Sex, Drugs, and Cancer: The Association between Oral Tobacco Use, HPV, and Head and Neck Cancer[§]

Jennifer Maki*

November 1, 2012

Abstract

This paper explores the relationship between oral tobacco use, oral human papillomavirus (HPV) infection, and oral cancer. Oral infection of HPV, a sexually transmitted disease, is an established causal factor for oral cancer. However, previous estimates of the relationship between oral tobacco use and oral cancer have failed to take into account risky health behavior (including HPV infection status), which may have led to biased estimates. Using individual level data, I explore the relationship between tobacco use and mortality. As mortality represents only the most extreme outcome, I then conduct a second analysis using aggregate state level data to estimate the relationship between oral cancer incidence and oral tobacco use. The methodology I employ accounts for both censoring and selection, which are present in the data. My research findings suggest that when controlling for smoking status, HPV infection, and both demographic and economic variables, the estimated causal effect of oral tobacco use on oral cancer is significantly diminished.

[§] Please do not cite without permission of the author.

*PhD Candidate, North Carolina State University, Department of Economics. Email: Jamaki@ncsu.edu. I would like to thank Barry Goodwin and Melinda Morrill for their encouragement, excellent advice, and many useful suggestions. I would also like to express appreciation for the valuable comments I received from Robert Clark, Allan Deserpa, Jeffrey Federspiel, Robert Hammond, and Thayer Morrill. I also thank Brad Rodu for providing information and suggesting sources that I referenced during this project.

I Introduction

Reducing smoking prevalence in the US is a major goal of national health policy.[1] Although the harmful side effects from smoking are well known, and despite public effort to discourage uptake and encourage cessation, the decline in the smoking rate has stagnated. Recent research indicates that some of the most promising methods to aid in smoking cessation may have little long term impact (see Alpert et al., 2012). This highlights a disconcerting reality facing smokers. Faced with increased risk of death and disease due to smoking, many smokers want to quit but are unable to do so. The commonly advocated methods to assist in smoking cessation are not only costly, but offer questionable effectiveness. As such, a harm reduction approach to tobacco consumption may provide a way to mitigate this problem.

Tobacco harm reduction involves advocating the use of a less harmful alternative for those smokers who are unwilling or unable to quit smoking. Within this framework, oral tobacco may be an acceptable alternative to cigarette smoking.[2,3] If the cost (in terms of health risks) associated with oral tobacco use is less than the cost associated with smoking, welfare gains may be possible by substituting one product for the other. However, many in the health community oppose efforts to encourage smokers to substitute oral tobacco for cigarettes. The reasons are two-fold: (1) oral tobacco is thought to be harmful, so while it may be less harmful than smoking, complete abstinence is preferable, and (2) framing a tobacco product as being less harmful than cigarettes may induce usage by some who would have otherwise abstained. In addition, it is possible that oral tobacco could act as a gateway to smoking if these new users become addicted to nicotine and progress to cigarette smoking.[4]

Evaluating the health risks due to oral tobacco use is complicated and public perception regarding the competing risks is uninformed. Phillips and Heavner (2009) note the misperceptions about the relative risks of these products are primarily due to three reasons. Despite substantial evidence indicating otherwise, many people believe that (1) oral tobacco use is as risky as cigarette smoking, (2) oral tobacco use is more likely to cause oral cancer than smoking, and (3) nicotine causes cancer. Incorrect information

[1] The US Department of Health and Human Services lists reducing tobacco use as a priority. Accordingly, they have a stated objective of decreasing adult smoking prevalence to 12 percent by 2020.

[2] In this paper, I use the term "oral tobacco" to refer to both snuff and chewing tobacco.

[3] It is particularly relevant to consider the substitution possibility at present as a new generation of oral tobacco products (called snus) that come in small, satchel-like packets and do not require spitting, have been recently introduced in the US market.

[4] Evidence collected from Sweden, a country where oral tobacco use is high and smoking prevalence is low, should alleviate some of the concerns about oral tobacco usage acting as a gateway to smoking (see Maki, 2012; Norberg et al., 2011; Rodu et al., 2002).

about the true risk associated with oral tobacco use prevents users from making consumption decisions that maximize utility.

Oral tobacco use is commonly believed to cause cancer of the oral cavity and pharynx, hereafter referred to as oral cancer. In prior research, estimates of the relative risk associated with use vary widely depending on the sample restrictions applied and controlling variables employed in the analysis (see Lee and Hamling, 2009). The studies that find a link between oral tobacco use and oral cancer were published prior to 1990 (for example, see Winn et al., 1981). More recent studies (that were conducted using more modern techniques) fail to support these earlier results. These early epidemiological studies conducted during the 1950s through the 1980s include very few (if any) controls and did not account for differences in individual risk behavior.

A risk factor for oral cancer has recently emerged that is little recognized by those outside of the health community. Oral infection of Human Papillomavirus (HPV), a sexually transmitted disease, has been identified as a significant risk factor for cancers of the head and neck (Auluck et al., 2010). As the link between oral tobacco use and oral cancer was established prior to the identification of the role of HPV in cancers of this region, it is possible that earlier risk estimates are incorrect due to omitted variables, and in particular, HPV infection status. If an important causal factor is omitted from a model, interpretation of the coefficients from an equation is problematic. This omitted variable bias (or specification bias) may lead to a bias in the estimated coefficients.

Despite the large number of epidemiologic studies exploring the association between oral tobacco use and oral cancer, the findings are not conclusive. Results from past studies are mixed, and furthermore, they failed to account for the role of oral HPV infection. Accordingly, further examination of the association between oral tobacco use and oral cancer is necessary in order to determine the appropriateness of advocating oral tobacco use as an alternative to cigarette smoking.

In this paper, I conduct two separate empirical exercises. I first use individual level data to estimate the relationship between oral tobacco use and mortality. The results fail to show a link between oral tobacco use and increased probability of death. This may be due to the fact that oral cancer is treatable and mortality represents only the most extreme outcome. As such, I conduct a second analysis using state level data to explore the connection between oral tobacco use and oral cancer incidence. This analysis is complicated by censored data and selection, which I address in Section IV. In exploring cancer incidence, both smoking and HPV emerge as important risk factors. The results provide suggestive evidence that the correlation between oral tobacco use and oral cancer can be explained, in part, by risky health behavior exhibited by oral

tobacco users.

The remainder of the paper is structured as follows. Section II contains background information and motivates the study. Section III presents individual level analysis evaluating the link between tobacco use and mortality. In Section IV, I focus on morbidity, using state level data to investigate the link between oral tobacco use, risky sexual behavior, and cancer incidence. Section V concludes.

II Background

To properly frame this discussion, I begin by explaining why individuals choose to consume tobacco and list the commonly noted health effects that result from its use. This will help the reader to understand why many smokers find quitting difficult and why a harm reduction approach to smoking cessation may be a reasonable alternative to currently recommended methods. Nicotine is a drug and can act as a stimulant. Studies show that it has been found to increase concentration and memory (Sherwood, 1993). Nicotine has been studied as a potential treatment for those with Alzheimers, Parkinsons, and Attention Deficit Hyperactivity Disorder (ADHD) (Sacco et al., 2004). Cigarettes and oral tobacco are the two most common methods used to consume nicotine.

Tobacco use is linked to adverse health effects and is cited as the leading cause of preventable mortality in the United States (Iwasaki et al., 2006). The harm from tobacco use can be broadly grouped as resulting from three factors: (1) nicotine, (2) the additives and chemical reactions that occur during the production process, and (3) the delivery method. Nicotine is addictive and although it is not carcinogenic, nicotine may slow the body's ability to destroy unwanted cells. Nicotine intake increases heart rate, blood pressure, respiration, and blood glucose levels (Marieb and Hoehn, 2007).[5]

The second factor, the production process, involves transforming the raw tobacco into the final product used in oral tobacco and cigarettes. This process involves fermentation, which leads to the development of carcinogenic tobacco-specific nitrosamines (TSN).[6] In addition, other additives that influence the taste or flavor of the product may also be harmful.

[5]There may be additional risk associated with nicotine use among pregnant women. Some studies have found an increased risk of preeclampsia, preterm delivery, and low birth weight in infants of mothers who use oral tobacco (England et al. 2010).

[6]A type of oral tobacco not included in this study but commonly used in Sweden and recently introduced in the US (snus), differs from oral tobacco commonly used in the US as it contains far lower levels of TSNs. Instead of fermentation, it is produced through a process similar to pasteurization which prevents the formation of the carcinogenic nitrosamines.

The final factor, the delivery method, is where the difference in harm between the two types of products is most evident. While consumers derive utility from nicotine consumption, they may experience adverse health effects due to the nicotine delivery system employed. Cigarette smoking requires the inhalation of smoke through the airways to the lungs. Through inhalation, the chemicals in cigarette smoke are transmitted through the bloodstream and reach many other parts of the body (Phillips and Heavner, 2009). Smoking leads to both cardiovascular disease and chronic obstructive pulmonary disease, which are major causes of death and disease.[7] With oral tobacco, nicotine extraction occurs through absorption across the oral mucus membranes. This method arguably results in less risk than does smoking.

Due to the low prevalence of oral tobacco use, most of the discussion about tobacco cessation revolves around cigarette smoking. Although the smoking rate in the United States has declined since reaching a peak in the early 1980s,[8] 19 percent of all adults over the age of 18 smoke cigarettes. Smoking cessation medications[9] are recommended by the US Department of Health and Human Services Clinical Practice Guide. Nicotine replacement therapies (NRT) are commonly advocated as an appropriate tool to help smokers quit. Funding for these products is publicly available with Medicaid programs covering one or more forms of NRT in 39 U.S. states (Alpert et al., 2012). NRTs, approved by the FDA, have increased in popularity with sales growing from 45 million dollars in 1984 to over 800 million dollars in 2007 (Cary, 2012).

