid expansion of tobacco taxation and other policies but have also not been able to outline many of the key determinants of tobacco use. With these clear limitations from the different sets of literatures across disciplines, additional understanding may require a more substantive merging of these and other findings and methodologies. However, economic analysis that combines a genetic perspective to further understand the biological determinants and sources of heterogeneous impacts of tobacco policies has not been undertaken.

The purpose of this paper is to present the first evidence in the literature of the existence of gene x environment (policy) interactions using a national sample of adults from the US. To do this, this paper focuses on a particular genetic variant thought to operate along the biological pathway between nicotine exposure and brain/body response. Using variation across states and over time in the state-level cigarette tax rate, this paper provides evidence that individuals at low genetic risk respond to tobacco taxes while individuals at higher risk have no such response. Importantly, the "low risk" individuals are approximately 50% of the sampled population. These results are robust to the inclusion of state fixed effects and measures of social norms of smoking. The implications of this interaction are several-fold. First, additional policies may need to be employed to more adequately target the "high risk" non-responders in order to reduce tobacco use in this population. The findings also suggest that state-level policies may (unintentionally) have the capacity to affect the expression of genetic endowments, which opens the possibility of explicitly crafting policies for this purpose in the future as well as increasing our understanding of the biological processes affecting substance use decisions. While this analysis is an initial step in the process of combining economics and genetics to further our understanding of tobacco use patterns, additional research is needed to both replicate as well as extend these basic results.

Background Literature

There are many well-documented determinants of substance use, including genetic endowments, governmental policies, and social, psychological and demographic factors, however much of the variation in use and dependence is still

unaccounted for in standard empirical models. For example, there is very little doubt that genetics plays an important role in substance use (Tyndale 2003, Dick and Toroud 2003). Family-based studies have demonstrated that approximately 50-60% of the variance in alcohol dependence can be accounted for by genes (McGue 1999). Twin studies have provided evidence of large genetic influences on many other dimensions of alcohol use such as quantity consumed, frequency of use, alcohol metabolism, and other factors (Heath 1995, Tyndale 2003). Likewise, some studies suggest that as much as 70% of the variance in nicotine dependence is due to genetic factors (Sullivan and Kendler 1999).

Although genetics is a key component determining tobacco use and cessation, few large scale policies have attempted to leverage this fact. Indeed, arguably the most successful public policy in the past century at increasing public health has taken a broad brush to achieve its aims—large increases in tobacco taxes. Tobacco taxes have been show to be responsible for large reductions in tobacco use. For example, Chaloupka (1998) suggests that a $1.50 increase in cigarette taxes and prices would reduce overall consumption by about thirty percent and cut youth smoking rates by nearly half. On the other hand, tobacco taxation has not proven to be a silver bullet in the fight against tobacco consumption. The smoking rate for adults was over 20%[2] and over 17%[3] of high school students reported smoking in 2009. In addition to the seemingly entrenched use by a sizable proportion of the population, there is also substantial heterogeneity in the responsiveness to taxation. For example, researchers have found higher responses for the poor (Farrelly and Bray 1998), girls (Lewit et al. 1997), women

[2] http://www.cdc.gov/Features/VitalSigns/TobaccoUse/
[3] http://www.cdc.gov/tobacco/data_statistics/fact_sheets/youth_data/tobacco_use/index.htm

(Chaloupka and Warner 2001), whites teens (DeCicca et al. 2000) and younger teens (Gruber 2000). Both the difficulty of attaining additional reductions in smoking rates as well as the unexplained and relatively unexamined heterogeneity in response suggests additional research is needed for new or retooled policy interventions that might achieve still lower rates of smoking. This paper seeks to combine knowledge from the social and biological sciences to increase our understanding of continued tobacco use and ways to potentially enhance the typical policy response by incorporating genetic variation into a standard policy analysis. This approach leverages a gene-environment framework that has grown rapidly in the biological sciences but has not been used in health economics.

In the biological sciences, a relatively new and rapidly increasing focus in determining substance use patterns is on *interactions* between genetic endowments and environmental factors. The classic example of this research design is from Caspi et al. (2003), where an interaction between differences in the 5-HTT gene and exposure to life stress was examined as a determinant of depressive symptoms in a cohort of white male New Zealanders. The authors reported that the main effect of maltreatment predicted depression, the main effect of the gene did not, but the interaction of the gene and environmental exposure was statistically significant, indicating a gene X environment (GxE) interaction. This general approach is useful because it can both outline important considerations in the heterogeneity of responses to environmental factors (and policies) and can also further develop our understanding of the critical biological pathways through which substance use is initiated and maintained.

One important disadvantage of much research that focuses on GxE interactions is the use of *endogenous* aspects of the environment. It is well known that gene-environment correlation (rGE) can mask true causal pathways as one is uncertain whether the genetic endowment affects *response* to the environmental factor or an increased *risk of exposure* to the environmental factor. Analysis that examines interactions between non-exogenous environmental factors and genetic variation are also limited by the potential of gene-gene interactions. This issue has recently been raised by several researches (e.g. Conley 2009), where the need to focus on *exogenous* environmental factors is suggested. This paper follows this new preferred methodology and focuses on adult decisions to use substances by examining the interactions between genetic endowments and economic and social policies. This research design increases the power of detecting GxE influences because it limits the influence of gene-environment correlation, which has problematized GxE work for many years. This paper utilizes this new direction of GxE research by focusing on *exogenous* environmental exposures, in the form of state-level tobacco policies.

