Determinants of Smoking Initiation in South Africa*

Department of Economics | University of Cape Town

Nicole Vellios | ng.vellios@uct.ac.za

Corné van Walbeek | cwalbeek@gmail.com

16 August 2013

Abstract: This paper investigates the individual and household variables that influence the decision to start smoking. The data was drawn from wave 1 of the National Income Dynamics Study (NIDS) of 2008 and the analysis was performed using survival analysis. Based on the international literature and the constraints of the NIDS survey, the following potential determinants of smoking onset were investigated: age, gender, population group, price of cigarettes, geographic location (urban/rural), socio-economic status of parents, whether the respondent's mother was alive when the respondent was aged 15 or not, literacy, parents' smoking behaviour, respondent's alcohol consumption and tobacco control legislation.

Smoking initiation in South Africa typically takes place in the late teenage years and early twenties. Smoking initiation amongst males is much higher than amongst females. For both males and females, the probability of starting smoking is highest amongst the Coloured population. African females have a very low uptake of smoking. Males are more responsive to price changes than females. Depending on the specification, a R1 increase in the price of cigarettes reduces the risk of smoking onset by between 1.1% and 2.8% for males. For females the impact of price on smoking initiation is insignificant. Males and females who have a parent who smokes are more likely to initiate smoking. Females whose mother died before the respondent was aged 15 are more likely to start smoking. The same effect was not found for males. Male and female respondents who currently drink alcohol one or more times a week were more likely to start smoking.

The policy impact of this study is that an increase in the price of cigarettes will decrease smoking initiation, especially amongst males.

* This paper is based on a Masters thesis in Economics by Nicole Vellios supervised by Corné van Walbeek. We would like to thank to thank Emmanuel Guindon, John Tauras, Erik Nesson, Vimal Ranchhod and Asha Sundaram for help in earlier drafts of the paper. We take full responsibility for any errors and omissions.

1. Introduction

A substantial body of literature shows that there is an inverse relationship between the price of tobacco and tobacco consumption (IARC, 2011). Higher tobacco taxes may either prevent people from starting to smoke, or may delay onset. Other than two studies on Vietnam, all studies that consider smoking onset have been performed in high income countries. The importance of the current study lies in the fact that it considers a middle-income country. Smoking is rapidly moving from the developed world to the developing world. Associated with this move in tobacco consumption is a rapid increase in non-communicable diseases. Within this context it is useful to understand what the determinants of smoking initiation are, and if these determinants imply certain policy responses, what these responses should be.

Survival analysis is used to investigate the impact of household and individual variables that might affect a person's decision to start smoking. The National Income Dynamic Study of 2008 is used, together with longitudinal price data on cigarettes. In the past 20 years there has been a substantial increase in the price of cigarettes in South Africa as a result of a comprehensive tobacco control policy.

Before 1990, the South African Government (the National Party) and the tobacco industry had a close and long relationship which resulted in a weak tobacco control policy, despite medical evidence that tobacco was harmful to people's health (Van Walbeek, 2005: 6, 13). In June 1994, the African National Congress came into power with a clean slate, having no historical ties with the tobacco industry. In the same year, the government announced that it would increase the excise tax on cigarettes to 50% of the retail price, to be phased in over a number of years. This new legislation reversed the trend of the previous 25 years of rapidly decreasing real tobacco excise taxes (Van Walbeek, 2005: 23).

In 2001 an amendment to the previous legislation came into effect which further strengthened tobacco control policy. All tobacco advertising and sponsorship was banned, smoking was prohibited in all public and work places and the distribution of free cigarettes and the sale of single cigarettes was prohibited. The hospitality industry was allowed to set aside a maximum of 25% of their floor space for smokers (Van Walbeek, 2005: 26, 27).

Excise taxes have effectively raised the price of cigarettes and other tobacco products which has resulted in a substantial decrease in smoking prevalence. The real price of cigarettes in South Africa increased by more than 100% during the period 1993 to 2003 (Van Walbeek, 2005: 33). Smoking prevalence among young adults (aged 16-24) decreased from about 24 per cent in 1993 to 17 per cent in 2003. The smoking prevalence among people aged 25 to 49 decreased from 39 per cent in 1993 to 28 per cent for the 25-34 age group and to 31 per cent for the 35-49 age group in 2003 (Van Walbeek, 2005: 37). Understanding the factors that influence the smoking onset decision will further strengthen policy aimed at decreasing smoking initiation.

2. Literature review

In 2012, Emmanuel Guindon conducted an extensive literature review on the impact of tobacco prices on smoking onset. Guindon reviews 27 studies that examine the transition between never smoking and smoking. Most studies use data from the United States and to a lesser extent data from other high-income countries, with only one study on a low-income country (Vietnam). In a different paper, Guidon (2009) also looks at Vietnam where he examines the impact of tobacco prices on smoking onset. Studies from high-income countries have weak generalizability to low- and middle-income countries (Guindon 2012:3). Guindon concludes that existing studies do not provide strong evidence that taxes or tobacco prices impact smoking onset. This is an unexpected result as most of the existing literature on tobacco control suggests that higher prices discourage people from smoking. Guindon points to important methodological limitations of these 27 studies which are categorised into two broad groups: data and measurement issues (e.g. how smoking is defined) and methodological issues (e.g. using price as a time invariant covariate).

Guindon (2012) argues that since smoking prevalence in most countries varies substantially between males and females (males tend to have a higher prevalence), estimation models should be analysed separately by gender. Of the few studies that account for gender heterogeneity, the effects of prices on smoking onset are found to be substantially different (Guindon, 2012: 9). Guindon (2012) also points to possible correlations between taxes (and thus price) and tobacco control measures. If taxes are correlated with tobacco control measures (advertising bans and smoke-free policies), estimates of the price or tax responsiveness will be inaccurate (Guindon, 2012: 10).

Survival analysis of smoking behaviour requires data on year of smoking initiation which is linked to prices of cigarettes in that year. Researchers typically create an artificial panel from cross-sectional data. Forster and Jones (2001) use information on smoking histories matched to a long time-series for the tax rate on cigarettes to construct an artificial longitudinal dataset. Kidd and Hopkins (2004) use retrospective time series data on the real price of tobacco (matched to the age at which the individual started smoking) to assess whether price plays an important role in the decision to start smoking.

Some researchers use smoking-dedicated surveys while others use general surveys which include a section on smoking behaviour. Choosing the sample of individuals is influenced by data restrictions. For example, some researchers do not consider respondents above a certain age because of a lack of price data further back in time (e.g. López Nicolás, 2002). Other researchers leave out respondents below a certain age so that the effect of education can be analysed (e.g. Kidd and Hopkins, 2004 and Douglas and Hariharan 1994). Guindon (2009) focuses on males only, since the prevalence among females is so low (1.5% compared to prevalence of 49.2% among males) (Guindon, 2009: 7,8).