Many studies using randomized controlled trials (RCTs) find that use of NRTs increase the initial probability of quitting by up to 100 percent compared to those using a placebo. While these results are promising, only a relatively small fraction of smokers achieve prolonged abstinence (see Eisenberg et al. 2008 for a meta-analysis of 69 trials involving a total of 32,908 patients). The modest long term results may still overstate the general effectiveness of NRTs as findings from RCTs may not translate well into a "real world" experience. A recent long-term population based study by Harvard's Center for Global Tobacco Control highlights this issue. In the study by Alpert et al. (2012), their findings suggest that NRTs are generally not a successful component of a smok-

[7] *http : //www.cdc.gov/tobacco/data − statistics/fact − sheets/health − effects/effects − cig − smoking/* (Accessed September 12, 2012)

[8] Trends in per capita cigarette consumption *http : //www.lung.org/finding − cures/our − research/trend − reports/Tobacco − Trend − Report.pdf* (Accessed July 23, 2012)

[9] Medications identified in the US Department of Health and Human Services Clinical Practice Guide include nicotine gum, nicotine inhaler, nicotine lozenge, nicotine nasal spray, nicotine patch, and two non-nicotine medications: Bupropion SR and Varenicline. Nicotine Replacement Therapy (NRT) involves the use of one or more of these medications. *http : //www.ahrq.gov/clinic/tobacco/treating − tobacco − use*08.*pdf* (Accessed September 6, 2012)

ing cessation strategy. This prospective cohort study that includes 787 adult smokers in Massachusetts finds that the long-term quit ratio of smokers utilizing NRT with or without counseling was no higher than those who attempted to quit without the use of these aids. These findings raise doubt regarding the overall efficacy of currently advocated smoking cessation methods and should motivate discussion regarding alternative approaches.

Bans on public smoking are another tool used to decrease smoking prevalence. However, recent research finds that these bans may lead to increased smokeless tobacco use, primarily through the dual use of oral tobacco products and cigarettes (McClave-Regan and Berkowitz, 2011). Dual use among tobacco users is a concern for those in the public health community as many believe it may hinder cessation and lead to increased health risks.[10] However, this dual use behavior provides suggestive evidence regarding the substitutability among tobacco products. If smokeless tobacco is equally as harmful as cigarette smoking, there may be no public health gain associated with a change in consumption behavior. But, if smokeless tobacco is less harmful, gains are possible.

If tobacco users are not fully informed about the differing health risks associated with tobacco products, they cannot make optimal consumption decisions. Becker's and Murphy's Rational Addiction framework (Becker and Murphy, 1988) posits that consumers make consumption decisions that are influenced not only by past behavior (which can be thought of as a habit stock in terms of an addictive good) and current utility, but also take into account the future cost of use. The future cost component includes health effects due to use. Within this framework, it is necessary that the consumer has reliable information about the full cost of the good (both current and future) in order to make utility maximizing decisions.

Becker et al. (1994) and Gruber and Koszegi (2001) utilize the rational addiction framework and explore the effect of increases in excise taxes on current cigarette demand. They find that consumers adjust current smoking levels in response to an anticipated price increase, providing evidence of future looking behavior and the effectiveness of tax policy to influence cigarette demand. Farrelly (2004) and Adda and Cornaglia (2006) also look at the effect of tax increases on demand, but find evidence of compensating behavior. They find that smokers react to an increase in price by choosing cigarettes

[10]Despite concerns regarding dual usage of oral tobacco and cigarette smoking, evidence from Sweden where a form of oral tobacco (snus) is commonly used suggests that dual use is relatively uncommon. In a sample of 6,055 40 year old males in 2002-2007, 33.7 percent identified themselves as current snus users with only 5.6 percent indicating dual use. Interestingly, 36.5 percent of those identifying themselves as current snus users were former smokers (Norberg et al., 2011). A second study focusing on Northern Sweden notes that dual use is infrequent and present in only 3-5 percent of the sample (Rodu et al., 2002).

higher in tar and nicotine (Farrelly, 2004) and smoke the cigarettes they do consume more intensively (Adda and Cornaglia, 2006). Although taxation has been shown to decrease demand, this compensating behavior highlights a limitation of using taxation as a means to reduce cigarette consumption.

Less studied, but also important, are the externalities created by tobacco use. These include the excess demand for medical care due to the adverse health effects from smoking, the loss of productivity in the workplace, and secondhand smoke. Manning et al. (1989) looks at the externalities of smoking and find that although smokers incur higher medical expenses, they also die sooner. This in turn leads to a fairly balanced trade-off: non-smokers subsidize smokers health care and smokers subsidize non-smokers pensions and long-term care. Gruber (2001) notes that smokers not only take additional sick days (compared to nonsmokers), but may also be less productive while at work.[11] He also notes that nonsmokers exposure to cigarette smoke results in a potentially nontrivial externality. In comparing cigarette smoking to oral tobacco use, the latter creates fewer externalities. Using oral tobacco in lieu of cigarette smoking may mediate some of the loss of productivity due to smoking and would alleviate exposure to secondhand smoke.

In order to approximate the gains possible due to substituting oral tobacco for cigarettes, an accurate assessment of the risk related to oral tobacco use is necessary. However, findings on the relationship between oral tobacco use and oral cancer are mixed. Lee and Hamling (2009) conducted a meta-analysis reviewing 89 studies exploring the link between oral tobacco use, cigarettes, and cancer in North America and Europe. They find that estimates of the relative risk of oral cancer vary widely depending on the timing of the study and the sample restrictions applied. Although cigarette smoking is consistently identified as a causal factor for oral cancer in all time periods, early studies (pre-1990) find an association between oral tobacco use and oral cancer while most studies conducted after 1990 do not. The reason for the difference in results is not clear, although a likely candidate is the use of controlling variables employed.

Epidemiological studies that established the link between oral tobacco use and oral cancer failed to account for oral HPV infection, a known causal factor for oral cancer. Including an explicit variable to proxy for HPV infection may lead to more precise estimates. Risky sexual behavior is one specific risk factor that can influence individual health, and distinct risky health behaviors may be correlated.[12] Early research using

[11] Presumably, one factor that lowers their productivity is the time that smoke breaks require.

[12] Risky health behavior can be thought of as making suboptimal health decisions, i.e. poor diet, failure to seek preventative medical care, failure to take safety precautions, and engaging in risky sexual behavior.

case-control analysis or cohort studies often did not control for confounding factors[13] while more recent studies generally include a richer set of controls. As oral tobacco users differ from non-users in many ways, failing to control for economic and demographic variables that proxy for time preference and risky behavior (which can affect health outcomes) complicate the interpretation of study findings. The use of economic variables like education and income may be correlated with risky health behavior, and in this way, may also collectively act as a crude proxy for HPV infection.

HPV is generally associated with genital infection, but oral infection is becoming a growing concern (see Chaturvedi et al., 2011; Kreimer et al., 2005; Gillison et al., 2012). There are many different strains of the HPV virus, some of which are harmful and others that are thought to be innocuous. Infection is site specific and may occur in the genitals or oral cavity and throat. Oral infection is generally through oral to genital contact but may also occur with heavy kissing. The body has the ability to clear the virus, but there is currently no treatment available to cure HPV infection.

Gillison et al. (2012) conducted a study using data from the 2009-2010 National Health and Nutrition Examination Survey (NHANES) and estimated that the oral HPV infection rate was 6.9 percent among individuals aged 14-69 years old. For HPV-16, one of the high risk strains, the prevalence rate of 1.0 percent. They note that prevalence increased with the "number of sexual partners and cigarettes smoked per day," but did not comment on the role of smokeless tobacco use. In a review of 60 studies, Kreimer et al. (2005) calculated that the average HPV positivity rate was 35.6 percent for cancerous tumors of the oral cavity. Chatuvedi et al. (2011) notes that the number of oral cancer cases linked to HPV has been increasing and if recent trends continue, the number of oral cancer cases due to HPV is expected to surpass the number of cervical cases by 2020.

Due to the issues stated above, previous research attempting to estimate the relationship between oral tobacco and oral cancer produce findings which are not sufficient to reach a conclusion. For this reason, further examination of the association between oral tobacco use and oral cancer is warranted. To estimate the population level health risk associated with oral tobacco use, while controlling for economic and demographic variables, I employ two different approaches. The first approach utilizes mortality linked individual level data while the second, addressed in Section IV, focuses on cancer incidence using aggregate state level data.

[13] As technology has improved greatly within the last few decades, failure to control for these variables in early studies was likely due, in part, to limitations in computing power or data restrictions.

III Individual Level Analysis: Mortality

Using data from the 1994 Integrated Health Interview Series (IHIS), I explore the association between tobacco use behavior and mortality. The data was collected by the CDCs National Center for Health Services (NCHS) for the National Health Interview Survey (NHIS). The interview covers a range of health topics as well as collects demographic and economic information of respondents. The 1994 series covered tobacco use in a special topic component that included questions regarding oral tobacco usage. Adults who provided sufficient information were eligible for inclusion in the mortality linkage follow-up. Linked mortality data indicate mortality status as of 2004. Mortality data include an indicator as to whether the individual was assumed deceased[14] at the time of follow up and cause of death, if available.

The IHIS mortality linked sample includes observations on 19,349 individuals age 18 and older.[15] The sample is 42 percent male and 81 percent white. Respondents located in the South make up 32 percent of the sample. As of 1994, 25 percent of respondents indicated that they were current smokers, while only 2.6 percent were current oral tobacco users. Approximately 16 percent of the sample were assumed deceased at the date of follow-up.

IHIS data allow for examination of the association between oral tobacco use and all-cause mortality within a richer context than that typically utilized in past epidemiological studies exploring the link between oral tobacco use and related morbidity/mortality. These dated case-control studies and cohort analysis pose several concerns. Although case-control studies include a large number of cases, the controls are usually matched only by age, and possibly race and/or gender. The two groups often differ markedly by education, income (if reported), and other variables. As the control group differs from the case group in meaningful way, results may be biased.