While the "E" in the GxE approach in this application is straightforward, focusing on a specific measure of "G" is somewhat less so. Although genetic influences are strongly implicated in many dimensions of substance use and misuse, detecting the specific genes responsible has been difficult, partly because there could be different genes involved in different dimensions of substance use and environmental interactions can make it difficult to estimate the direct effects of genes. However, there is some convincing evidence for the role of specific genes in substance use behaviors. Several candidate genes in neurobiological pathways that play a role in nicotine's reinforcing

and addictive effects, such as dopamine and serotonin, have also been found, although nonreplication has been an important limitation in much current work (IOM 2006, p. 71).

This paper will focus on a single nucleotide polymorphism (SNP) in the genetic code as the measure of "G" in the GxE framework. A SNP is measured by variation in a single letter in the chain of over 3 billion letters (either A,G,C,T) that make up the human DNA code, lying within one of the over 25,000 genes, which itself lies within one of the 46 chromosomes. In particular, this study uses variation within the CHRNA6 gene, which is located on chromosome 8, to examine differential responses to tobacco taxation. The CHRNA6 (**Ch**olinergic **r**eceptor, **n**icotinic, **a**lpha **6**) gene encodes an alpha subunit of neuronal nicotinic acetylcholine receptors. This gene is among a class of genes, which are the primary targets for nicotine in the brain, involved in nicotine-related behaviors and is in the family of nicotinic acetycholine receptors (nAChRs), where each subunit is encoded by a single gene (Mineur and Picciotto 2008).[4] [5]

Briefly, the gene has been shown to mediate the "pleasure" (transmission of dopamine in the brain) from exposure to nicotine.[6] This (and related) genes have been shown in numerous studies to be related to tobacco use and dependence. Stevens et al. (2008) shows associations between the CHRNA5-CHRNA3-CHRNB4 gene cluster and heavy smoking. More specifically, Hoft et al. (2009) show a replicated association

[4] See Dajas-Bailador ad Wonnacott (2004) for general overview of neuronal nicotinic acetylcholine receptors.

[5] Mineur and Picciotto (2008) discuss that nicotine is believed to act in part through activation of the mesocorticolimbic system and that activation of nAChRs on dopaminergic neurons of the ventral tegmental area (VTA) increases their firing rate and stimulates dopamine release from their terminals in the nucleus accumbens. Interestingly lesions of nicotinic antagonists into the VTA have been shown to prevent the development of behaviors related to nicotine addiction, particularly using animal models.

[6] More formally, these receptors consist of five subunits and function as ion channels involved in neurotransmission. The encoded protein is a subunit of neuronal nicotinic acetylcholine receptors that mediate dopaminergic neurotransmission and are activated by acetylocholine and exogenous nicotine. See: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=8973

between the CHRNA6 gene with tobacco dependence in two samples, see also Zeiger et al. (2008). Saccone et al. (2007) and Greenbaum and Lerer (2009) show that the specific SNP used in this study (rs2304297), is associated with tobacco use outcomes in several samples.

In addition to human studies, genetic engineering studies using animal (mouse) models have identified a number of subunits that are critical for activation of the reward system in the brain following nicotine exposure (Mineur and Picciotto 2008). In particular, Champtiaux et al. (2002) show, using knockout mice, that the alpha-6 subunit plays a role in addiction related behaviors, including nicotine induced dopamine release.[7]

Thus, this paper will combine findings from the genetics/biology and economics/policy literatures in order to examine whether there are gene-environment interactions in determining tobacco use in a national sample of adults from the US.

Data and Empirical Methods

Even with the abundance of genetic data in the biological sciences and the rapidly growing number of datasets from the social sciences that include genetic information, surprising few datasets contain the necessary information to examine a potential gene-taxation interaction in determining tobacco use. The core requirements include measurements of (1) tobacco use, (2) tobacco-related genes, and (3) state tobacco tax rates. Unfortunately, most datasets cannot accommodate these requirements. Many datasets from the biological sciences contain (1) and (2) but are

[7] While nAChRs are the primary targets for nicotine, they are also expressed in most tissues and organs and can be detected on presynaptic terminals, cell bodies and dendrites of many neuronal subtypes (see Dajas-Bailador and Wonnocott 2004).

convenience samples from a small geographic area (e.g. a county or city) and thus do not have variation in (3). On the other hand, many social science datasets contain (1) and (3) but do not yet have relevant information for (2). Therefore, this paper uses Phase II of the National Health and Nutrition Examination Survey (NHANES) III dataset, which is a nationally representative sample of individuals from 1991-1994[8]. Although the NHANES is available from the early 1970s to the present, only the 1991-1994 data has been genotyped at this time, and only for a small number of genes. Luckily, other researchers have made the investment to genotype a specific polymorphism in the CHRNA6 gene that has been linked with tobacco use. As discussed above, the type of variation available in the NHANES III Data for this gene is a single nucleotide polymorphism (SNP) (rs2304297), meaning there is a single letter change in the genetic code (in this case either a C (cytosine) or a G (guanine)).

In addition to containing genetic data, the NHANES also is a survey that asks respondents to report participation and frequency of use of tobacco, but this information is primarily collected for individuals over the age of 18, which will be the focus of this study. Specifically, this analysis will focus on smoking participation, number of cigarettes consumed per day, and a separate measure of nicotine taken from blood samples—A major component of the NHANES study is the collection of biological specimens, including blood and urine. These specimens are then analyzed in order to create objective measures of exposure to nicotine (which can include secondhand

[8] The Add Health data contains similar information as the NHANES, but the sample size of the Add Health is only a third of the size as the NHANES. Even so, it would be useful to use alternative datasets to examine the robustness of the findings in this paper.

smoke) through cotinine[9], which is an alkaloid found in tobacco and a metabolite of nicotine. Several papers in the economics literature have used cotinine as an objective measure of tobacco use (e.g. Adda and Cornaglia 2006, 2010). Finally, in order to merge the NHANES with state-level cigarette tax information, the confidential version of the NHANES is used.[10]

Summary statistics for the sample with genetic information is provided in Table 1 below. Predictors of whether an individual agreed to participate in genotyping are provided in Table 1A in the appendix—approximately 6,500 of the 10,000 respondents from Phase II have genotypic data. I find very small differences in the availability of genotypic data based on demographics and previous smoking behavior, suggesting that the sample is approximately a random subset of the national NHANES sample.