Coefficients across studies cannot be compared if the assumption of when individuals are at risk of initiating smoking differs. Douglas and Hariharan (1994) and López Nicolás (2002) assume that individuals are at risk of initiating smoking from birth while Kidd and Hopkins (2004) and Madden (2002) assume that individuals are first exposed to the risk of initiating smoking at age 10. Guindon (2008) assumes that people are at risk at age 14 (because of the limited number of years that price data is available), Forster and Jones at birth and age 4. The mean starting age across studies is generally between 17 and 18 years although variations do occur (e.g. Grignon 2009).

Results on the effect of price on the decision to start smoking are mixed. Grignon (2007) finds statistically significant and moderately large effect sizes: a 10% increase in prices would delay starting by about three to six months, depending on the specifications. Kidd and Hopkins (2004) estimate that a 10% increase in price results in a 1.6% decrease in the demand for cigarettes for males and a 1.2% decrease for females, implying that males are slightly more responsive to price changes than females. Douglas and Hariharan (1994), Forster and Jones (2001) and Madden (2007) find small and/or insignificant effect sizes. Guindon (2009) finds that tobacco prices in Vietnam have a statistically significant and fairly substantial effect on the age of starting smoking.

Considering other variables shows interesting and significant results. Existing literature suggests that education is a key determinant of smoking behaviour. Guindon (2012) argues that education is generally not exogenously determined before or at the time when a respondent initiates smoking. For example, a person who currently has a Masters degree may have been in grade 10 when he or she started smoking at age 16. The level of education recorded as a Masters degree would be inaccurate since the respondent was only in grade 10 at the time the decision to start smoking was made (Guindon 2012:8).

Kidd and Hopkins (2004) and Douglas and Hariharan (1994) include individuals who are assumed to have completed their lifetime educational attainment by restricting their sample to those aged 27 and older and 25 years and older respectively and find that individuals with higher educational attainment are less likely to start smoking, and if they do start, start later. Madden (2002) finds limited evidence that education changes the effect of taxes on smoking onset. Grignon (2007) finds that education tends to decrease the probability to start and delays onset. Grignon also finds that males are more likely to become smokers and start smoking earlier than females. Guindon (2009) estimates that being a student hastens smoking onset by 14.8% compared to individuals who are no longer in school (Guindon, 2009: 16).

Some authors (e.g. Madden 2002) use marriage as a covariate. Madden states that although respondents may not be married at the time when the decision to start smoking was made, marriage might be correlated with an individual's underlying rate of time preference. Presumably people who marry (or intend to marry) have lower discount rates, which would explain why they are less likely to start smoking.

Guindon (2009) includes several interesting covariates including wealth, membership in organisations, peer smoking, employment and literacy. Guindon (2009) finds that wealth (proxied by an asset index) is statistically significantly associated with smoking onset. Individuals in higher wealth categories (especially those in middle wealth categories) tend to initiate smoking earlier compared to those in the bottom quintile. Individuals who are members of mass organisations or clubs and who start smoking delay onset by 14.7%. Individuals who have at any time worked to earn money and who start smoking delay onset by 25.7%. Youths who have friends that smoke have a hazard that is nearly 15 times higher than youths who do not have friends that smoke. Individuals who have ever worked to earn money and those who know how to read and write are more likely to initiate smoking (Guindon, 2009).

Guindon (2009) includes several interesting covariates including wealth, membership in organisations, peer smoking, employment and literacy. Guindon (2009) finds that wealth (proxied by an asset index) is statistically significantly associated with smoking onset. Individuals in higher wealth categories (especially those in middle wealth categories) tend to initiate smoking earlier compared to those in the bottom quintile. Individuals who are members of mass organisations or clubs and who start smoking delay onset by 14.7%. Individuals who have at any time worked to earn money and who start smoking delay onset by 25.7%. Youths who have friends that smoke have a hazard that is nearly 15 times higher than youths who do not have friends that smoke.

Individuals who have ever worked to earn money and those who know how to read and write are more likely to initiate smoking (Guindon, 2009).

The split population duration (SPD) model relaxes the assumption that all subjects will eventually experience the event by 'splitting' the observations into two subpopulations, one that will eventually experience the event of interest and one that will never experience the event. SPD models estimate two sets of coefficients: coefficients for the effects of covariates on the *incidence* of the event occurring (probit component), and coefficients for the effects of the covariates in the *timing* of the event, conditional on the probability of the event occurring (duration component).

Kidd and Hopkins (2004) compare split models to non-split models with estimates for males and females combined, together with separate results by gender. In general the size and level of significance of coefficients in the split model tend to be larger. The non-split model finds that price delays smoking onset but is insignificant, across all three estimates. The split model finds that price significantly delays smoking onset (Kidd and Hopkins, 2004:187, 188).

3. Survival analysis methodology

Survival analysis has several names: duration analysis, event history analysis, reliability analysis or failure time analysis with diverse applications in engineering (e.g. failure of mechanical systems), medical research (e.g. time from diagnosis to death), politics (e.g. survival of political regimes) and economics (e.g. length of employment). Duration models are designed to measure the probability of transition between states, which must non-overlap and must cover all possible states (Douglas, 1998: 49). In the context of this study, the two possible states are whether respondents are smokers or not.

Survival data are described and modelled in terms of two related probabilities, namely survival and hazard. The hazard rate is the subject's risk of experiencing the event of interest, given that he or she has not yet experienced the event. The hazard rate is the fundamental dependent variable in survival analysis. The survival function is derived from the hazard rate. A high hazard rate relates to a rapidly falling survival function and a low hazard rate relates to a gently decreasing survival function (Allison, 1984:11).

It is common in many survival analysis models that the survival times are unknown for a subset of the study group. This is referred to as censoring. A censored observation is defined as an observation with incomplete information, which can be either left or right censored. Right censoring occurs when a person does not experience the event before the study ends, or a person is lost to follow-up during the study period or withdraws from the study. Left censoring occurs when a state is observed but when it began is unknown (Clark et al, 2003: 232). In the analysis presented below, respondents who had not started smoking at the time of the interview are right-censored.

There is a distinction between *informative* and *non-informative* censoring. Non-informative censoring can occur at random or at a time dictated by design. For example, in an alcohol relapse study, an individual who is censored because he has moved cities is non-informative, but if he dropped out of the study because he started drinking again and stopped notifying investigators of his whereabouts, then censoring is informative. If censoring is informative, the hazard rate would be biased. The validity of survival analysis rests on the assumption that censoring in non-informative. This assumption presumably holds true for the dataset used in our analysis, since the probability of sample selection or agreeing to participate in the survey is unrelated to smoking behaviour (Singer and Willett, 2003:318, 319).

Censoring makes standard statistical tools inappropriate even for simple analyses of event occurrence data. A censored event time provides only partial information in that an individual did not experience the target event by the time of censoring. Traditional statistical methods do not provide a way of simultaneously analysing observed and censored event times, while survival methods do (Singer and Willett, 2003:325).