Cohort analysis generally includes a larger sample than what is typical for a case-control study, but the cohort is not reflective of the US population. For example, the two major cohort studies investigating the link between oral tobacco and oral cancer are US Veterans Study and the Cancer Prevention Study. The Veterans study, as the name

[14]Per IHIS documentation, mortality status was determined by NCHS based on probabilistic matches of survey participants NHIS records to National Death Index (NDI) records. $http://www.ihis.us/ihis-action/variables/MORTSTAT/description-tab$(Accessed September 7, 2012)

[15]This sample differs from the underlying population as only those respondents who provide sufficient information for the mortality follow-up are included. As adult respondents who provide sufficient information to be included in the mortality follow-up may differ from those who do not, NCHS developed eligibility weights to correct for this potential bias. In results not shown, the use of the eligibility adjusted weights did not materially change the regression results.

suggests, was limited to veterans who as a whole, differ from the general US population (Zahm et al., 1992).[16] The Cancer Prevention Study II (CPS-II) cohort consists of friends, neighbors, and acquaintances of American Cancer Society volunteers. The study participants were more educated, had higher incomes, were more likely to be married, and were less racially diverse than the general US population (Henley et al., 2005). The CPS-II study compares men who used tobacco with men who never used other tobacco products (including cigarettes). Individuals who use tobacco differ from non-users in certain ways (a commonly cited example is general time preference and self-control (Kan, 2007)). Tobacco users may be more likely to engage in risky sexual behavior or other activities detrimental to one's health. The use of this tobacco-free reference group may lead to biased findings.[17]

The IHIS sample corresponds closely to a cohort study in design, but the sample is more representative of the US population than those used in either of the cohort studies referenced above. As the IHIS dataset does not contain a measure of risky sexual behavior or a proxy for HPV infection status, I do not consider the relationship between oral HPV infection and oral cancer incidence here. However, the role of HPV is addressed in the second approach which uses state level data. The analysis presented in this section demonstrates how the relationship between oral tobacco use and mortality changes as more controls are added to the model. Using the IHIS data, I estimate the increase in mortality attributable to tobacco use while controlling for economic and demographic characteristics of respondents. I estimate the following equation using logistic regression:

$$Pr(Death_i = 1) = \beta_0 + \beta_1 \mathbf{X}_i + \beta_2 White_i + \beta_3 Male_i + \beta_4 Smoke_i + \beta_5 OralTobacco_i + \varepsilon_i \quad (1)$$

where i refers to the individual and $\mathbf{X}$ is a vector of economic and demographic characteristics including age, years of schooling, married, income categories as used in the

[16] The Veterans study included a total of 248,046 veterans who completed a tobacco use history questionnaire by mail in 1954 or in 1957. Findings show that those who had ever used oral tobacco experienced a 40 percent excess risks of oral cancer (though the result was not statistically significant). As of 1980, there were 119 deaths due to oral cancer among this cohort. However, of the 2,038 veterans who indicated that they used oral tobacco only (no cigarette smoking), there were no reported deaths due to oral cancer.

[17] Although the CPS-II has several limitations (as noted above), it is the most recent study conducted exploring the link between oral tobacco use and cancer. The initial survey was conducted in 1982 and the study has a sample size of 114,809 individuals. The sample consists of males that either used smokeless tobacco or never used any type of tobacco. There are 3,327 respondents who report exclusive use of oral tobacco. The 18 year follow-up period identified 19,588 deaths (17 percent), only one of which is attributed to oral cancer.

survey, and region indicators. The dependent variable indicates the probability of dying within ten years of completing the survey.

To highlight the difference in findings that result from including non-biomedical controls (i.e., income, education, etc.) and to allow for a comparison between my findings and those from early epidemiological studies, which use few controls, I consider several specifications of the equation.[18] Table 1 presents the regression results with the coefficients presented as average marginal effects. Column (1) contains the base specification which controls only for age. We see that when including only the control for age, both oral tobacco use and cigarette smoking are associated with a greater probability of death. In column (2), I add race and gender as additional controls. While there is little change in the coefficient on *Smoke* or *Age*, the estimated coefficient on *Oral Tobacco* becomes smaller. In comparing the relative magnitude of the estimated coefficients on the tobacco use variables, we see that the coefficient on *Smoke* is nearly three times the size of that of *Oral Tobacco*. In column (3) I include the full model with controls for education, income, marital status, and region. Controlling for income is important as the correlation between health and income is well established. Education is an important variable as it can be used as a measure of time preference. Including region can control for differences in lifestyle and population health. Including these additional controls leads to a significant reduction in the size of the estimated coefficient on *Oral Tobacco* and the variable is no longer statistically significant.[19]

This final set of results indicates that cigarette smoking is associated with a statistically significant increase in the probability of death. The average marginal effect is seven percentage points, which is approximately 43 percent of the mean death rate. Males and those residing in the South also experience an increase in the probability of death. On the other hand, having additional years of schooling, having income above 40,000 dollars per year, and being married are associated with a reduction in the probability of death. Within this specification, when controlling for confounding factors, I fail to reject the null hypothesis that oral tobacco influences mortality.

[Table 1]

[18] When comparing the relationship between tobacco use and mortality, most epidemiological studies that have been conducted in the past (1950s-1990s) implement controls for age, race, and gender while more recent studies may include controls for income or education, but generally not both. Region is generally not accounted for. See Lee and Hamling (2009) for a review of 89 epidemiological cohort and case control studies investigating smokeless tobacco use and cancer. Their paper summarizes the study design, sample, and controlling variables employed in each of the studies they review.

[19] As using oral tobacco may be collinear with gender, in results not shown I estimate the full set of regressions on a sample limited to males only. Within this specification, the estimated average marginal effect of oral tobacco is smaller (0.002) and remains insignificant. The remaining estimated coefficients are similar to those presented in Table 1 using the full sample.

The IHIS dataset also includes cause of death. Of the 3,200 deceased individuals in the dataset, 11 cases list oral cancer as the cause of death. It is interesting to note that none of those with oral cancer as the cause of death indicated current or former oral tobacco use at the date of the 1994 interview. However, 9 of the individuals were smokers.

IV Aggregate State Level Analysis: Morbidity

Although the coefficient on oral tobacco use is not statistically significant in the full model using IHIS data, it may be that a relationship between oral tobacco use and death does not emerge because the disease is treatable. In this case, exploring the relationship between use and incidence of cancer would be more informative. In this second part of my analysis, I turn my attention to incidence.

Individual level data that includes behavioral observations and linked health outcomes over an extended period of time (i.e., 30 years or more) would be the preferred method to investigate the health effects related to oral tobacco use. However, population wide individual level data that include information on tobacco use, health related metrics including sexually transmitted disease, demographic and economic information, and cancer incidence does not exist. Therefore, I utilize an approach based on aggregated state level data. Reporting of oral cancer diagnosis allows patients to remain anonymous, and as such, case data cannot be linked to individuals. The cancer incidence data (the outcome of interest) are reported by geographic region and the analysis presented here uses data at the state level.[20] However, counts are suppressed if there were fewer than 16 cases for that race-gender-state cell. As the data are censored, sections IV.3 and IV.4 include a discussion of conducting regression estimation when missing values of the dependent variable coupled with selection is present.[21]

Since tobacco use behavior and demographics vary widely among race and gender, I construct four groups of individuals from each state and present average economic, demographic, and tobacco use behaviors for white females, white males, black females, and black males, which results in sample size of 204. Grouping by race and gender maintains the observed differences present in the sample; aggregating over all groups would result in a loss of meaningful variation.

[20] While county level data are available, I use state level cancer incidence data as the tobacco use data are not available at the county level.

[21] Access to confidential data that contain the full set of counts is subject to approval by National Center for Health Statistics Research Data Center Review Committee. I currently have an application under review. If approved, future work will include an analysis using the full dataset.

I use the state-level oral cancer incidence rates for each of the four race-gender groups used in the analysis. Incidence is recorded as the rate per 100,000 so as to make rates comparable across populations of various sizes. While the cancer incidence data is at the state level (as obtained from the data source), demographic, economic, and tobacco use variables are at the individual level. In order to make the independent and outcome variables comparable, I aggregate demographic, economic, and tobacco use data at the state level. The demographic, economic, and tobacco use data are transformed so that they present not individual responses, but the rate for that specific group. Tobacco use rates can be interpreted as a measure of intensity of use. For example, if 20 percent of black males in Oregon smoke cigarettes, their group level use rate is recorded as .2. This format allows the researcher to explore how variation in intensity of use of tobacco corresponds to oral cancer incidence.

IV.1 Data

A benefit of using state level data is that it allows me to draw data from many different sources to construct a rich dataset. Demographic, economic, and tobacco use data were obtained from the 2006-2007 and 2010 waves of the Current Population Survey - Tobacco Use Supplement (CPS-TUS). The 2006-2007 waves include observations on 237,199 individuals in all 50 states and the District of Columbia. The May 2010 wave includes 84,180 observations. Although there were a total of 321,379 individuals interviewed for the tobacco use supplement, some race-gender groups in some states comprised of only a very small number of observations. This makes group level tobacco use behavior, demographic, and economic information for these cells somewhat unreliable. As such, I developed sample inclusion rules. Groups with either a 0 or 100 percent marriage rate, or a smoking rate of 0 or greater than 40 percent[22] were deemed to be unsound and dropped from the sample. The final sample size after excluding these unreliable observations was 196.[23]

[22]The upper limit of 40 percent was based on the maximum smoking rate (37 percent) for each race/gender/state group in the CDC's 2008 Behavioral Risk Factor Surveillance System (BRFSS) dataset. The BRFSS sample is much larger and includes more observations for each group which leads to a higher precision in estimates. However, the BRFSS data was not suitable for the analysis presented here as oral tobacco use was covered in a supplemental module administered in only a small number of states.

[23]The 321,379 individual observations were grouped by race and gender and collapsed at the state level. The 50 states and the District of Columbia result in 51 distinct geographic locations. The four race-gender groups (White Male, White Female, Black Male, and Black Female) provide four sets of observations for each state, for a total of 204 observations. Eight observations were dropped which resulted in a final sample size of 196. To test the sensitivity of results to the removal of these 8 observations, regressions

Sampling weights were used to account for complex survey design and were used in collapsing the CPS-TUS data to the state level. The weighting variable is a composite that includes the inverse of the probability of inclusion in the sample and ratio estimates which account for any difference in the distribution of the population selected for the sample with the characteristics of the population of that state (US Department of Commerce, Census Bureau, 2008).