Table 1 shows that during this period, 25% of individuals aged 17-90 report smoking tobacco. The average number of cigarettes consumer per day is nearly four for the full population and is over 15 for smokers. The average serum cotinine level is 68.4 ng/mL—a typical level used to distinguish smokers from non-smoker is 14 ng/mL (Florescu et al. 2009).[11] The variation in state-level smoking is between 17-34% and the cigarette tax is on average $0.24 per pack. There is also substantial genetic variation in the sample. Eleven percent were genotyped as “CC”, 38% were “CG” and 51% were “GG” for the rs2304297 SNP.

[9] Cotinine has a half-life of approximately 20 hours and is able to be detected for several days after the use of tobacco. The level of cotinine in the blood is proportional to the amount of exposure to tobacco smoke (Florescu et al. 2009).

[10] Although the confidential data is generally available for use in any of the Census Restricted Data Centers (RDC), the genetic data is only available for use at the National Center for Health Statistics in Hyattsville, MD or the Centers for Disease Control and Prevention in Atlanta, GA.

[11] Over 92% of the self-reported non-smokers in the data have levels of cotinine below 14. The average cotinine level in the sample for smokers is 219, while for non-smokers it is 17 (the median value for non-smokers is 0.2).

One important issue in examining gene-environment interactions is that of population stratification. It could be the case that the distribution of genetic variants differ by population subgroups (e.g. by race/ethnicity) and these differences in the distribution of genetic variation could be correlated with differences in environmental exposures.[12] Table 2 shows evidence of substantial population stratification in the SNP of interest, where 57% of the white population are carries of the GG variant and only 7% of the black population are carries of the GG variant. Thus, we might be worried that if exposure to high (or low) tax levels is correlated with race, the standard framework may provide spurious results. The two primary methods of overcoming this problem are to control for race/ethnicity in the results as well as examine the robustness of the results when they are stratified by race/ethnicity. The application of either of these methods does not change the main results (shown below). I also extend the analysis below by explicitly showing that the genetic variants are uncorrelated with the policies in the data.

Conceptually, this paper follows the standard gene-environment approach of modeling tobacco use/initiation as a function of vectors of individual, family, and environmental level characteristics and interactions:

$$Y = f(X, E, G, GxE) \quad (1)$$

where the determinants of use are a function of genetic endowments (G), environmental factors (E), and their interactions (GxE). Combining this approach with the standard economic framework of analyzing the effects of tobacco taxation (Chaloupka and Warner 2000), I estimate specifications of the following form:

[12] The classic example of population stratification is the story of the "chopstick" gene (Hamer and Sirota 2000). Here the point is that any genes that are more prevalence in Asian populations will seem to be predictive of chopstick use (as the outcome of interest). However, the clear confounder is Asian ethnicity and culture.

$$Smoking_{ist} = \beta_0 + \beta_1 Tax_{st} + \beta_2 SNP_i + \beta_3 Tax_{st} * SNP_i + \beta_4 X_i + \tau_t + \varepsilon_{ist} \quad (2)$$

where the tax rate will be entered in log form, the SNP reflects variation in the CHRNA6 gene, the X vector controls for demographic variables (age, race, gender), year fixed effects are controlled, and the error term is clustered at the state level. The coefficient of interest is β_3 , which tests whether there is evidence of gene-environment interaction in predicting tobacco use outcomes.

A limitation with the NHANES data is the short panel (1991-1994) available, whereas most investigations control for state-level fixed effects in order to separate the effects of taxation with other state-level unobservables (DeCicca et al. 2002). As sensitivity analyses, this paper will estimate two auxiliary specifications

$$Smoking_{ist} = \beta_0 + \beta_1 Tax_{st} + \beta_2 SNP_i + \beta_3 Tax_{st} * SNP_i + \beta_4 X_i + \tau_t + \lambda_s + \varepsilon_{ist} \quad (3)$$

$$Smoking_{ist} = \beta_0 + \beta_1 Tax_{st} + \beta_2 SNP_i + \beta_3 Tax_{st} * SNP_i + \beta_4 X_i + \tau_t + \beta_5 Z_s + \varepsilon_{ist} \quad (4)$$

Where equation (3) includes state-fixed effects and equation (4) controls for state level smoking rates in the regression. The estimates are not sensitive to these specifications.

Results

Table 3 reports the primary findings of this paper. The first column reports the unconditional effect of higher taxes of smoking participation—a 100% increase in the tax rate is associated with a 3.1 percentage point reduction in smoking; the baseline rate is 25%. The second column reports the unconditional effect of carrying the GG genetic variant, which reduces smoking participation by more than that of a 100% tax increase. Column three shows that the results are relatively robust to including both tax

and SNP variables. Column 4 presents the main coefficient of interest, the gene-environment interaction. The results suggest that, for individuals with either the CC or CG genotypes, the participation elasticity is essentially zero. However, for those "low risk" individuals with the GG genotype, which is approximately 50% of the population, the reduction in the likelihood of reporting smoking is nearly 7.3 percentage points.