The hazard function assesses the risk associated with each time period. The survivor function cumulates the period-by-period risks of event occurrence to assess the probability that a randomly selected individual will "survive" (not experience the event). At the beginning of time nobody has experienced the event (by definition).

4. Data

The National Income Dynamics Study (NIDS) focuses on income, consumption, expenditure, fertility, mortality, health and education. The base wave of NIDS is a representative sample of the South African population. Although NIDS is designed as a panel survey, we use wave 1 as a cross-sectional survey. We are using the retrospective information which allows the examination of the impact of price and a host of other variables on the decision to start smoking.

Wave 1, completed in 2008, consists of a combination of household and individual level questionnaires which includes 7301 households, 16 878 adult (aged 15 on the day of the interview) individuals and 9616 children. If a person was unavailable at the time of interview a proxy questionnaire was completed for them (n=1753). The total number of individuals is 28 247.

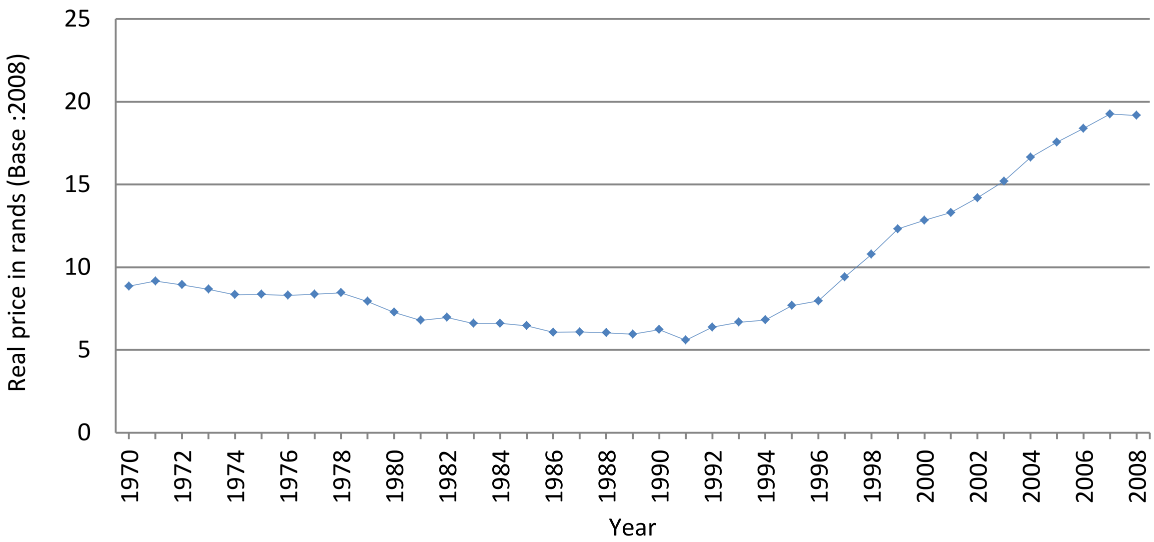
The final sample of 10 873 is obtained by subtracting from the full sample (n=28 258) children (n=9616), proxies (n=1753), respondents older than 48 years (n=4518), non-responses (n=998) and other excluded data (n=500). All respondents aged 48 in 2008 were dropped (n=4518) to reduce the recall error of older respondents. We exclude 500 observations with information that is illogical or missing. For example, ten respondents are dropped because age of smoking onset is reported to be greater than age at time of survey and 273 observations are dropped where we do not have the age of smoking onset. The final sample comprises 4495 males and 6378 females.

There are five smoking-related questions in the NIDS adult questionnaire. A dichotomous variable is created for respondents who are either current smokers or ex-smokers. This variable relates to the questions *"Do you smoke cigarettes?"* and *"Did you ever smoke cigarettes regularly?"*. We constrain the data to all people between the ages of 15 and 48. Of the 10 873 respondents in this age interval, 2069 (19%) indicated that they were current smokers, 358 (3%) indicated that they were ex-smokers and 8446 (78%) indicated that they had never smoked. Combining the current and ex-smokers yields an ever-smoking prevalence of 22%. Ex-smokers are asked the question *"How old were you when you last smoked cigarettes regularly?"*. Both current smokers and ex-smokers are asked *"How old were you when you first smoked cigarettes regularly?"*. The last question asks: *"On average, how many cigarettes per day did you/do you smoke?"*.

Data on cigarette prices are obtained from Van Walbeek (2005) who in turn obtained the data from various sources including the Auditor General, Statistics South Africa, Republic of South Africa and the Tobacco Board. In order to obtain real values in 2008 terms the nominal price series was deflated by the CPI of all goods and services (base 2008).

Figure 1 presents the real price of cigarettes from 1970 to 2008. The real price of cigarettes (2008 base) was R8.85 in 1970 gradually decreasing to R5.59 in 1991 and steadily increasing subsequently. In fact between 1991 and 2007 the real price increased by 245%. The sharp increase in the real retail price of cigarettes since the early 1990s was triggered by rapid increases in the excise tax, but the industry also aided the process by increasing the real retail price by far more than the increase in the excise tax (Van Walbeek, 2005:114).

Figure 1: Real price of a pack of 20 cigarettes (base: 2008)



Source: Van Walbeek 2005, Statistics South Africa (various issues)

5. Descriptive statistics

Table 1 presents unweighted and weighted data showing smoking prevalence by gender for the age group 15 – 48 years. Weights are used to account for different response rates across different race, age and gender groups. We use these weights for the descriptive statistics but not in the survival analysis estimation (Leibbrandt et al, 2009: 28).

Table 1: Unweighted and weighted data showing smoking prevalence by gender for those aged between 15 and 48.

	Males		Females		Males and females	
	Number	Prevalence percentage	Number	Prevalence percentage	Number	Prevalence percentage
Unweighted						
Never smoker	2 791	62.09%	5 655	88.66%	8 446	77.68%
Ever smoker	1 704	37.91%	723	11.34%	2 427	22.32%
Total	4 495	100%	6 378	100%	10 873	100%
Weighted						
Never smoker	5 595 465	62.39%	10 219 127	90.15%	15 814 592	77.89%
Ever smoker	3 373 680	37.61%	1 116 425	9.85%	4 490 105	22.11%
Total	8 969 145	100%	11 335 552	100%	20 304 697	100%

Source: NIDS 2008 data

Ever-smoking (current smokers and ex-smokers) unweighted prevalence among males (37.9%) is much higher than among females (11.3%). The weighted data (37.6% for males and 9.9% for females) is very close to the unweighted data. This suggests that the sample does well in representing the true population. The weighted data reveals that in 2008, there were a total of 4.49 million smokers (3.37 million male; 1.12 million female) between the ages of 15 and 48. Table 2 shows that Coloured males and females are more likely to smoke than other demographic groups.