As state-level data on HPV infection was not available, I employ the use of a proxy. The CDC estimates that approximately 20 million Americans are infected with HPV and that 50 percent of people who are sexually active will contract the virus at some point in their lives.[24] Despite the high prevalence, no large scale population level data is collected. The National Cancer Institute notes that although DNA tests can be used to determine HPV infection status, the FDA approves these tests only under two circumstances: (1) as a follow up test for women with an abnormal Pap Test and (2) for cervical cancer screening for women over the age of 30. The FDA has not approved a method to diagnose HPV infection in males, nor is there any recommended screening method.[25] Due to the poor quality of HPV infection statistics, it is necessary to use a proxy for HPV in this analysis.

Chlamydia is the second most common STD and is an appropriate choice as a proxy for two reasons: there are similar risk factors for both HPV and chlamydia infection, and there is an established association between chlamydia infection and HPV infection (Samoff et al., 2005; Oakeshott et al., 2012). A major risk factor for both HPV and chlamydia infection is risky sexual activity reflected by a high number of sexual partners (Moscicki et al., 2001). Both types of infection are more likely to occur in teens and young adults. Research has shown that concurrent chlamydia infection is a risk factor for vaginal HPV infection (Oakeshott et al., 2012) and that chlamydia infection is associated with persistence of high risk HPV infection (Samoff et al., 2005). In the simplest terms, both chlamydia and HPV are associated with risky sexual behavior and having one infection increases the risk of contracting the other. If the relative level of risky sexual behavior is high, it is reasonable to assume that so is the relative level of

using the full sample (N=204) were computed and are available upon request. Comparing the regression results using the full sample to those using the final sample (N=196) demonstrates that the exclusion of these 8 observations does not materially alter the results.

[24]Estimates are based on findings from the 2003-2004 National Health and Nutrition Examination Survey (NHANES). Residents in 15 counties were selected to participate in the health examination survey. The small sample on which this estimation is based comprises of females aged 14-59 years old residing within the selected counties. $http://www.cdc.gov/std/HPV/STDFact-HPV.htm$ (Accessed July 27, 2012)

[25]$http://www.cancer.gov/cancertopics/factsheet/Risk/HPV$ (Accessed May 17, 2012)

HPV infection. I obtained state level data by race and gender on chlamydia infection for 2006-2008 from the CDC.

I use data on cancer incidence which I obtained from the United States Cancer Statistics (USCS), which compiles the official statistics from the cancer registries of all 50 states and the District of Columbia[26]. The data provide measures of the average annual cancer incidence (rate per 100,000) for the period 2006-2008. The data are aggregated by site (for example, incidence data are reported for cancers of the oral cavity and pharynx jointly) and are available for both lung cancer and oral cancer, as well as a host of other types of cancers. An advantage of using the aggregate level data is that they allow me to achieve broader coverage than obtained in any research study on this topic. While the number of oral cancer cases included in many epidemiological studies is usually in the range of 100-350, my study design allows me to include all cases reported in all 50 states and the District of Columbia during 2006-2008, over 33,000 cases. However, cancer incidence rates are suppressed if the actual number of cases does not reach a preset threshold. Accordingly, groups with low incidence *or* with low underlying population sizes are more likely to be recorded as missing. The sample size limited to those observations with non-missing data is 164.

Table 2 presents sample means. The sample is balanced by race and gender due to sample construction method employed. The sample-wide oral tobacco use rate is less than 2 percent and the smoking rate is approximately 18 percent. The average age of the respondent is 44 with an average of 12.9 years of schooling. The remaining columns illustrate the difference in demographics and tobacco use behavior by race and gender. We see that oral tobacco use is more widespread among white males, while cigarette smoking is more prevalent among black males. The chlamydia infection rate is higher among females than males of the same race.

[Table 2]

IV.2 Econometric Framework

As oral cancer is not instantaneous, most research incorporates a lag between use and diagnosis. In the research presented here, statistics pertaining to use behavior and cancer incidence are contemporaneous. As such, identification relies on the preservation of relative differences in use patterns between groups over time.

[26]The United States Cancer Statistics (USCS) compiles the official statistics from the cancer registries of all 50 states and the District of Columbia. They are produced in collaboration with the CDC, the National Cancer Institute (NCI), and the North American Association of Central Cancer Registries (NAACCR).

Nelson et al. (2006) investigate the trend in oral tobacco use over time. They find that oral tobacco use has steadily declined between 1987 and 2000. The average annual change among men during this period was .14 percentage points. Although the decline has been somewhat steady, the relative rate of decline has varied by race/Ethnicity, education, and age groups. While this has served to magnify the difference in use behavior by group, the relative ranking in use prevalence between groups has remained unchanged. For example, although the rate of decline among black males has exceeded that of white males (-0.24 vs. -0.09), oral tobacco use by white males is at least 3 percentage points higher during each year for which use statistics were reported.[27] There findings indicate that between-group and within-area variation in oral tobacco use behavior persist over time.[28]

Identification relies on geographic variation in tobacco use behavior, risky sexual behavior, and oral cancer incidence. Figure 1 presents a representation of the geographic variation in the intensity of oral tobacco use for each race/gender group. We see that oral tobacco use among black females is heavily confined to the South, while use among white females is relatively more dispersed with high use states located in both the Midwest and the South. Difference in use patterns also emerge when comparing black and white males. There is a relatively high level of oral tobacco use among white males in the Midwest, while use among black males in this region is comparatively low.

[Figure 1]

Figure 2 illustrates the geographic variation in the HPV Proxy (chlamydia infection) for the race/gender groups. We see distinct variation in relative prevalence of infection by race. Within each race group, there is positive but imperfect correspondence between males and females. Figure 3 presents the geographic variation in oral cancer incidence. The states without shading indicate areas for which incidence data are suppressed.

[Figure 2]

[Figure 3]

I exploit geographic variation in prevalence of oral tobacco use and the incidence of oral cancer at the state level to explore the relationship between oral tobacco and

[27] Statistics are reported for 1987, 1994, and 2000. Prevalence of smokeless tobacco for black males is estimated at 3.9 (2.8-5.0) in 1987, 2.8 (1.6-4.0) in 1994, and 1.4 (0.8-2.0) in 2000. Prevalence among whites during the same period was 6.9 (6.2-7.6), 7.4 (6.4- 8.4), 5.5 (5.0-6.0), with confidence intervals reported in parenthesis.

[28] As the tobacco use statistics gathered in 2006-2010 will not be perfectly correlated with oral tobacco use from the appropriate period (i.e., 20 to 30 years in the past, an appropriate amount of time to develop cancer attributed to tobacco use), the estimated coefficients may be biased. As such, the magnitude of my coefficients should be interpreted with some degree of caution.

cancer.[29] I utilize a reduced form model using the following equation:

$$OralCancer_{g,s} = \beta_0 + \beta_1 \mathbf{X}_{g,s} + \beta_2 OralTobacco_{g,s} + \beta_3 Smoke_{g,s} + \beta_4 White_{g,s} + \beta_5 Male_{g,s} + \beta_6 HPVProxy_{g,s} + \varepsilon_{g,s} \quad (2)$$

Where g refers to the race/gender group (white female, black female, white male, black male) and s refers to state. $\mathbf{X}$ variables include: age, a quadratic in average weekly earnings, years of schooling, a region indicator, and marital status. Weekly earnings are presented in hundreds of dollars. *Married* specifies the percent of the population who indicate that they are currently married. The tobacco use variables represent the percent of the population within that race-gender-state cell that indicate current use. The region variable is an indicator function that identifies the location of the observational unit. *White* and *Male* are dummy variables that represent the race/gender identity of the group. The *HPV Proxy* variable accounts for the prevalence of risky sexual behavior. Oral HPV infection is widespread and has been identified as a causal factor for oral cancer so the coefficient on this variable is expected to be positive and significant.

I present regression results estimating equation 2 using OLS on the sample of non-censored data (N=164) in Table 3. In column (1) of Table 3, both oral tobacco use and cigarette smoking are positive and significant predictors of increased oral cancer incidence rates. The coefficient on *Currently Use Oral Tobacco* is over 3 times as large as the coefficient on *Currently Smoke Cigarettes*. However, controlling for race and gender in column (2) significantly reduces the size of the coefficient on *Currently Use Oral Tobacco*.

Column (3) includes the *HPV Proxy* variable. HPV is an established risk factor for oral cancer that has not been controlled for in previous studies exploring the link between oral tobacco use and oral cancer. Failure to account for this important risk factor may have resulted in biased estimates if risky sexual behavior is correlated with some of the variables in the model. The inclusion of this proxy in the this column reduces the coefficient on *Currently Use Oral Tobacco* and results in a loss of statistical significance for this variable. The proxy itself is statistically significant at the 1 percent level.

Column (4) includes the full set of demographic and economic controls. *Earnings*

[29] As identification relies on geographic variation, I do not use population weights in estimating the regression equation. Although characteristics of smaller groups may be less precisely measured, weighting by population size would allow those states with the largest populations (California, New York, Florida, and Texas) to largely determine the results.

Squared and *Married* are both negatively associated with oral cancer incidence, while *Currently Smoke Cigarettes*, *Resides in the South*, and *Age* are positively associated with oral cancer incidence. The positive estimated coefficient on *Age* is not surprising as the probability of developing cancer increases with age. Region indicators may be acting as a proxy for environment or as an additional health metric. According to the CDC, the obesity rate in the south is 29.4 percent, the highest in the US.[30] As obesity is associated with poor health, this may explain the significance of the *Resides in the South* variable.

Tobacco users vary from non-users in many ways, a major difference being that users *chose* to engage in an activity that they believe is detrimental to their health. This willingness to engage in risky behavior may be captured in part by the tobacco use variables, but these variables cannot alone explain, in the entirety, the difference in risk behavior between different individuals. Including additional controls such as income and marital status also account for individual differences that may affect health behavior and, as a result, health outcomes.

The results in this final column indicate that smoking is associated with increased cancer incidence. It is surprising to find that the coefficient of oral tobacco is effectively zero and is not statistically significant. However, measurement of the coefficient suffers from a lack of precision as indicated by the large standard errors.