Columns 5-7 show that this finding is robust to the inclusion of control variables, state fixed effects, and state-level smoking rates. In particular, Column 5 controls for individual level demographic and socioeconomic characteristics. Column 6 adds the "State Anti-Smoking Sentiment" measure used by DeCicca et al. (2002).[13] Because the NHANES data includes only a short panel of states, the primary analysis is unable to use state-level fixed effects, as is common in the literature. Work by DeCicca has shown that, in these cases, the SASS measure provides a good approximation to typically unobserved state-level factors that may co-vary with state tobacco tax rates. The results suggest the results are insensitive to this addition. Finally, column 7 includes state-level fixed effects (with the caveat of a short panel) as well as state-specific smoking rates (aggregated from the NHANES data). Again, the results suggest the core finding is robust across these specifications. Appendix Table 2A shows that these results are similar if the analyses are stratified by genotype and estimated separately.

This finding suggest important heterogeneity in the response to tobacco taxation based on a single change in the genetic code related to nicotine metabolism in the brain and body. Column 5 also places the importance of the genetic risk in perspective—carriers of the GG genotype have an increased risk of smoking that is 50% of the

[13] The author thanks Phil DeCicca for generously providing these data.

magnitude of the black-white difference in smoking rates, is similar in magnitude to a (conditional) 1.5 increase in years of schooling, or the conditional increase of over $10,000 in income per year.

This interaction has never been reported in the extant literature but has several implications. The results suggest that tobacco tax policy may be completely irrelevant to a large segment of the population, who are characterized based on a "high risk" genotype. The implication is that additional policies will be necessary to more effectively reduce smoking rates for these entrenched smokers. The flipside of the first implication is that these results suggest that state-policy variables may be able to modify the expression of genetic risk. The mechanisms for this empirical finding cannot be uncovered based on these data. One possibility, following DeCicca et al. (2002), is that taxation is partially capturing differences in social norms across states. However, the magnitude of the interaction is not dampened in Column 7 after the inclusion of state-level smoking rates or in Column 5 with the use of the SASS measure. An alternative may be that there are threshold effects (or other non-linearities) from taxation, and those who gain the most from smoking are insensitive to taxes at low levels. However, in results presented in Appendix Table 4A, these non-linearities do not appear in the data over the range available (2-50 cents per pack).

Each of these implications require additional research with alternative datasets and methods, but the main finding of an interplay between genetic endowment and state-level policy suggests that this and related future research could be very illuminating about both the fundamental and complex determinants of substance use as well as in making potential policy recommendations.

Additional Results

Table 4 presents results for alternative measures of tobacco use. Column 1 shows that the results are similar if I use an objective measure of tobacco exposure, serum cotinine, where only those individuals with the "low risk" genotype are affected by the tax level. Here the black-white differences are not as stark as the self-reported differences in smoking rates; one explanation for this difference, which is common in the literature, is that menthol cigarettes contribute higher levels of cotinine than non-menthol cigarettes, and since blacks are more likely to smoke menthol cigarettes, their levels of cotinine are higher than whites. On the other hand, the magnitude of the reduction in cotinine associated with taxation for those with the GG genotype is similar to 1.5 years of schooling, like the previous table. Column 2 shows that the number of reported cigarettes is affected in the same way as smoking participation and the level of measured cotinine, where only individuals with the GG genotype seem to be affected by tobacco taxation.

Table 5 presents the results from Table 3 stratified by important demographic subgroups. Columns 1-3 present the results stratified by race/ethnicity. While the results are qualitatively the same across groups, the results for blacks are considerably weaker than for whites. One explanation of this finding is the relatively small number of blacks in the sample with the low risk genotype. An alternative explanation is that blacks could respond differently to taxation than whites. In contrast, Column 3 and 4 shows that the results for Hispanic and "Other Race" individuals largely mirror those for whites. Columns 5 and 6 show evidence of no gender differences in the main results.

In Panel 2, the results are stratified based on age categories and education level. With the exception of the age 50-65 group, the findings are quite robust across the columns. Overall, the robustness of the results across important demographic sub-groups of the population strengthens the main findings and suggests important interaction between genotype and tobacco policy in determining tobacco use patterns in the data. In the final pair of columns, the findings suggest that poor individuals (<median income) have large elasticities based on both tax rates as well as genetics, where the implication is that individuals with more resources may be able to protect themselves from their genetic risk of smoking and are also less responsive to taxes.

Table 6 examines an additional potential issue with the framework—that of gene-environment correlation, where "genes select environments." This could occur if individuals with genetic predispositions choose to move to states with low tobacco taxes or, potentially more likely, that there may be some correlation between state level taxes and some population characteristics, such as race/ethnicity, that are also correlated with genotype. Table 6 examines this issue and shows that there is some evidence of correlations between the "risky genotype" and state level tax levels, although the correlation is positive, where individuals at higher risk for tobacco use live in states with higher tax levels. In Column 3, the results suggest that this correlation is indeed a result of other individual level variables (e.g. race/ethnicity) and that any remaining associations are small and statistically insignificant (about ½ of one cent).

Finally, although there are no clear potential replication samples available for these results, I take a step in this direction by leveraging the repeated cross-sectional nature of data and form three separate "replication samples" based on the year of the

survey. In each year, a new set of individuals are sampled, creating a "quasi-replication" opportunity over similar populations. Thus the principle analysis is undertaken on each year of data separately, and the results are presented in Appendix Table 3A. Overall, the results are highly consistent in each of the three years of the data, where the GG genotype is protective for current tobacco use and only those with this genotype appear to reduce their smoking likelihoods as taxes are increased.