Table 2: Smoking prevalence by race and gender (weighted and unweighted data)

	Unweighted		*Weighted*	
	Male	Female	Male	Female
African	32.9%	2.8%	34.1%	2.8%
Coloured	62.4%	49.6%	57.6%	46.1%
Asian	47.4%	11.1%	40.1%	8.1%
White	51.5%	47.0%	56.2%	47.6%
Total	**37.9%**	11.3%	**37.7%**	**9.8%**

Source: NIDS 2008 data

The year in which individuals started smoking is calculated using the answers to the initiation age question. Other than the heaping effect previously mentioned, using annual data creates room for measurement error. For example, a person born in 1982 who reported that he or she started smoking at 18 years could have started smoking on his or her 18th birthday (2000) or the day before turning 19 (2001). This could result in measurement error of almost a year.

Matching this information with cigarette price data provides the opportunity to exploit a long time series with sufficient variability in the price of cigarettes to identify the sensitivity of smoking initiation to price.

The mean starting ages (table 3) observed in the NIDS data is similar to the ages observed in the literature, with an average mean starting age of around 18. Looking at the weighted data, it appears that Asians start smoking at the earliest age, but this is estimated on a very small sample (27 males and 10 females) and therefore might be inaccurate. African males have a mean starting age of 18.68, which is the highest compared to other population groups. White females have a slightly higher mean starting age than African females

(19.61 and 19.45 respectively). Coloured females start smoking at an early age (17.62) compared to African and White females.

Table 3: Weighted data for mean starting age by race and gender

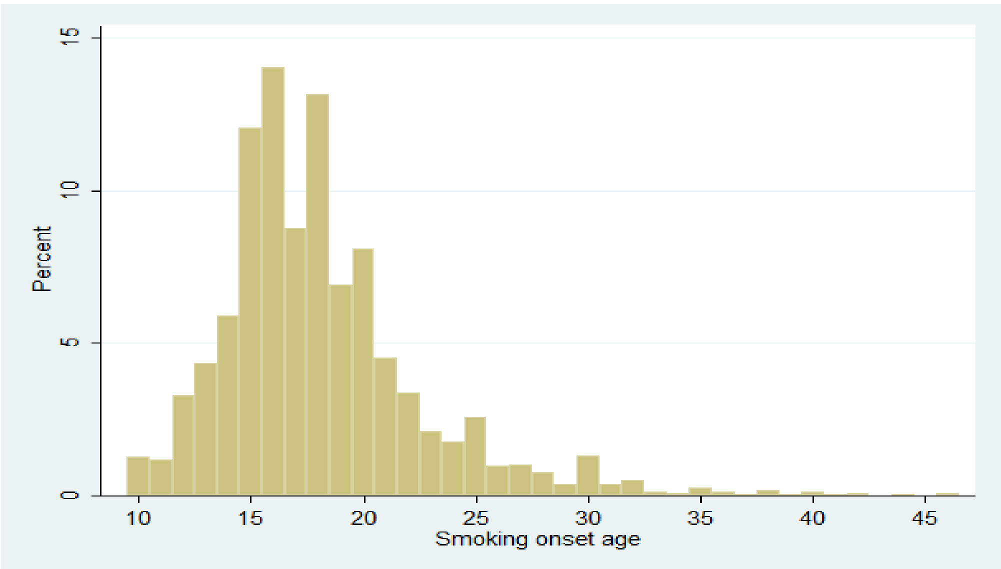
	Male		Female		Male and female	
	Mean initiation age	Std dev	Mean initiation age	Std dev	Mean initiation age	Std dev
African	18.68	4.65	19.45	5.52	18.75	4.74
Coloured	16.81	4.01	17.62	4.64	17.23	4.36
Asian	14.83	2.92	16.34	3.99	15.13	3.17
White	16.52	3.67	19.61	5.33	18.15	4.86
All races (WA)	**18.12**	**4.56**	**18.73**	**5.16**	**18.27**	**4.72**

Source: NIDS 2008 data

Singer and Willett (2003) describe the methods to calculate the hazard and survivor functions. The hazard function can be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve. The non-monotonic shape (figure 3) of the function suggests a log-normal or log-logistic specification might provide a suitable for the data.

Since the question on age of initiation is retrospective, individuals may not accurately recall the age of initiation. Recall bias results in respondents "heaping" their answer to rounded off values (figure 2). Recorded smoking initiation at ages 16, 18, 20, 25 and 30 are much more likely, with a deficit on their neighbouring values. Tauras and Chaloupka (1999) stress that incorrect recall by participants can introduce substantial measurement error. Recall bias is problematic especially when respondents are asked to recall the exact age or year when they started smoking when the event occurred decades earlier (Guindon 2012:7). By constraining our analysis to those who are aged 48 and less, we hope to eliminate some of this recall bias.

Figure 2: Distribution of age when respondent began to smoke cigarettes showing heaping effect



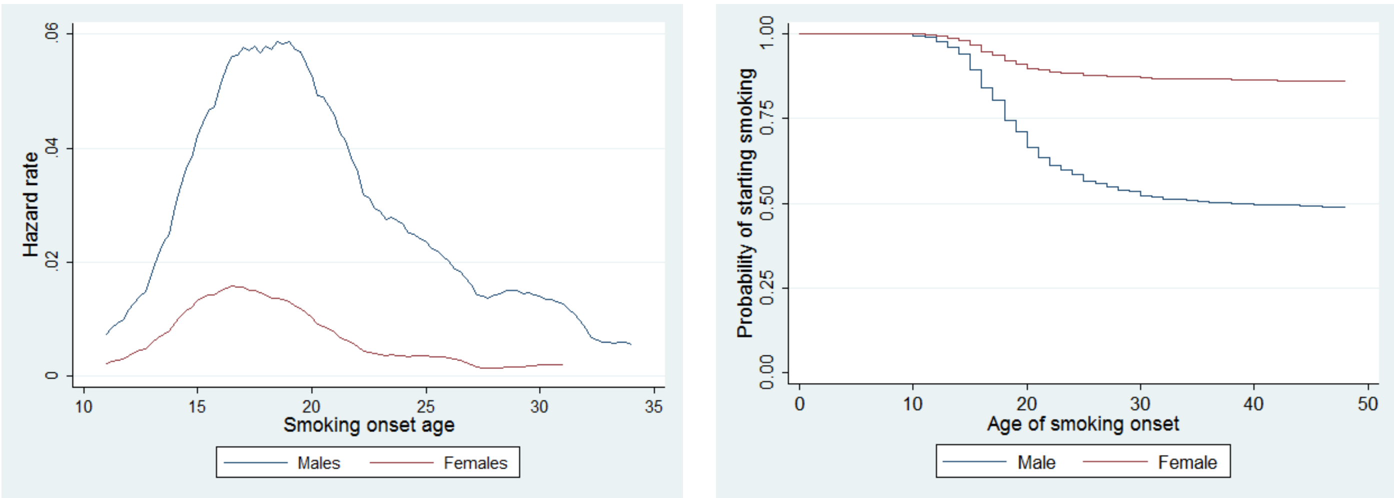
Source: NIDS 2008 data

The survivor function by gender is shown in figure 4. Most respondents (n=10 842) survive the first time period. Over time, as more people start smoking, the percentage of survivors declines. During the teenage years when the hazard of starting smoking is high, the survivor function drops rapidly, especially for males (figure 4). After age 20 when the hazard decreases, the survivor function declines slowly and eventually flattens. Unlike the hazard function which can increase, decrease or remain the same between intervals, the survivor function will never increase. When passing through time periods when no events occur, the survivor function remains steady at its previous level (Singer and Willett, 2003:334).