[Table 3]

IV.3 Estimation: Multiple Imputation

A complication arising from the cancer incidence data source is that counts are suppressed if there were fewer than 16 cases reported for that specific race/sex/region group. Table 4 presents means by censoring status as implemented by the NCI. The table highlights the difference in characteristics between the censored and non-censored groups. Those included in the censored group are, on average, lower paid, younger, more likely to be black, less likely to be married, and have a higher chlamydia infection rate. The table also includes the lung cancer incidence rate for these two groups. Although this measure is somewhat incomplete as it is subject to the suppression criteria described above, a larger number of counts were released (N = 183). By comparing the groups divided by oral cancer censoring status, we see that those in the censored group experience, on average, lower lung cancer incidence rates.

[Table 4]

Although regression analysis can be performed on this sample of non-missing data,

[30] Obesity rates by region, CDC 2010 *http* : *//www.cdc.gov/obesity/data/adult.html/* (Accessed July 23, 2012)

the results may be biased. As demonstrated in Table 4, there are non-trivial differences in the demographics and tobacco use behavior between the observations with and without censored data that may lead to biased parameter estimates. To deal with the censored data, I use two separate approaches: Multiple Imputation and Heckman Selection.

Multiple Imputation (MI) is a simulation based approach to analyzing incomplete data. It uses observed data to predict missing values. This approach involves the development of an imputation model that is used to create multiple imputed datasets. The imputation model includes not only those variables used in the outcome equation, but also additional variables such as population size and the relative proportion of the state population represented by that particular group. The imputed datasets are used individually for estimation, after which the estimation parameters are pooled to create a single set of estimates. The standard error of these estimates contains two components: within imputation variance (the average of the standard error squared across imputations) and the between imputation variance (the variance of the parameter estimates across imputations). Incorporating the additional variance when using imputed values properly accounts for the uncertainty introduced by imputation (Little and Rubin, 2002). MI regression estimates were performed using OLS.[31]

I present regression results using MI in Table 5. The results are similar to those found when using OLS regression on the non-censored data (Table 3). However, there are some small differences in the magnitude of the estimated coefficients. By comparing the results from each of the four specifications presented in Table 5 (MI), we see that, consistent with the findings presented in Table 3, including variables that account for risky heath behavior and differences in economic and demographic characteristics of individuals systematically affects the magnitude of the estimated coefficient on *Currently Use Oral Tobacco*.

[Table 5]

IV.4 Estimation: Heckman Selection

The validity of the results presented using MI relies on the assumption that the censored data are missing at random (MAR) (See Kenward and Carpenter, 2007, for further discussion.). If the data are missing not a random (MNAR), valid results can only be obtained by accounting for the mechanism that causes the missing data. Oral cancer incidence rates are censored (suppressed) if there are fewer than 16 cases present in that

[31]The results presented in this subsection were estimated with the use of the MI function in Stata 12. To ensure that imputed values were non-negative, an option specifying a lower limit of 0 was used. For each specification, twenty imputations were performed.

race-gender-region group. Accordingly, groups with small underlying populations or that experience low incidence are less likely to be observed. In this case, the missingness mechanism is a function of the case count, therefor, the missing data are MNAR. If an unobserved factor is correlated with both the probability of inclusion in the sample and the cancer incidence rate, the estimates produced using MI may be affected by sample selection bias. To account for the selection that may be present, a Type II Tobit (Heckman) estimation model may provide more reliable results.

The Heckman estimator is widely used to correct for selection. Within this context, the latent variable, Y_1^*, will denote the annual oral cancer incidence rate. This variable is observed only if a second latent variable, Y_0^* is greater than 0. In this case, Y_0^* indicates whether the data collector allows the rate to be revealed in the data set. Although the typical case of sample selection occurs when there is self-selection by the individuals being studied, Heckman (1979) notes that "sample selection decisions by analysts or data processors operate in much the same fashion as self-selection."
The two equation model is comprised of a selection equation where

$$Y_0 = \begin{cases} 1 & \text{if } Y_0^* > 0 \\ 0 & \text{if } Y_0^* \leq 0 \end{cases} \tag{3}$$

and an outcome equation where

$$Y_1 = \begin{cases} Y_1 & \text{if } Y_0^* > 0 \\ unobserved & \text{if } Y_0^* \leq 0 \end{cases} \tag{4}$$

The cancer incidence rate is observed only if $Y_0^* > 0$, and has a binary form, 1 if observed, 0 if not. The model is assumed to be linear with additive error terms and can be represented as

$$Y_0^* = \mathbf{Z}\delta + \mu \tag{5}$$

$$Y_1^* = \mathbf{X}\beta + \varepsilon \tag{6}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are vectors of regressors. $\mathbf{X}$ is a subset of $\mathbf{Z}$, where $\mathbf{Z}$ includes the additional variable used for identification. The jointly normal distributed error terms, μ and ε have a mean expected value of zero and have a correlation coefficient ρ. The variance of the selection equation error term is normalized so that $\sigma_\mu^2=1$. Conditioning on the subset of the population for which $Y_0=1$, the expected incidence rate can be

expressed using the following formula (Greene, 2003):

$$\mathbb{E}\left[Y_1^* \mid Y_0 = 1, X, Z\right] = \mathbb{E}\left[X'\beta + \varepsilon \mid Y_0^* > 0\right] \tag{7}$$

$$= X'\hat{\beta} + \mathbb{E}\left[\varepsilon \mid \mu > -Z'\delta\right] \tag{8}$$

$$= X'\hat{\beta} + \hat{\rho}\hat{\sigma}_\varepsilon \frac{\phi}{\Phi} \tag{9}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution and $\phi(\cdot)$ is the corresponding probability density function. Simplifying, it can be written as

$$\mathbb{E}\left[Y_1^* \mid Y_0 = 1, X, Z\right] = X'\hat{\beta} + \hat{\rho}\hat{\sigma}_\varepsilon \hat{\lambda}\left(-Z'\delta\right) \tag{10}$$

where $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$, the inverse of the mills ratio (the nonselection hazard).

In the case presented here, I use the fraction of the entire population represented by group g in state s as the exclusion restriction. This should have a direct effect on selection, as population size influences the probability of selection, as noted above. However, there is no apparent reason why the relative size of the group in state s should itself influence oral cancer incidence. As such, this variable meets the qualifications required for it to act as an appropriate exclusion restriction.

I estimate equation 2 using the Heckman Selection Two-Step method outlined above. The estimated coefficients in Table 6 are similar to those in Tables 3 and 5. In the final two specifications presented in Table 6, we see that the estimated coefficient on *Currently Smoke Cigarettes* is larger than the coefficient on *Currently Use Oral Tobacco*. The relatively large magnitude of this variable indicates that smoking appears to have a strong causal role in oral cancer incidence while the role of oral tobacco use is less certain. We also see that the *HPV Proxy* variable is statistically significant at the 1 percent level. In comparing the estimated coefficient on *Currently Use Oral Tobacco* by column, we see that the variable continually decreases as more controls are added. This provides some suggestive evidence that the risk associated with oral tobacco use as found in previous studies may be due, in part, to differences in the individual's risk behavior profile. The resulting omitted variable bias may have been responsible for an over-weighting of the health risk associated with oral tobacco use.

[Table 6]

IV.5 Falsification Test: Lung Cancer

As a placebo robustness check, I estimate the relationship between tobacco use and lung cancer using the same methodology as described in Section IV.2.[32] Cigarette smoking is an established causal factor for lung cancer. Oral tobacco use and HPV infection are associated with risky health behavior, but are not thought to directly influence lung cancer incidence. Theoretically, the inclusion of the HPV proxy should produce little effect when modeling lung cancer incidence. To examine the relationship between tobacco use and lung cancer, I estimate the following equation using OLS:

$$LungCancer_{g,s} = \beta_0 + \beta_1 \mathbf{X}_{g,s} + \beta_2 OralTobacco_{g,s} + \beta_3 Smoke_{g,s} + \beta_4 White_{g,s} + \beta_5 Male_{g,s} + \beta_6 HPVProxy_{g,s} + \varepsilon_{g,s} \quad (11)$$

Where g refers to the race/gender group (white female, black female, white male, black male) and s refers to state. $\mathbf{X}$ variables include: age, a quadratic in average weekly earnings, years of schooling, a region indicator, and marital status. Weekly earnings are presented in hundreds of dollars. *Married* specifies the percent of the population who indicate that they are currently married. The tobacco use variables represent the percent of the population within that race-gender-state cell that indicate current use. The region variable is an indicator function that identifies the location of the observational unit. *White* and *Male* are dummy variables that represent the race/gender identity of the group. The *HPV Proxy* variable accounts for the prevalence of risky sexual behavior.

Results presented in Table 7 indicate a clear association between cigarette smoking and lung cancer incidence. Comparing the results in columns (2) and (3) of Table 7, we see that the inclusion of the HPV proxy does not result in a significant change in the size of the estimated coefficient on *Currently Smoke Cigarettes*, nor is the *HPV Proxy* variable itself statistically significant. This provides strong support for the claim that the significance of the HPV proxy in the oral cancer model is not spurious. Including additional controls in column (4) does decrease the estimated coefficient on *Currently Smoke Cigarettes*, but it remains large and statistically significant. Although risky health behavior may increase the probability of disease, cigarette smoking appears to be the main determinant of lung cancer.

The results presented in Table 7 differ markedly from those estimating the link between oral tobacco use and oral cancer. As we see in Tables 3, 5, and 6, including

[32] Lung cancer incidence rates are also subject to suppression if they do not meet the criteria described earlier. However, fewer than 7 percent of the lung cancer incidence rates are censored and in results not presented here, selection does not appear to be an issue. Therefore, I estimate the equation using OLS.

controls to account for risky health behavior (including risky sexual behavior) reduces the size of the coefficient on oral tobacco use. This indicates that while risky health behavior is an important confounding factor when estimating oral cancer incidence, it appears to play a minor role when estimating lung cancer incidence.

[Table 7]

V Conclusion

Tobacco harm reduction as a means to increase smoking cessation has been largely dismissed by respected health agencies due to concerns regarding the health risks of oral tobacco use.[33] The results presented here provide suggestive evidence that while smoking clearly leads to adverse health, oral tobacco use may be significantly less harmful.