Conclusions

Although tobacco use has fallen dramatically in the past half a century, there continues to be a large share of the population who initiates and is unable or unwilling to quit as they age. These facts contribute to tobacco's remaining status as the leading preventable cause of death in the US, with over 400,000 American deaths each year (NIDA 2009). The underlying inability for taxation to further reduce use is of great concern for both theory and policy. Researchers are still attempting to uncover all the critical determinants of initiation and continued use and policymakers struggle to leverage the determinants in new policies to further reduce use and increase cessation. One area that has received little attention is combining biological and social science methods and findings to further examine the determinants of use and the responses to policy. This paper presents the first evidence in the literature of the existence of interaction effects between state-level tobacco policy and genetic susceptibility in predicting patterns of tobacco use. The results suggest that carriers of a particular variant of a nicotinic receptor gene respond to taxation as economists and policymakers would predict—by reducing participation and consumption. However, carriers of a

different variant in the same gene appear to be completely unresponsive to taxation. These results are robust to several auxiliary analyses but also need to be verified in independent samples. Unfortunately, at present, few available datasets contain all the information necessary to accommodate a replication exercise.

While the results are unique and quite interesting, there are a few caveats to consider. First, this analysis was opportunistic, and thus was limited to considering a SNP that had already been sequenced in the data. Although this SNP has considerable scientific credibility from human and animal studies as an important component in the determination of tobacco use and dependence, genotyping alternative SNPs based on additional findings from the genetic literature would be a useful next step. Even with the reductions in cost of genotyping, analyzing nearly 6,500 specimens remains expensive and time consuming. Similarly, the reader should be reminded that the particular SNP under investigation in this study could be in linkage disequilibrium (LD) with other, nearby SNPs (see Hoft et al. 2009). While this circumstance is an important limitation in gene discovery exercises, where the precise identity of the causal SNP is required, in gene-environment investigations, issues of LD may not be as consequential for a proof-of-concept examination like the current paper. That is, the main point of this paper is the existence of GxE rather than the specific variant within the gene (or in LD with the gene) that is causing the interaction.

With these limitations in mind, the implications of the analysis in this paper for both the determinants of tobacco use and the potential for policy intervention are substantial. For the first time in the literature, this paper uncovers a potential source for the important lack of response to tobacco taxation by large sub-groups of the population

found in the literature. Understanding these sources of response and non-response should prove useful when crafting additional policies and retooling current policies to reduce tobacco consumption. A second implication of the findings is the potential for policies to shape the expression of genetic endowments and predispositions. That is, results suggest that for individuals who live in low tax states, genetic predispositions do not differentially affect smoking rates; however, the effects of these genetic differences do appear in higher tax states. These findings are then consistent with the idea that the policy environment could shape the expression of underlying genetic liabilities—reducing the effects of genetics in some instances and amplifying them in other instances. Thus, the consequence of policies may well extend farther than initially thought and may even have epigenetic consequences. Future research should verify these findings in alternative datasets and for additional phenotypes of interest.

References

Adda, J. and F Cornaglia. (2006). "Taxes, Cigarette Consumption and Smoking Intensity", *American Economic Review*, 96, 4, 1013-1028,

Adda, J. and F Cornaglia. (2010). "The Effect of Taxes and Bans on Passive Smoking". *The American Economic Journal-Applied Economics.* Vol. 2, No. 1, 1-32.

Carmelli, D. G.E. Swan, D. Robinette and R. Fabsitz, Genetic influence on smoking—a study of male twins, *N Engl J Med* 327 (1992), pp. 829–833.

Caspi, Avshalom, Karen Sugden, Terrie E. Moffitt, Alan Taylor, Ian W. Craig, HonaLee Harrington, Joseph McClay, Jonathan Mill, Judy Martin, Antony Braithwaite, Richie Poulton. (2003). Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene." *Science* 301 (5631): 386-389

Chaloupka FJ. *The Impact of Proposed Cigarette Price Increases.* Policy Analysis No. 9, Health Sciences Analysis Project. Washington: Advocacy Institute, 1998.

Chaloupka FJ and Warner KE. (2000). *The economics of smoking*. In: The Handbook of Health Economics 1B. JP Newhouse and AJ Culyer (eds.) Amsterdam: North-Holland, pp. 1539–1612.

Champtiaux, N., Z.Y. Han, A. Bessis, F.M. Rossi, M. Zoli and L. Marubio *et al.*, Distribution and pharmacology of alpha 6-containing nicotinic acetylcholine receptors analyzed with mutant mice, *J Neurosci* **22** (2002), pp. 1208–121

Conley, D. F.C. (2009). "How (Not) to Study Genes as a Social Scientest or Learning to Love Animal Models," in Bernice Pescosolido (ed.) *Handbook of Medical Sociology*.

Dajas-Bailador F. and S. Wonnacott, Nicotinic acetylcholine receptors and the regulation of neuronal signalling, *Trends Pharmacol Sci* **25** (2004), pp. 317–324

DeCicca P, Kenkel D and Mathios A. (2000). Racial difference in the determinants of smoking onset. *Journal of Risk and Uncertainty* 21(2-3):311-340.

Dick D, Foroud T. 2003. Candidate genes for alcohol dependence: A Review of genetic evidence from human studies. *Alcoholism: Clinical and Experimental Research* 27(5):868-879.

Farrelly MC and Bray JW. (1998). Response to increases in cigarette prices by race/ ethnicity, income, and age groups - United States, 1976–1993. *JAMA* 280(23):1979-1980.

Fischer, R. Lung cancer and cigarettes, *Nature* 182 (1958), p. 108.

Fletcher, Jason M., Partha Deb, and Jody L. Sindelar. (2009). “Tobacco Use, Taxation, and Self Control in Adolescence.” NBER Working Paper 15130

Florescu A, Ferrence R, Einarson T, Selby P, Soldin O, Koren G (2009). "Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology". Therapeutic Drug Monitoring 31 (1): 14–30.