The value of the survivor function at "the end of time" (age 35) estimates the proportion of the population that will survive past this last observed period (Singer and Willett, 2003:335). 72.6% of respondents are estimated to survive (not start smoking), while 27.4% are estimated to start smoking.

Males are at a much greater risk of becoming smokers than females are (figures 3 and 4). These different hazard functions show that males and females should be analysed separately. The hazard for both males and females reaches a maximum in the late teenage years and declines steeply thereafter. This pattern is similar to that previously reported in the literature. Looking at the Kaplan-Meier curves, we see that 50% of males do not initiate smoking by age 40 while 88% of females survive, i.e. do not initiate smoking. Since these ratios are significantly different, from here onwards we analyse males and females separately.

Figure 3 and 4: Smoothed hazard and survivor functions by gender



Source: NIDS 2008 data

The hazard functions for males and females by population group indicate that the population is heterogeneous: different individuals distinguished on the basis of population group have different hazard functions. The hazard function by population group (not shown) indicates that Coloureds are more likely to start smoking compared to Whites, Africans and Asians for both males and females. The Kaplan-Meier curves show that Coloured and White females have similar survival functions but Coloured females tend to imitate smoking a year or two earlier than White females. The hazard rate for Africans it is extremely low.

6. Empirical methodology

Respondents in the NIDS study were asked at what age they began smoking. This implies annual data. Since respondents are unable to recall precise times, it is not be possible to ask the exact day (or even month) that respondents started smoking. Discrete time is therefore used in this analysis.

Since tobacco prices are recorded at national level in South Africa, the only source of price variation is variation over time. When the price of cigarettes is treated as a time-varying covariate, data are required for each year the individual is at risk, not just the year of starting. The model assumes that each person decides each year whether they want to start smoking or not, taking account of, amongst others, the price of cigarettes in that year. This approach follows that of Forster and Jones (2001). Tax/price enters the likelihood function as a time-varying covariate in all the years during which the respondent is at risk of starting.

We assume that subjects are at risk of starting smoking from the age of 10, in line with Kidd and Hopkins (2004) and Madden (2002). Since males and females have very different smoking prevalence rates, separate models by gender are estimated, as in Forster and Jones (2001) and López Nicolás (2002).

Variables were chosen based on existing literature and within the constraints of the data. The "urban" variable is coded as a dichotomous variable (if respondent lives in a formal rural or tribal authority area the value of "urban" is 0; if respondent lives in a formal or informal urban area the value of urban is 1). The "literacy" variable indicates whether a respondent is literate in either English or in his or her home language (respondent reads or writes very well/fair in home language/English = 1; respondent does not read or write well/not at all in home language/English).

Parents' education is used as a proxy for household income at the time when the respondent was a child or teenager. Given the well-established relationship between education and income, parents' education is likely to be a better indicator of household income at the time the person was susceptible to starting smoking than current income levels. The variable is coded as either parent's highest level of education.

Tobacco control policies in South Africa are controlled for by two variables, one which accounts for legislation in 1995 and one for legislation in 2001. In 1995 the government forced the tobacco industry to print warning labels on tobacco packaging and advertising material. The 2001 legislation banned all tobacco advertising and sponsorship, prohibited smoking in all public and work places, and prohibited the distribution of free cigarettes and the sale of single cigarettes. Both variables are dichotomous.

Two dichotomous variables capture the impact of parental smoking, one for the respondent's mother's smoking behaviour and one for the father's (1=ever smoker, 0=never smoker). Parents in each household were matched to respondents, which was only possible in situations where parents lived in the same household as respondents. Mothers of 3457 respondents (31.8%) were successfully matched. Of these, 450 mothers were smokers, and 3007 were non-smokers (smoking prevalence of 12%). Information on fathers' smoking behaviour was available for only 1293 respondents (11.9%). Smoking prevalence amongst fathers is much higher than for mothers (48%).

The "alcohol" variable is used to establish if there is a link between alcohol consumption and smoking behaviour. Although alcohol consumption at the time of the interview might be different to alcohol consumption at the time that the respondent started smoking, both smoking and drinking are habitual and most people who initiate either habit do so in their teens and early twenties. Also, there is evidence (Terblanche, 2011) that suggests that smokers and drinkers have expenditure patterns that are different to non-smokers and non-drinkers. In particular, smokers and drinkers tend to spend relatively more money on goods and services which have immediate gratification, while non-smokers and non-drinkers spend relatively more money on goods and services which have longer term benefits (e.g. education). The "alcohol" variable is dichotomous (respondent drinks less than one day a week = 0, respondent drinks one or more days per week = 1).

Table 4 and 5 present odds ratios for males and females showing the determinants of smoking initiation controlling for age (using appropriate age dummy variables to account for time throughout). In these tables, each predictor is regressed on the start variable (1 if started smoking, 0 otherwise). Regressions are calculated as follows:

$$logit\ h(t_j) = [a_{10}D_{10} + a_{11}D_{11} + a_{12}D_{12}\ \ldots + a_{35}D_{35}] + B_iX_i \quad \text{where } X_i \text{ is the variable of interest}$$

Only odds ratios are presented, since they can be interpreted more easily. The coefficients can be obtained by taking the natural log of the odds ratio. In tables 4 and 5 the focus is on the individual effect of an independent variable (or set of variables as is the case for race). Each independent variable is regressed on the start variable so that the maximum number of observations for each independent variable can be observed. For example, there are 16 004 observations in the regression 9 of table 4 where the independent variable is the respondent's mother's smoking behaviour.

In the multiple regression models (tables 6 and 7), race is included as a covariate in the model for males (table 6). However, for females (table 7) this was practically impossible because the racial differences were so large that they completely dominated the regression. The pseudo-R^2 for the hazard equation including just race and age dummies (time) was 0.231 (table 5, equation 3). Adding any other variables did not improve the pseudo-R^2. As a result, separate regression models are estimated for African, Coloured and white females (table 7) because the smoking initiation rates between races are so different. As we saw in table 5, Coloureds are 25 times more likely to initiate smoking than Africans, while Whites are 20 times more likely to initiate smoking than Africans. Asian females are excluded from the analysis because there were too few Asian females in the sample who smoked.