If individuals are unwilling or unable to give up smoking, advocating a less harmful alternative may lead to a reduction of smoking prevalence. There is the potential for massive savings in health care cost that would result from a decline in smoking. Lung cancer is just one of the many diseases caused by smoking. The CDC estimates that smoking is responsible for 90 percent of all lung cancer cases.[36]

In 2007, there were 153,017 hospital discharges for lung cancer treatment with a mean charge per discharge of 45,473 dollars (Newton and Ewer, 2010). If all smokers use oral tobacco (thus eliminating the smoking related lung cancer costs), this could represent potential cost saving in the excess of 6 billion dollars in one year alone. This back of the envelope cost benefit analysis provides only a rough measure as it does not include the loss of productivity or the pain and suffering inflected on the patient by the disease. In addition, the substantial funding directed toward tobacco cessation [37] could be used to fund education, improvements in public infrastructure, or a host of other important causes.

In this paper, I explore the link between oral tobacco use and oral cancer. I first use individual level data to estimate the relationship between oral tobacco use and mortal-

[33]The CDC discourages the use of smokeless tobacco in a harm reduction approach and states that "smokeless tobacco is not a safe alternative to smoking cigarettes."[34] The National Cancer Institute states "All tobacco products are harmful and cause cancer, and the use of these products is strongly discouraged. There is no safe level of tobacco use. People who use any type of tobacco product should be urged to quit."[35]

[36]$http://www.cdc.gov/cancer/lung/basicinfo/riskfactors.htm$. Accessed July 26, 2012

[37]Rhoads (2012) notes that the CDC recommends an average funding level of $12.34 per capita annually for tobacco control. However, only one state, North Dakota, provides funding at the suggested level. 9 states fund at levels between 50 percent and 99 percent of the recommend amount while 31 states and the District of Columbia fund at levels which equate to less than 25 percent of the recommended amount.

ity. My results fail to show a link between oral tobacco use and increased probability of death. This may be due to the fact that oral cancer is treatable, so focusing on mortality may not be appropriate. As such, I conduct a second analysis to test the relationship between the oral tobacco use and incidence of cancer by exploiting geographic variation in the intensity of oral tobacco use and the incidence of oral cancer at the state level. My findings show that controlling for differences in risky health behavior (including risky sexual activity) reduces the estimated causal effect of oral tobacco use on oral cancer incidence. Within this framework, both cigarette smoking and HPV emerge as important risk factors for oral cancer incidence.

Previous findings that link oral tobacco use to oral cancer did not control for differences in an individual's risk behavior profile or account for oral HPV infection (for example, see Winn et al., 1981). Tobacco users differ from nonusers in many ways, and these differences may influence individual health. Tobacco users willingly engage in an activity they believe to be detrimental to their health, so findings that tobacco users may suffer increased morbidity or mortality should not be surprising. Failing to directly consider difference in an individuals risk behavior may have resulted in an over-weighting of the risk associated with oral tobacco use. Oral HPV infection, one specific risk factor, is widespread and has been identified as a causal factor for oral cancer. To my knowledge, no study that has found that oral tobacco use leads to oral cancer has taken HPV infection into account.

An accurate assessment of the health risks due to oral tobacco use is important as oral tobacco may be used as a substitute for cigarettes. If consumers are provided with accurate information about the harm associated with oral tobacco use, gains may be possible. Smokers may elect to substitute oral tobacco for cigarettes, which would allow them to continue consumption of tobacco but do so through a method that has a lower cost in terms of health impact. The focus of current policy revolves around cessation and leaves little room for alternative approaches. An alternative approach would center on managing nicotine dependence. This can include optimal selection of products that present the lowest costs in terms of health.[38]

It should be noted that the work presented here has several limitations. The use of a proxy for HPV infection was necessary as population level estimates are not collected. As the reliability of the estimates depends on the strength of the proxy, better data on HPV infection would allow for a more robust analysis. In addition, analysis of the oral

[38]Considering the substitution possibility is particularly relevant at this point in time as a new oral tobacco product, snus, has been recently introduced in the US market. Snus is a form of oral tobacco which does not require spitting. It is contained in a small, teabag-like satchel and due to the small size, can be discretely used.

cancer incidence data is complicated by the presence of suppressed counts that lead me to the use of imputation techniques. As these methods are sensitive to the assumptions used, more robust treatment could be performed with a complete dataset. Oral tobacco use and oral cancer are both relatively rare events that could complicate identification. In addition, I do not address the concern that oral tobacco use may lead to an increase in cigarette uptake.

My results suggest the need for more extensive data and more research on the link between oral tobacco use and oral cancer. The findings presented in this paper should be cause for public health policy officials to increase efforts to collect data on population level HPV infection. Oral HPV infection is a significant causal factor of oral cancer, but quality data on population level infection rates do not exist. A better understanding of the prevalence and trends over time may help in highlighting the importance and dual benefit of the HPV vaccine as well as allow for further study of the relationship between oral tobacco use, HPV, and oral cancer.

References

Adda, J. and F. Cornaglia (2006). Taxes, cigarette consumption, and smoking intensity. *American Economic Review 96*(4), 1013–1013.

Alpert, H. R., G. N. Connolly, and L. Biener (2012). A prospective cohort study challenging the effectiveness of population-based medical intervention for smoking cessation. *Tobacco control*.

Auluck, A. A., G. G. Hislop, C. C. Bajdik, C. C. Poh, L. L. Zhang, and M. M. Rosin (2010). Trends in oropharyngeal and oral cavity cancer incidence of human papillomavirus (hpv)-related and hpv-unrelated sites in a multicultural population: the british columbia experience. *Cancer 116*(11), 2635–2644.

Becker, G. S., M. Grossman, and K. M. Murphy (1994). An empirical analysis of cigarette addiction. *American Economic Review 84*(3), 396–418.

Becker, G. S. and K. M. Murphy (1988). A theory of rational addiction. *Journal of Political Economy 96*(4), 675–700.

Cary, B. (2012). Nicotine gum and skin patch face new doubt. *New York Times*.

Chaturvedi, A. K., E. A. Engels, R. M. Pfeiffer, B. Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M. T. Goodman, M. Sibug-Saber, W. Cozen, L. Liu, C. F. Lynch, N. Wentzensen, R. C. Jordan, S. Altekruse, W. F. Anderson, P. S. Rosenberg, and M. L. Gillison (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the united states. *Journal of Clinical Oncology 29*(32), 4294–4301.

Eisenberg, M. J., K. B. Filion, D. Yavin, P. Belisle, S. Mottillo, L. Joseph, A. Gervais, J. O'Loughlin, G. Paradis, S. Rinfret, and L. Pilote (2008). Pharmacotherapies for smoking cessation: a meta-analysis of randomized controlled trials. *Canadian Medical Association journal 179*(2), 135–144.

Farrelly, M. C. (2004). The effects of higher cigarette prices on tar and nicotine consumption in a cohort of adult smokers. *Health Economics 13*(1), 49–58.

Gillison, M. L., T. Broutian, R. K. L. Pickard, Z.-y. Tong, W. Xiao, L. Kahle, B. I. Graubard, and A. K. Chaturvedi (2012). Prevalence of oral hpv infection in the united states, 2009-2010. *JAMA : The Journal of the American Medical Association 307*(7), 693–703.

Greene, W. (2003). *Econometric Analysis, 5th Edition*. Prentice Hall.

Gruber, J. (2001). Tobacco at the crossroads: The past and future of smoking regulation in the united states. *Journal of Economic Perspectives 15*(2), 193–212.

Gruber, J. and B. Koszegi (2001). Is addiction "rational"? theory and evidence. *Quarterly Journal of Economics 116*(4), 1261–1303.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Henley, S., M. Thun, C. Connell, and E. Calle (2005). Two large prospective studies of mortality among men who use snuff or chewing tobacco (united states). *Cancer Causes and Control 16*(4), 347–358.

Iwasaki, N., C. H. Tremblay, and V. J. Tremblay (2006). Advertising restrictions and cigarette smoking: Evidence from myopic and rational addiction models. *Contemporary Economic Policy 24*(3), 370–381.

Kan, K. (2007). Cigarette smoking and self-control. *Journal of Health Economics 26*(1), 61–81.

Kenward, M. G. and J. Carpenter (2007). Multiple imputation: current perspectives. *Statistical methods in medical research 16*(3), 199–218.

Kreimer, A., G. Clifford, P. Boyle, and S. Franceschi (2005). Human papillomavirus types in head and neck squamous cell carcinomas worldwide: A systematic review. *Cancer Epidemiology Biomarkers and Prevention 14*(2), 467–475.

Lee, P. N. P. and J. J. Hamling (2009). Systematic review of the relation between smokeless tobacco and cancer in europe and north america. *BMC medicine 7*, 36–36.

Little, R. J. and D. B. Rubin (2002). *Statistical Analysis with Missing Data, Second Edition.* Whiley-InterScience.

Maki, J. (2012). The swedish experience and the incentives created by harm reduction. *Working Paper*.

Manning, W., E. Keeler, J. Newhouse, E. Sloss, and J. Wasserman (1989). The taxes of sin - do smokers and drinkers pay their way. *Jama-Journal of the American Medical Association 261*(11), 1604–1609.

Marieb, E. N. and K. Hoehn (2007). *Human Anatomy and Physiology, 7th Ed.*

McClave-Regan, A. and J. Berkowitz (2011). Smokers who are also using smokeless tobacco products in the us: a national assessment of characteristics, behaviours and beliefs of 'dual users'. *Tobacco control 20*(3), 239–242.

Minnesota Population Center and State Health Access Data Assistance Center (2011). *Integrated Health Interview Series: Version 4.0.* Minnesota Population Center and State Health Access Data Assistance Center.

Moscicki, A., N. Hills, S. Shiboski, K. Powell, N. Jay, E. Hanson, S. Miller, L. Clayton, S. Farhat, J. Broering, T. Darragh, and J. Palefsky (2001). Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. *JAMA : The Journal of the American Medical Association 285*(23), 2995–3002.

Nelson, D., P. Mowery, S. Tomar, S. Marcus, G. Giovino, and L. Zhao (2006). Trends in smokeless tobacco use among adults and adolescents in the united states. *American Journal of Public Health 96*(5), 897–905.

Newton, A. N. and S. R. Ewer (2010). Inpatient cancer treatment: An analysis of financial and nonfinancial performance measures by hospital-ownership type. *Journal of health care finance 37*(2), 56–80.