Greenbaum, L and B Lerer. (2009). Differential contribution of genetic variation in multiple brain nicotinic cholinergic receptors to nicotine dependence: recent progress and emerging open questions. *Molecular Psychiatry*, 14: 912-945

Gruber J. (2000). Youth smoking in the U.S.: prices and policies. NBER Working Paper No. W7506. Available at SSRN: http://ssrn.com/abstract=217832

Hamer, D and L Sirota. (2000). Population stratification is a potential source of error in psychiatric genetics. New study designs and statistical methods can help guard against this problem. *Molecular Psychiatry* 5, 11-13.

Heath AC. 1995. Genetic influences on drinking behavior in humans, in *The Genetics of Alcoholism* (Begleiter H, Kissin B eds), pp 82–121. Oxford University Press: New York. Institute of Medicine. 2006. *Genes, behavior, and the social environment: Moving beyond the nature/nurture debate*. (Hernandez l, Blazer D eds). Washington, DC: The National Academies Press.

Hoft, Nicole R, Robin P Corley, Matthew B McQueen, Isabel R Schlaepfer, David Huizinga and Marissa A Ehringer. Genetic Association of the CHRNA6 and CHRNB3 Genes with Tobacco Dependence in a Nationally Representative Sample. *Neuropsychopharmacology* (2009) 34, 698–706

Lewit EM, Hyland A, Kerrebrock N and Cummings KM. (1997). Price, public policy, and smoking in young people. *Tobacco Control* 6(suppl. 2):S17–S24.

McGue M. 1999. The behavioral genetics of alcoholism. *Current Directions in Psychological Science* 8:109–115.

Mineur YS, Picciotto MR (2008). Genetics of nicotinic acetylcholine receptors: relevance to nicotine addiction. *Biochem Pharmacol* 75: 323–333.

NIDA. (2009). Research Report Series: Tobacco Addiction. NIH Publication Number 09-4342

Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. Hum Mol Genet 2007; 16: 36–49.

Stevens, Victoria L., Laura J. Bierut, Jeffrey T. Talbot, Jen C. Wang, Juzhong Sun, Anthony L. Hinrichs, Michael J. Thun, Alison Goate, and Eugenia E. Calle. Nicotinic

Receptor Gene Variants Influence Susceptibility to Heavy Smoking. Cancer Epidemiol Biomarkers Prev December 2008 17:3517-3525
http://cebp.aacrjournals.org/content/17/12/3517.long

Sullivan, P.F. and K. S. Kendler (1999). The genetic epidemiology of smoking. Nicotine and Tobacco Research, 1(Suppl 2): S51-S57

Thorgeirsson, Thorgeir E et al. (2010). Sequence variants at CHRNB3–CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics*, 42: 448-453

Tyndale R. 2003. Genetics of alcohol and tobacco use in humans. *Annals of Medicine* 35:2: 94-121.

Zeiger, Joanna S., Brett C. Haberstick, Isabel Schlaepfer, Allan C. Collins, Robin P. Corley, Thomas J. Crowley, John K. Hewitt, Christian J. Hopfer, Jeffrey Lessem, Matthew B. McQueen, Soo Hyun Rhee, and Marissa A. Ehringer. The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco *Hum. Mol. Genet*. (2008) 17(5): 724-734

Tables

Table 1
Summary Statistics
NHANES 1991-1994 Genetic Sample (N~6,200)

Variable	Mean	Std Dev	Min	Max
Smoke	0.25	0.43	0	1
Number of Cigarettes	3.73	8.76	0	140
Numberof Cigarettes|Smoke	15.2	11.78	1	140
Cotinine	68.41	139.29	0.35	1890
Age	42.83	17.09	17	90
Female	0.52	0.50	0	1
Black	0.10	0.30	0	1
Hispanic	0.10	0.30	0	1
Other Race	0.05	0.22	0	1
Education	12.52	3.06	0	17
Income ($1000s)	35.32	18.39	0	60
Married	0.61	0.49	0	1
Missing Information	0.06	0.24	0	1
State % Smoke	0.25	0.04	0.17	0.34
Cigarette Tax	24.53	11.34	2.0	56.0
Year = 1991	0.02	0.15	0	1
Year = 1992	0.38	0.49	0	1
Year = 1993	0.31	0.46	0	1
Year = 1994	0.28	0.45	0	1
SNP:				
rs2304297=="CC"	0.11	0.31	0	1
rs2304297=="CG"	0.38	0.49	0	1
rs2304297=="GG"	0.51	0.50	0	1

Table 2
Racial Differences in Genetic Endowments

Full		Freq.	Percent	Cum.
	CC	1,278	11.46	11.46
	CG	2,328	38.41	49.88
	GG	2,572	50.12	100
	Total	6,178	100	
White		Freq.	Percent	Cum.
	CC	123	5.74	5.74
	CG	875	37.4	43.14
	GG	1,374	56.86	100
	Total	2,372	100	
Black		Freq.	Percent	Cum.
	CC	986	54.81	54.81
	CG	642	38.24	93.05
	GG	118	6.95	100
	Total	1,746	100	
Hispanic		Freq.	Percent	Cum.
	CC	163	7.36	7.36
	CG	775	45.92	53.28
	GG	1,017	46.72	100
	Total	1,955	100	