Guindon (2009) argues that a split population model, which accounts for the fact that some people will never experience the event (i.e. start smoking) is appropriate. However, a substantial number of studies (see Guindon 2009) use techniques that do not account for split populations. In particular, Guindon was critical of studies that did not account for time in the empirical estimation of the hazard equation. At a purely practical level, although we tried to use a split population model, using the 'spsurv' command in Stata, the model did not converge for any but the most basic specification. Thus in the analysis going forward, we use a non-split survival model, but we take great consideration in accounting for time. Although a split model is not used, many of the techniques and variables used are similar.

7. Results

Higher prices (see equation 1 of table 4 and 5) are associated with a decreased probability of smoking initiation for both males and females, not controlling for any other factor except age. The odds ratio for males is 0.966 and 0.965 for females. The odds ratio of prices for males is 0.966, and the natural log is -0.0346. Thus, in each year from age 10 to 35, the estimated odds of starting smoking are 3.5% (1-0.966) and 3.6% (1-0.965) lower for males and females respectively for every R1 increase in the price of cigarettes. A R10 increase in price is associated with a 29.5% ($1 - e^{-0.0346*10}$) lower odds of smoking initiation for males and a 29.7% lower odds of smoking initiation for females. This implies that males and females are approximately equally price sensitive, when other regressors are not taken into account. More complete specifications are considered later in this section.

When the independent variable is dichotomous, the odds ratio compares the odds of event occurrence for the two groups in every time period. All the independent variables in table 4 and 5 other than price are dichotomous. Odds ratios are symmetric about 1 (Singer and Willett, 2003: 389). If a dichotomous predictor has an odds ratio of 1.0, the odds of event occurrence in the two groups are equal. If the odds ratio is greater than 1.0, the event is more likely to occur in the second group while an odds ratio less than 1.0 indicates that the event is less likely to occur in the second group (Singer and Willett, 2003: 389).

Table 4: Determinants of smoking initiation among males, looking at independent variables separately

Predictor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Price	0.966*** (0.005)										
Urban		1.315*** (0.066)									
Coloured			2.469*** (0.153)								
Asian			1.666** (0.334)								
White			1.569*** (0.167)								
Literate				0.674*** (0.036)							
Parent education (primary)					1.031 (0.073)						
Parent education (secondary)					1.110 (0.074)						
Parent education (tertiary)					0.950 (0.119)						
Without mother at age 15						0.923 (0.108)					
Tobacco control measures (1995)							0.906 (0.059)				
Tobacco control measures (2001)							0.757*** (0.053)				
Father smoker								1.983*** (0.318)		1.899*** (0.361)	
Mother smoker									2.865*** (0.338)	1.703** (0.378)	
Current alcohol use											2.695*** (0.152)
Constant	0.007*** (0.002)	0.005*** (0.001)	0.004*** (0.001)	0.007*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.004*** (0.002)	0.004*** (0.002)	0.003*** (0.002)	0.004*** (0.001)
Controls for age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58 038	58 038	57 494	58 016	50 220	57 073	58 038	5 920	16 004	4 542	57 969
Pseudo R-squared	0.0655	0.0645	0.0754	0.0659	0.0632	0.0621	0.0655	0.0916	0.0795	0.112	0.0802

Standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1
Source: NIDS 2008 data

Table 5: Determinants of smoking initiation among females, looking at independent variables separately

Predictors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Price	0.965*** (0.008)										
Urban		3.521*** (0.313)									
Coloured			24.507*** (2.386)								
Asian			3.675*** (1.269)								
White			19.833*** (2.525)								
Literate				0.699*** (0.055)							
Parent education (primary)					1.116 (0.137)						
Parent education (secondary)					1.978*** (0.209)						
Parent education (tertiary)					2.928*** (0.477)						
Without mother at age 15						1.293* (0.190)					
Tobacco control measures (1995)							0.889 (0.088)				
Tobacco control measures (2001)							0.789** (0.088)				
Father smoker								2.543*** (0.645)		1.397 (0.417)	
Mother smoker									9.408*** (1.457)	5.540*** (1.574)	
Current alcohol use											8.632*** (0.850)
Constant	0.001*** (0.001)	0.000*** (0.000)	0.000*** (0.000)	0.001*** (0.001)	0.000*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.001)	0.001*** (0.000)	0.001*** (0.001)	0.001*** (0.000)
Controls for age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	101 820	101 820	100 985	101 797	88 858	100 105	101 820	8 048	24 298	6 317	101 731
Pseudo R-squared	0.0578	0.0838	0.231	0.0580	0.0665	0.0557	0.0576	0.0546	0.131	0.103	0.0929

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: NIDS 2009 data

As an example, consider the variable "urban" in equation 2 of table 4. An odds ratio of 1.315 indicates that in every year when males are aged 10 to 35, the estimated odds of starting smoking are just over 1.3 times higher at all ages between 10 and 35 for males who live in urban areas compared to males who live in rural areas. The effect of living in urban areas on smoking initiation is greater for females (3.53) compared to males (1.32).

For variables that take on several discrete values, the group representing the omitted level serves as the reference category (Singer and Willett, 2003: 390). The base population group for both genders is African. For both males and females, the probability of starting smoking is highest amongst the Coloured population. The estimated odds of smoking onset are nearly 2.5 times higher for Coloured males than for African males. Asian and White males tend to smoke more than African males (odds ratio of 1.67 and 1.57 respectively). The prevalence of smoking amongst African females is extremely low (2.8%), which results in very high estimated odds ratios, especially for Coloureds and Whites. The odds of smoking onset among Coloured females are almost 25 times higher than among African females. Similarly, among White females, the odds of starting smoking are about 20 times higher than among African females.

The effect of literacy for males and females is consistent. The odds ratio for males (0.67) and females (0.70) indicates that respondents who are literate are less likely to start smoking.

Parents' education levels do not have a significant impact on males' smoking onset decision. In contrast, females who have a parent with secondary or tertiary education are more likely to start smoking compared to females whose parents have no education. At first glance this looks counterintuitive. More educated people have lower smoking prevalence, because they are more aware of the health consequences of smoking (Jha & Chaloupka, 1999: 18). Presumably, they pass this knowledge to their children. However, in South Africa with its large income and educational disparities, especially along racial lines, the level of education of the parents is strongly correlated with race. Thus, in the absence of controls for race, parents' education picks up the racial differences in smoking initiation among females. In a more complete specification (see table 7) the effect of parents' education levels on females smoking initiation becomes much more nuanced.

Females whose mothers died before the respondent was aged 15 are more likely to start smoking compared to females who had not lost their mother by age 15 (odds ratio of 1.293). The odds ratio for males is insignificant which suggests that losing a mother by age 15 does not affect the decision to start smoking for males.