Norberg, M., G. Lundqvist, M. Nilsson, H. Gilljam, and L. Weinehall (2011). Changing patterns of tobacco use in a middle-aged population - the role of snus, gender, age, and education. *Global Health Action 4*, 5613.

Oakeshott, P., A. Aghaizu, F. Reid, R. Howell-Jones, P. E. Hay, S. T. Sadiq, C. J. Lacey, S. Beddows, and K. Soldan (2012). Frequency and risk factors for prevalent, incident, and persistent genital carcinogenic human papillomavirus infection in sexually active women: community based cohort study. *British medical journal 344*, e4168.

Phillips, C. V. and K. K. Heavner (2009). Smokeless tobacco: the epidemiology and politics of harm. *Biomarkers 14*, 79–84.

Rhoads, J. K. (2012). The effect of comprehensive state tobacco control programs on adult cigarette smoking. *Journal of health economics 31*(2), 393–405.

Rodu, B., B. Stegmayr, S. Nasic, and K. Asplund (2002). Impact of smokeless tobacco use on smoking in northern sweden. *Journal of internal medicine 252*(5), 398–404.

Sacco, K., K. Bannon, and T. George (2004). Nicotinic receptor mechanisms and cognition in normal states and neuropsychiatric disorders. *Journal of Psychopharmacology 18*(4), 457–474.

Samoff, E., E. Koumans, L. Markowitz, M. Sternberg, M. Sawyer, D. Swan, J. Papp, C. Black, and E. Unger (2005). Association of chlamydia trachomatis with persistence of high-risk types of human papillomavirus in a cohort of female adolescents. *American Journal of Epidemiology 162*(7), 668–675.

Sherwood, N. (1993). Effects of nicotine on human psychomotor performance. *Human Psychopharmacology: Clinical and Experimental 8*(3), 155–184.

United States Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute (2011). *United States Cancer Statistics: 1999-2008, WONDER On-line Database*. United States Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute.

United States Department of Health and Human Services, Centers for Disease Control and Prevention, and National Center for HIV, STD and TB Prevention (NCHSTP), Division of STD/HIV Prevention (2011). *Sexually Transmitted Disease Morbidity 1996*

- *2009, by gender, age group and race/ethnicity, CDC WONDER On-line Database.* United States Department of Health and Human Services, Centers for Disease Control and Prevention, and National Center for HIV, STD and TB Prevention (NCHSTP), Division of STD/HIV Prevention.

US Department of Commerce, Census Bureau (2008). *National Cancer Institute and Centers for Disease Control and Prevention Co-sponsored Tobacco Use Supplement to the Current Population Survey (2006-07).* US Department of Commerce, Census Bureau.

Winn, D., W. Blot, C. Shy, L. Pickle, A. Toledo, and J. Fraumeni (1981). Snuff dipping and oral-cancer among women in the southern united-states. *New England Journal of Medicine 304*(13), 745–749.

Zahm, S. H., E. F. Heineman, and J. B. Vaught (1992). Soft tissue sarcoma and tobacco use: Data from a prospective cohort study of united states veterans. *Cancer Causes and Control 3*(4).

Table 1: Mortality Analysis Means and Regression

	Means	(1)	(2)	(3)
Currently Use Oral Tobacco	2.6%	0.046**	0.027*	0.013
		(0.013)	(0.013)	(0.013)
Currently Smoke Cigarettes	25.4%	0.082**	0.079**	0.069**
		(0.005)	(0.005)	(0.005)
Age	46.2	0.010**	0.010**	0.009**
		(0.000)	(0.000)	(0.000)
White	81.2%		−0.030**	−0.012*
			(0.006)	(0.006)
Male	42.1%		0.043**	0.055**
			(0.004)	(0.004)
Income $20,000 - $40,000	7.2%			−0.003
				(0.013)
Income $40,000 plus	88.1%			−0.039**
				(0.013)
Years of Schooling	13.7			−0.003**
				(0.001)
Reside in the South	31.7%			0.011+
				(0.006)
Reside in the Northeast	20.9%			−0.008
				(0.006)
Reside in the Midwest	26.0%			−0.004
				(0.006)
Married	53.1%			−0.037**
				(0.005)
Observations	19,349	19,349	19,349	19,349

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: The sample is comprised of 19,349 individuals age 18 and up who were interviewed for the 1994 National Health Interview Survey. Coefficients are presented as average marginal effects where the dependent variable is *Assumed Deceased*, which refers to individual all-cause mortality as of 2004. 16.4 percent of the sample were assumed deceased at the date of the mortality follow up. All other variables are as of the date of interview in 1994. Tobacco use refers to whether the individual indicated he was a current user at the time of the interview. Age indicates the respondent's age in 1994, while schooling, income categories, married, race-gender, and region variables are dummy variables representing the individual's economic and demographic information. Logistic regression was used to estimate the probability of death. The omitted categories are: Black, Female, Income less than $20,000, and Reside in the West.

Table 2: Aggregate State Level Data Sample Means

	Full Sample	White Males	White Females	Black Males	Black Females
Currently Use Oral Tobacco	1.60%	4.75%	0.13%	1.00%	0.32%
Currently Smoke Cigarettes	17.50%	19.18%	15.92%	19.80%	14.88%
Age	44.07	45.43	46.92	41.25	42.39
Average Weekly Earnings	97.98	139.17	85.65	97.12	66.23
Years of Schooling	12.87	13.12	13.22	12.59	12.51
Reside in the South	34.69%	33.30%	33.30%	34.69%	37.78
Married	47.36%	59.00%	55.19%	40.83%	32.44
Chlamydia Infection (Cases per 100,000)	854.98	78.52	236.94	1192.44	2067.95
Oral Cancer Incidence (Rate per 100,000)	15.46	23.64	9.75	18.43	7.04
Lung Cancer Incidence (Rate per 100,000)	98.41	121.32	93.45	106.05	67.69
Observations	196	51	51	49	45

Notes: The table represents the state level means for each race-gender group. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. These demographic, economic, and tobacco use variables were recorded for those individuals age 18 and up at the time of the interview. Chlamydia infection, the HPV proxy variable, is the average number of cases per 100,000 for the race-gender-state cell for individuals age 15 and up. Cancer incidence is the average rate per 100,000 for each group of individuals age 15 and up. Oral and lung cancer incidence rates are limited to those race-gender-state cell's with non-missing data (N=164 and N=183, respectively).

Table 3: OLS Regression, Sample Limited to Those Observations with Non-Missing Data

	(1)	(2)	(3)	(4)
Currently Use Oral Tobacco	142.179**	19.995+	9.053	3.995
	(17.917)	(11.221)	(11.293)	(11.452)
Currently Smoke Cigarettes	44.110**	19.480**	14.551*	10.730+
	(10.698)	(5.685)	(5.675)	(6.040)
Age	0.005	0.994**	1.222**	0.913**
	(0.171)	(0.160)	(0.168)	(0.182)
White		−0.480	1.623	6.774**
		(0.826)	(1.001)	(1.681)
Male		13.219**	15.088**	16.610**
		(0.643)	(0.822)	(0.925)
HPV Proxy			0.002**	0.002**
			(0.001)	(0.001)
Average Weekly Earnings (in Hundreds)				2.765
				(2.299)
Average Weekly Earnings Sq				−1.947*
				(0.914)
Years of Schooling				0.667
				(0.722)
Reside in the South				2.190**
				(0.518)
Married				−17.671**
				(5.941)
Constant	4.873	−39.293**	−52.090**	−60.615*
	(8.208)	(7.037)	(7.733)	(27.770)
Observations	164	164	164	164
Adjusted r2	0.439	0.849	0.859	0.873

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: See Table 2 for a description of the sample. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. Chlamydia infection, the HPV proxy variable, is the average number of cases per 100,000 for the race-gender-state cell. The dependent variable is Oral Cancer Incidence. The mean incidence rate is 15.46 cases per 100,000. OLS regression was used to estimate the coefficients. The omitted categories are: Black, Female, Reside outside the South, and Not Married.

Table 4: Means by Censoring Status

	Censored	Non-Censored	Difference
Currently Use Oral Tobacco	0.52%	1.81%	−1.29**
Currently Smoke Cigarettes	16.65%	17.67%	−1.02
Age	41.39	44.59	−3.20**
Average Weekly Earnings	91.07	99.33	−8.26
Years of Schooling	12.80	12.89	−0.09
Reside in the South	12.50%	32.02%	−26.52**
Reside in the Northeast	21.88%	17.07%	4.81
Reside in the Midwest	21.88%	24.39%	−2.51
White	3.12%	61.56%	−58.44**
Male	43.75%	52.44%	−8.69
Married	38.13%	49.17%	−11.04**
Chlamydia Infection (Cases per 100,000)	1732.70	683.71	1048.99**
Lung Cancer Incidence (Rate per 100,000)	72.68	101.39	−28.71*
Observations	32	164	

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: See Table 2 for a description of the sample. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. Lung cancer incidence rates are subject to suppression if there are fewer than 16 cases for each race-gender-state cell. As lung cancer is more common and less likely to be subject to the censoring criteria, means for those non-censored cells (N=183) are presented above. The chlamydia infection (HPV proxy) and the lung cancer incidence variables present the average number of cases per 100,000 for the race-gender-state cell.