Percentages are weighted

Table 3
Gene-Environment Interactions in Predicting Tobacco Use

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	Tax Only	Gene Only	Both	Interaction	Xs	SASS	State FE/Smoking
Log (Tax)	-0.031*		-0.030*	0.002	0.016	0.009	0.014
	(0.017)		(0.016)	(0.013)	(0.014)	(0.014)	(0.012)
rs2304297=="GG"		-0.037*	-0.035*	-0.037**	-0.032*	-0.032*	-0.032
		(0.019)	(0.020)	(0.017)	(0.018)	(0.018)	(0.019)
Interaction				-0.073***	-0.072***	-0.073***	-0.076***
				(0.024)	(0.024)	(0.024)	(0.023)
Age					0.016***	0.016***	0.016***
					(0.003)	(0.003)	(0.003)
Age-squared					-0.000***	-0.000***	-0.000***
					(0.000)	(0.000)	(0.000)
Female					-0.064***	-0.064***	-0.064***
					(0.018)	(0.018)	(0.018)
Black					-0.062**	-0.063**	-0.068***
					(0.022)	(0.022)	(0.022)
Hispanic					-0.184***	-0.188***	-0.182***
					(0.046)	(0.047)	(0.047)
Other Race					-0.101*	-0.105*	-0.102*
					(0.053)	(0.054)	(0.054)
Education					-0.024***	-0.024***	-0.024***
					(0.004)	(0.004)	(0.004)
Income ($1000s)					-0.003***	-0.003***	-0.003***
					(0.001)	(0.001)	(0.001)
Married					-0.024*	-0.024*	-0.024*
					(0.013)	(0.013)	(0.014)
Missing Information					0.011	0.012	0.011
					(0.043)	(0.043)	(0.043)
SASS Measure						0.055	0.110*
						(0.063)	(0.053)
State Level Smoking							0.602**
							(0.238)
Constant	0.251***	0.270***	0.268***	0.270***	0.458***	0.469***	0.316**
	(0.015)	(0.017)	(0.018)	(0.018)	(0.111)	(0.111)	(0.139)
Observations	6178	6178	6178	6178	6163	6163	6163
R-squared	0.002	0.002	0.004	0.007	0.094	0.094	0.096

Robust standard errors in parentheses clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 4
Gene-Environment Interactions in Predicting Tobacco Use
Alternative Measures of Tobacco Use

Outcome	Number of Cigs	Cotinine
Log (Tax)	0.139	-3.046
	(0.344)	(3.129)
rs2304297=="GG"	-0.25	0.063
	(0.370)	(6.789)
Interaction	-0.782***	-12.188***
	(0.265)	(4.146)
Age	0.511***	5.792***
	(0.065)	(0.939)
Age-squared	-0.006***	-0.068***
	(0.001)	(0.010)
Female	-2.121***	-30.823***
	(0.429)	(5.472)
Black	-3.811***	-2.145
	(0.649)	(5.729)
Hispanic	-5.894***	-70.994***
	(1.106)	(15.340)
Other Race	-3.189***	-27.386
	(0.876)	(17.451)
Education	-0.557***	-8.777***
	(0.107)	(1.749)
Income ($1000s)	-0.072***	-0.596***
	(0.018)	(0.193)
Married	-0.219	-5.721
	(0.330)	(4.517)
Missing Information	0.073	-8.33
	(0.842)	(10.750)
Constant	7.880**	121.797***
	(2.753)	(39.737)
Observations	6,130	6,089
R-squared	0.096	0.085

Robust standard errors in parentheses clustered at the state level.
*** p<0.01, ** p<0.05, * p<0.1

Table 5
Gene-Environment Interactions in Predicting Tobacco Use
Results Stratified by Demographic Subgroups

Sample	White	Black	Hispanic	Other Race	Female	Male
Log (Tax)	0.011	-0.003	0.114*	0.152*	0.004	-0.003
	(0.020)	(0.010)	(0.064)	(0.075)	(0.014)	(0.020)
Genotype = GG	-0.028	-0.091**	0.075	-0.148**	-0.038*	-0.032
	(0.022)	(0.033)	(0.064)	(0.067)	(0.018)	(0.026)
Interaction	-0.082***	-0.023	-0.193	-0.191*	-0.070***	-0.072**
	(0.025)	(0.037)	(0.127)	(0.107)	(0.020)	(0.029)
Observations	2,372	1,746	1,955	286	3,524	2,654
R-squared	0.008	0.002	0.002	0.075	0.007	0.007

Sample	<30	30-50	50-65	65+	Less HS	HS +	Poor	Rich
Log (Tax)	-0.014	0.017	-0.012	0.022	-0.067**	0.02	0.018	-0.004
	(0.035)	(0.019)	(0.046)	(0.020)	(0.026)	(0.014)	(0.023)	(0.012)
Genotype = GG	-0.014	-0.061	-0.044	0.013	-0.013	-0.04	-0.051*	-0.027
	(0.035)	(0.045)	(0.043)	(0.025)	(0.031)	(0.025)	(0.026)	(0.021)
Interaction	-0.083	-0.090	0.016	-0.105	-0.074*	-0.082***	-0.116***	-0.051
	(0.049)	(0.063)	(0.035)	(0.063)	(0.041)	(0.028)	(0.026)	(0.038)
Observations	1,639	2,178	1,111	1,250	2,427	3,751	2,655	3,523
R-squared	0.013	0.01	0.003	0.015	0.018	0.008	0.013	0.004

Robust standard errors in parentheses clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 6
Correlations between Genotype and State Level Tax Levels

Outcome	State Tax Level	State Tax Level	State Tax Level
Specification	Basic Weighted	Clustered	Xs
Genotype = C/G	1.985***	1.985	0.58
	(0.577)	(1.449)	(0.574)
Genotype = G/G	2.143***	2.143	0.478
	(0.556)	(1.796)	(0.708)
Age			0.017
			(0.027)
Black			-1.771
			(2.486)
Hispanic			8.807***
			(2.224)
Other Race			2.402
			(1.753)
Education			0.243
			(0.159)
Income			0.027
			(0.029)
Missing Information			0.674
			(1.191)
Year = 1992			13.757***
			(4.051)
Year = 1993			15.542***
			(2.783)
Year = 1994			13.482***
			(3.371)
Constant	26.096***	26.096***	7.815**
	(0.483)	(3.697)	(3.093)
Observations	7,008	7,008	6,178
R-squared	0.004	0.004	0.113

Robust standard errors in parentheses clustered at the state level in columns 2 and 3.
*** p<0.01, ** p<0.05, * p<0.1. Genetic sample weights used.