Tobacco control measures resulted in a decrease in smoking initiation amongst both males and females. The odds ratio for the 1995 and 2001 tobacco control variables are approximately 0.9 and 0.8 respectively for both males and females. The 2001 legislation appears to have been more effective than the 1995 legislation (column 8 of tables 4 and 5).

Males who have a father who smokes are 1.98 times more likely to start smoking compared to males who have a father who does not smoke. The effect of father smoking is greater for females, with an odds ratio of 2.54. Mothers' smoking behaviour also impacts females more than males (9.41 compared to 2.87).

Looking at the effect of parental smoking on males when information is known for both parents, the effect of the father's smoking behaviour appears to be greater than the mother's (1.9 compared to 1.7), which is surprising since the individual effects show that the effect of mothers smoking is greater than fathers (odds ratio of 2.865 and 1.983 respectively). This might be because there are fewer observations in the regression with both mother's and father's smoking behaviour (4 542 compared to 16 002). When information is known for both parents of females, mothers who smoke have more of an impact than fathers (odds ratio of 5.540 and 1.397 respectively).

Both males and females who drink one or more days a week were more likely to start smoking. The effect is greater for females (odds ratio of 8.6 compared to 2.7).

Table 7 presents different model specifications for males using multiple regression logit analysis. The first model (equation 1) includes price and race. The "urban" variable is then added to the first model (equation 2), followed by parents' education (equation 3). Individual variables (losing a mother before age 15 and alcohol consumption) are added in equation 4 and 5 and the tobacco control policy variables are added in equation 6.

Table 6: Regression results for multiple regression logit analysis for males

Predictors	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Price	0.973*** (0.005)	0.973*** (0.005)	0.972*** (0.006)	0.977*** (0.006)	0.977*** (0.006)	0.989* (0.006)	0.978 (0.017)
Coloured	2.376*** (0.148)	2.315*** (0.149)	2.277*** (0.169)	2.230*** (0.166)	2.243*** (0.168)	2.160*** (0.163)	2.163*** (0.164)
Asian	1.609** (0.323)	1.567** (0.316)	1.831*** (0.395)	1.992*** (0.431)	1.989*** (0.431)	2.055*** (0.447)	2.062*** (0.449)
White	1.469*** (0.158)	1.421*** (0.156)	1.532*** (0.189)	1.595*** (0.197)	1.598*** (0.198)	1.296** (0.163)	1.301** (0.164)
Urban		1.086 (0.058)	1.114* (0.067)	1.190*** (0.073)	1.181*** (0.073)	1.124* (0.070)	1.125* (0.070)
Parent education (primary)			1.019 (0.074)	1.103 (0.082)	1.110 (0.083)	1.113 (0.083)	1.111 (0.083)
Parent education (secondary)			1.006 (0.074)	1.150* (0.089)	1.144* (0.089)	1.112 (0.087)	1.109 (0.087)
Parent education (tertiary)			0.866 (0.118)	1.012 (0.141)	1.015 (0.142)	0.985 (0.138)	0.983 (0.138)
Literacy				0.650*** (0.043)	0.654*** (0.044)	0.688*** (0.046)	0.686*** (0.046)
Without mother at age 15					0.853 (0.125)	0.859 (0.127)	0.860 (0.127)
Alcohol						2.510*** (0.163)	2.511*** (0.163)
Tobacco control measures (1995)							1.045 (0.101)
Tobacco control measures (2001)							1.094 (0.150)
Constant	0.005*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.007*** (0.002)	0.007*** (0.002)	0.005*** (0.001)	0.006*** (0.001)
Controls for time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	57,494	57,494	49,753	49,740	49,173	49,150	49,150
Pseudo R-squared	0.0772	0.0773	0.0774	0.0806	0.0802	0.0944	0.0944

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: NIDS 2008 data

Equation 1 (table 6) presents the effect of price and population group on the decision to start smoking amongst males. The odds ratio of 0.973 (coefficient of -0.027, not reported) indicates that at every age between 10 and 35, the estimated odds of starting smoking are 2.7% lower for every R1 increase in the price of cigarettes. A R10 increase in price is associated with a 29.5% (1 - $e^{-0.027*10}$) lower odds of starting smoking. Irrespective of the specification, the odds ratio is less than 1 – in fact the point estimates varies between 0.972 and 0.989 – and is significant in all but one case.

The estimated odds of starting smoking are nearly 2.4 times higher for Coloured males than African males. Asian and White males initiate more often than African, but less than Coloured males. Equation 2 adds in the urban variable. Living in an urban area does not appear to impact the decision to start smoking, although the race effect might be dominating the impact of the urban variable.

Equation 3 adds in parents' education which does not seem to affect the decision to start smoking among males. Males who are literate are less likely to start smoking compared to males who are illiterate (equation 4). This is in line with existing literature, e.g. Kidd and Hopkins (2004) and Douglas and Hariharan (1994) that find education to be a key determinant of smoking initiation.

Being without a mother at age 15 (equation 5) does not appear to affect the decision to start smoking among males. Adding in the alcohol variable in equation 6 indicates that those who currently consume alcohol one or more days per week were at greater risk of smoking onset compared to those who drink less than once a week. Inclusion of the two tobacco control variables in equation 7 does not show significant results. This is likely due to the high correlation between tobacco control measures and price. The correlation coefficient between price and tobacco control measures in 1995 is 0.78 while the correlation coefficient between price and tobacco control measures in 2001 is 0.89. The multicollinaerity between these variables is probably responsible for the sizeable increase in the standard error of the price variable in equation 6, making the impact of price on smoking initiation insignificant.

Table 7 presents 3 separate multiple regression logit models for African, Coloured and White females. The first model (equation 1 to 3) includes price, either parent's highest level of education and the urban variable. Equations 4 to 6 include variables for the respondents' level of literacy, whether the person is without his/her mother at age 15 and whether or not they drink alcohol. Tobacco control measures are included in equations 7 to 9.

Since few females started smoking after t_{25}, and because the hazard rate looks parabolic, degrees of freedom were gained by using time and time squared instead of a dummy for each time period. Estimates of the age-related hazard rate using different time indicators (with either time dummies or time and time squared) show similar results. The advantage of the model based on the time dummies is that it does not presuppose a functional form and as such is a non-parametric estimate of the impact of age on smoking initiation. The advantage of using time and time-squared is that is saves degrees of freedom and probably counteracts the effects of heaping, recall bias and measurement error. Sample sizes vary widely, with many more observations for Africans than for Coloureds and Whites.

Price is not significant in the first model (equations 1 to 3). This is in contrast to the finding for males. African females who have a parent with tertiary education are more likely to start smoking than African females whose parents have no education (odds ratio of 1.871). The odds ratio for White females with parents with primary education is very high. This is a function of the small sample size (n=14). African and White females in urban areas are more likely to start smoking, while Coloureds in rural areas are more likely to start smoking relative to Coloureds in urban areas.