Table 5: Multiple Imputation Estimation

	(1)	(2)	(3)	(4)
Currently Use Oral Tobacco	150.995**	22.524+	14.964	8.241
	(18.033)	(12.179)	(12.350)	(11.405)
Currently Smoke Cigarettes	37.819**	16.874**	13.437*	10.190
	(9.060)	(6.201)	(6.473)	(6.181)
Age	0.299*	0.831**	0.931**	0.716**
	(0.143)	(0.130)	(0.140)	(0.136)
White		0.604	2.365*	6.252**
		(0.828)	(1.103)	(1.411)
Male		12.650**	13.830**	15.495**
		(0.697)	(0.894)	(0.879)
HPV Proxy			0.001*	0.001*
			(0.001)	(0.001)
Average Weekly Earnings (in Hundreds)				1.781
				(1.727)
Average Weekly Earnings Sq				−1.481**
				(0.516)
Years of Schooling				0.242
				(0.552)
Reside in the South				1.881**
				(0.474)
Married				−13.604**
			(3.910)	(3.947)
Observations	196	196	196	196

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: See Table 2 for a description of the sample. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. Chlamydia infection, the HPV proxy variable, is the average number of cases per 100,000 for the race-gender-state cell. The dependent variable is Oral Cancer Incidence, presented as the rate per 100,000. The mean incidence rate is 15.46 cases per 100,000. Multiple Imputation was used to predict missing values and OLS regression was used to estimate the coefficients. The standard error of these estimates contains two components: within imputation variance (the average of the standard error squared across imputations) and the between imputation variance (the variance of the parameter estimates across imputations). Incorporating the additional variance when using imputed values properly accounts for the uncertainty introduced by imputation. The omitted categories are: Black, Female, Reside outside the South, and Not Married.

Table 6: Heckman Two-Step, Oral Cancer Incidence

	(1)	(2)	(3)	(4)
Currently Use Oral Tobacco	134.848**	19.649^{+}	6.717	4.043
	(18.554)	(11.042)	(11.203)	(11.030)
Currently Smoke Cigarettes	42.355**	19.408**	13.742*	10.780^{+}
	(10.544)	(5.565)	(5.541)	(5.819)
Age	−0.274	0.944**	1.152**	0.907**
	(0.231)	(0.165)	(0.166)	(0.177)
White		−0.530	1.851^{+}	6.700**
		(0.811)	(0.976)	(1.653)
Male		13.129**	15.175**	16.596**
		(0.637)	(0.803)	(0.893)
HPV Proxy			0.002**	0.002**
			(0.001)	(0.001)
Average Weekly Earnings (in Hundreds)				2.626
				(2.303)
Average Weekly Earnings Sq				−1.892*
				(0.915)
Years of Schooling				0.703
				(0.716)
Reside in the South				2.168**
				(0.508)
Married				−17.343**
				(5.913)
λ	−4.177^{+}	−1.039	−2.101*	−0.261
	(2.260)	(1.093)	(1.016)	(1.198)
Observations	196	196	196	196

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: See Table 2 for a description of the sample. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. Chlamydia infection, the HPV proxy variable, is the average number of cases per 100,000 for the race-gender-state cell. The dependent variable is Oral Cancer Incidence, presented as the rate per 100,000. The mean incidence rate is 15.46 cases per 100,000. The Heckman Selection Two-Step method was used to estimate the coefficients. I use the fraction of the entire population represented by group g in state s as the exclusion restriction. The omitted categories are: Black, Female, Reside outside the South, and Not Married.

Table 7: OLS Regression, Lung Cancer Incidence

	(1)	(2)	(3)	(4)
Currently Use Oral Tobacco	337.902**	51.478	58.011	−79.147
	(69.930)	(74.661)	(77.155)	(72.095)
Currently Smoke Cigarettes	281.830**	238.415**	241.728**	181.160**
	(35.479)	(31.476)	(32.953)	(31.112)
Age	3.164**	5.002**	4.895**	3.962**
	(0.602)	(0.810)	(0.868)	(0.862)
White		−1.688	−3.107	22.720*
		(4.621)	(6.165)	(8.969)
Male		29.289**	28.197**	37.601**
		(3.844)	(4.965)	(5.085)
HPV Proxy			−0.001	0.002
			(0.003)	(0.003)
Average Weekly Earnings (in Hundreds)				18.460
				(11.593)
Average Weekly Earnings Sq				−10.622*
				(5.078)
Years of Schooling				−9.917*
				(3.919)
Reside in the South				15.107**
				(3.083)
Married				−26.310
				(27.859)
Constant	−96.665**	−179.622**	−173.274**	255.281+
	(28.181)	(35.367)	(39.852)	(151.849)
Observations	183	183	183	183
Adjusted r2	0.465	0.592	0.591	0.679

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: See Table 2 for a description of the sample. The tobacco use variables indicate the percent of current users within the race-gender-state cell. Married indicates the percent married, while age, earnings, and schooling are the average values for the race-gender-state cell as reported on of the date of the survey. The race-gender and region variables are indicator functions that identify the race, gender, and location of the observational unit. Chlamydia infection, the HPV proxy variable, is the average number of cases per 100,000 for the race-gender-state cell. The dependent variable is Lung Cancer Incidence. The mean incidence rate is 98.41 cases per 100,000. OLS regression was used to estimate the coefficients. The omitted categories are: Black, Female, Reside outside the South, and Not Married.

Figure 1: Geographic Variation in Oral Tobacco Use by Race and Gender

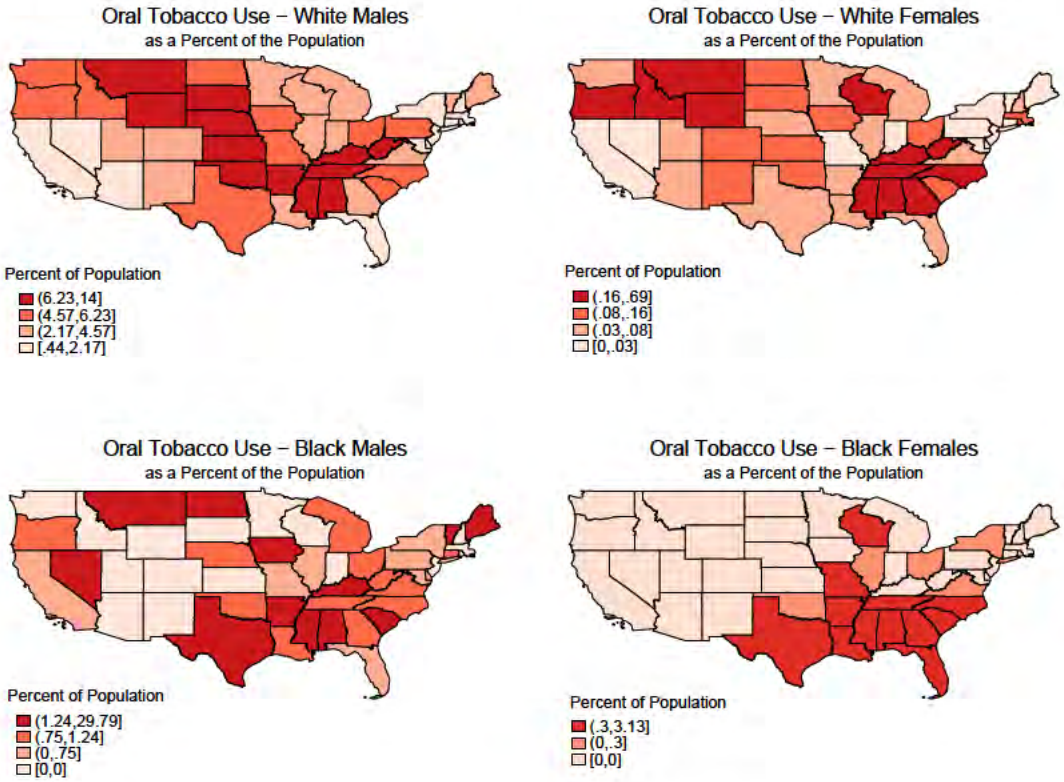

Figure 2: Geographic Variation in Chlamydia Infection by Race and Gender

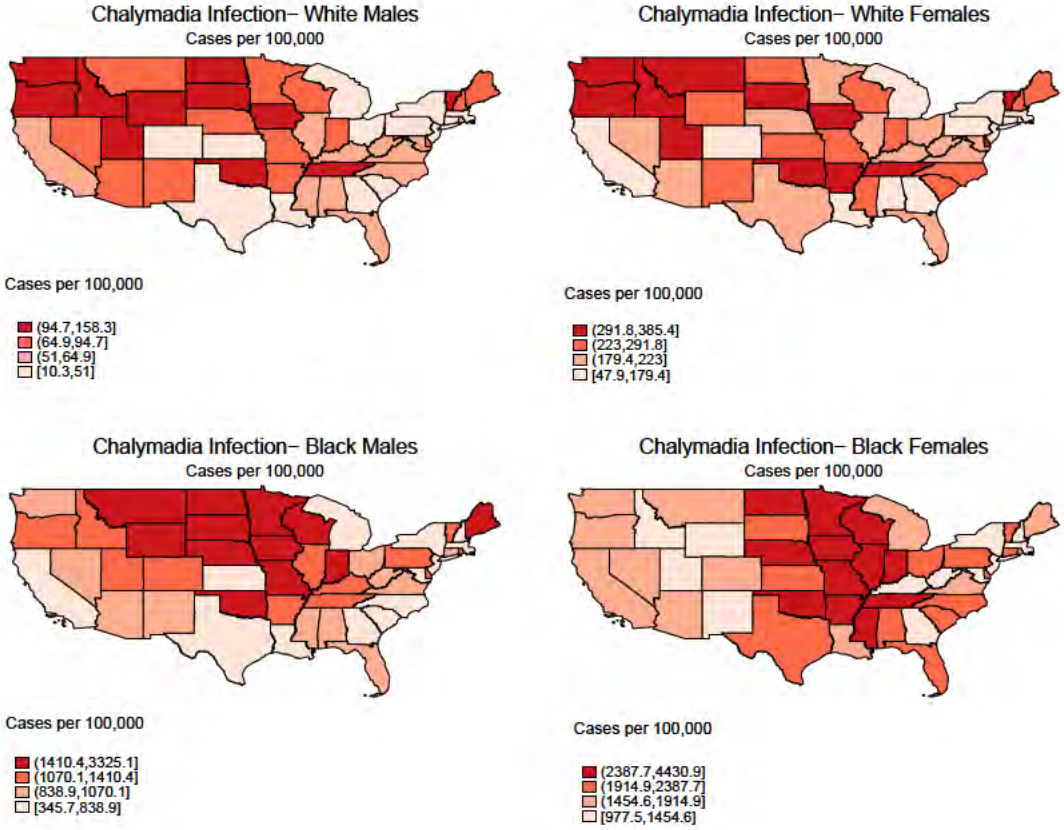

Figure 3: Geographic Variation in Oral Cancer Incidence by Race and Gender