Appendix Tables

Table 1A
Predictors of Availability of Genetic Data: NHANES III (1988-1994)

Outcome	Genes Available	Genes Available
Specification	OLS	Fixed Effects
Age	0.00219***	0.00226***
	(0.000644)	(0.000665)
Age-Squared	-3.22e-05***	-3.23e-05***
	(6.77e-06)	(6.69e-06)
Female	-0.00373	-0.00417
	(0.00395)	(0.00401)
Black	-0.0161	-0.00296
	(0.0131)	(0.0108)
Hispanic	-0.0108	0.0259**
	(0.0118)	(0.0108)
Other Race	-0.0144	-0.00123
	(0.0141)	(0.0141)
Education	0.00137*	0.00198**
	(0.000779)	(0.000764)
Income	0.000190	0.000298
	(0.000276)	(0.000245)
Married	0.0141***	0.0123***
	(0.00305)	(0.00348)
Missing	-0.0267***	-0.0252***
	(0.00696)	(0.00638)
Year=1989	0.00259	-0.00760
	(0.00495)	(0.0188)
Year=1990	0.000356	-0.0113
	(0.00579)	(0.0321)
Year=1991	0.0383*	0.0151
	(0.0221)	(0.0327)
Year=1992	0.588***	0.549***
	(0.0498)	(0.0380)
Year=1993	0.679***	0.666***
	(0.0373)	(0.0466)
Year=1994	0.788***	0.780***
	(0.0123)	(0.0154)
Smoke 100 Cigarettes?	-0.0121***	-0.0125***
	(0.00390)	(0.00415)
Constant	-0.0260	-0.0345
Observations	19,945	19,945
R-squared	0.526	0.536

Robust standard errors in parentheses clustered at the state level
. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2A
The Effects of State Level Taxation on Tobacco Use
Stratified by Genotype

Outcome	Smoke Now	Smoke Now	Smoke Now
Genotype	CHRNA6=CC	CHRNA6=GC	CHRNA6=GG
Log (Tax)	-0.016	0.014	-0.071**
	(0.018)	(0.017)	(0.029)
Constant	0.316***	0.256***	0.233***
	(0.025)	(0.024)	(0.016)
Observations	1,278	2,328	2,572
R-squared	0.001	0.001	0.012

Robust standard errors in parentheses clustered at the state level
. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 3A
The Effects of State Level Taxation on Tobacco Use
Stratified by Year of Survey

Outcome	Smoke	Smoke	Smoke
Sample	Year 1	Year 2	Year 3
Log (Tax)	-0.002	-0.013	0.019
	(0.017)	(0.047)	(0.017)
Genotype = GG	-0.039	-0.028	-0.053
	(0.032)	(0.018)	(0.035)
Interaction	-0.111***	-0.029	-0.050*
	(0.019)	(0.032)	(0.026)
Constant	0.283***	0.263***	0.268***
	(0.030)	(0.030)	(0.024)
Observations	1,682	2,367	2,038
R-squared	0.02	0.002	0.005

Robust standard errors in parentheses clustered at the state level
. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 4A
The Effects of State Tax Rates on Tobacco Use
Non-Linear Results Stratified by Genotype

Outcome	Smoke	Smoke	Smoke
Sample	Genotype = CC	Genotype = GC	Genotype = GG
Tax Rate (5-10)	0.035	0.063**	-0.196***
	(0.036)	(0.025)	(0.020)
Tax Rate (10-15)	-0.024	0.031	-0.200***
	(0.027)	(0.027)	(0.024)
Tax Rate (15-20)	0.172**	0.05	-0.281***
	(0.078)	(0.044)	(0.033)
Tax Rate (20-25)	0.007	0.089***	-0.155***
	(0.026)	(0.014)	(0.014)
Tax Rate (25-30)	0.048	0.172***	-0.265***
	(0.039)	(0.020)	(0.017)
Tax Rate (30-35)	0.115***	0.096**	-0.268***
	(0.038)	(0.040)	(0.028)
Tax Rate (35-40)	0.005	0.080**	-0.281***
	(0.025)	(0.032)	(0.022)
Tax Rate (40-45)	-0.026	0.057*	-0.268***
	(0.030)	(0.029)	(0.023)
Age	0.034***	0.019***	0.009***
	(0.005)	(0.004)	(0.002)
Age-sq	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Male	0.128***	0.104***	0.083***
	(0.030)	(0.022)	(0.022)
Black	-0.064*	0.01	-0.112**
	(0.032)	(0.030)	(0.043)
Hispanic	-0.104**	-0.131***	-0.102***
	(0.043)	(0.034)	(0.022)
Other Race	0.135**	-0.021	-0.052
	(0.052)	(0.049)	(0.039)
Education	-0.019***	-0.012***	-0.010***
	(0.005)	(0.004)	(0.002)
Income	-0.004***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)
Married	-0.072***	-0.058***	-0.031**
	(0.017)	(0.015)	(0.011)
Missing Info	0.022	-0.004	0.068
	(0.045)	(0.023)	(0.039)
Observations	1,272	2,322	2,569
R-squared	0.121	0.097	0.07

Robust standard errors in parentheses clustered at the state level
. *** p<0.01, ** p<0.05, * p<0.1