Table 7: Regression results for multiple regression logit analysis for females

Predictors	(1) African	(2) Coloured	(3) White	(4) African	(5) Coloured	(6) White	(7) African	(8) Coloured	(9) White
Price	1.013	1.016	0.960	1.026	1.028**	0.967	1.088	0.996	0.792***
	(0.020)	(0.013)	(0.028)	(0.022)	(0.014)	(0.028)	(0.063)	(0.037)	(0.065)
Parent education (primary)	0.615*	0.801	9.175*	0.749	0.887	19.018**	0.753	0.880	17.247**
	(0.154)	(0.124)	(11.947)	(0.192)	(0.145)	(26.880)	(0.194)	(0.144)	(24.257)
Parent education (secondary)	0.784	0.757*	1.283	1.108	0.873	2.003	1.117	0.867	1.959
	(0.188)	(0.115)	(1.322)	(0.277)	(0.144)	(2.305)	(0.280)	(0.144)	(2.211)
Parent education (tertiary)	1.871*	1.100	1.214	2.954***	1.247	1.443	3.002***	1.230	1.320
	(0.684)	(0.368)	(1.263)	(1.143)	(0.434)	(1.679)	(1.163)	(0.429)	(1.506)
Urban	2.427***	0.708**	3.336**	2.372***	0.810	3.154**	2.365***	0.815	3.379**
	(0.475)	(0.098)	(1.723)	(0.483)	(0.117)	(1.638)	(0.482)	(0.118)	(1.770)
Literate				0.487***	0.690***	0.521	0.494***	0.681***	0.442
				(0.108)	(0.098)	(0.390)	(0.111)	(0.097)	(0.333)
Without mother at age 15				2.335***	1.353	2.209	2.352***	1.369	2.314
				(0.757)	(0.410)	(1.325)	(0.762)	(0.415)	(1.380)
Alcohol				15.502***	1.878***	2.935***	15.592***	1.870***	3.219***
				(3.722)	(0.323)	(0.689)	(3.751)	(0.322)	(0.768)
Tobacco control measures (1995)							0.822	1.198	3.337***
							(0.277)	(0.247)	(1.154)
Tobacco control measures (2001)							0.615	1.194	2.486
							(0.278)	(0.349)	(1.605)
Time	1.832***	3.562***	4.384***	1.842***	3.563***	4.588***	1.852***	3.574***	4.639***
	(0.223)	(0.438)	(0.978)	(0.226)	(0.441)	(1.038)	(0.228)	(0.444)	(1.061)
Time^2	0.985***	0.966***	0.963***	0.985***	0.966***	0.962***	0.985***	0.966***	0.961***
	(0.003)	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)
Constant	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	79 384	8 930	3 535	78 376	8 783	3 535	78 376	8 783	3 535
Pseudo R-squared	0.0378	0.0821	0.111	0.0989	0.0897	0.132	0.0996	0.0900	0.144
Age turning point	20.0	18.4	19.6	20.2	18.4	19.7	20.4	18.4	19.3

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: NIDS 2008 data

Model 2, which includes variables for literacy, being without one's mother at age 15 and current alcohol use, shows that the coefficient on literacy for all races is less than one and for Africans and Coloured it is significant, implying that those who are literate are less likely to start smoking. The coefficient for White females is not significant because of the small number of observations in this category (243 White females are literate while only 4 are illiterate). African females who have lost their mother by age 15 are more likely to start smoking compared to other African females who have not lost their mother by age 15. Counterintuitively, the odds ratio for price for Coloured females in the second model is significantly greater than 1, implying that an increase in price increases smoking onset. Current alcohol consumption is associated with an increase in smoking onset for all 3 races, with a very high odds ratio (15.5) for African females.

Results from adding the tobacco control measures (model 3) imply that the tobacco control measures of 1995 resulted in White females initiating more often than before 1995. This is a surprising result which might be due to the high correlation between price and the tobacco control variables. The correlation coefficient for all races (excluding Asian) between price and the 1995 and 2001 tobacco control policy variables is 0.78 and 0.89 respectively. The fact that the odds ratio on price for White females in this specification changed to sharply (from being insignificant in equations (3) and (6) to being very significant in equation (9) suggests that the collinearity between the tobacco control intervention variables and price has greatly influenced the regression results.

8. Conclusion

The first wave of the National Income Dynamics Study of 2008 is used to estimate the determinants of smoking initiation amongst the South African population using a survival analysis model. Looking at respondents between the ages of 10 and 48, different model specifications are estimated with the aim of determining which variables have a significant effect on the decision to start smoking.

Both smokers and non-smokers are considered. For smokers, information on their age of initiation was self-reported. This allowed us to create a longitudinal person-period dataset where each individual was followed from age 10 until they started smoking, or if they never started smoking until the age of 48 or the age they were at the time of the interview in 2008.

The study has limitations. There is likely to be recall bias amongst respondents about their age of initiation given the heaping that is observed in the initiation data. Also, there are some variables that one would want to include in the analyses that were not available, e.g. data on conditions in the household when the respondent was a teenager. In some cases, there were an insufficient number of observations, e.g. parents' smoking behaviour. Some demographic groups, especially Asian females, have limited representation in the survey and thus one cannot conclude anything about such groups with any degree of confidence. Whereas the standard estimation technique for studies like smoking initiation assumes a split population model, we have used an estimation technique that does not account for this.

Considering the overall results, and ignoring some of the peculiarities, a number of general conclusions can be drawn. There is a large gender disparity in smoking initiation. At all ages, males initiate at a higher rate than females. There are very significant racial differences both for males and females, but especially so for females. African females are very unlikely to initiate smoking, while on the other hand White and Coloured females are at a high risk of initiating smoking. There is strong evidence that an increase in the price of cigarettes reduces smoking initiation amongst males. However, there is not much evidence that females respond to changes in prices.

Living in an urban area marginally increases the risk of smoking for males and significantly increases the risk of smoking for African and White females. Coloured females living in rural areas are somewhat more likely to initiate smoking. Being literate decreases the risk of smoking initiation for both males and females. Losing a mother before the age of 15 does not have a significant effect on males' decision to start smoking, but increases the risk of smoking for females. Being a current consumer of alcohol greatly increases the likelihood of initiating smoking for both males and females. It is impossible to determine the impact of the tobacco control legislation of 1995 and 2001 on smoking initiation given the high degree of collinearity between cigarette prices and these legislative interventions.

Findings from this study provide additional evidence of the effectiveness of tobacco prices in reducing tobacco use. Tobacco taxation should remain a major public policy instrument to discourage smoking. Further increases in the excise tax on cigarettes are likely to discourage the smoking habit and to delay onset for those who decide to start.

References

Allison, P. D. 1984. *Event History Analysis: Regression for Longitudinal Event Data.* California: Sage Publications: Issue 46.

Bartels, B. *Advances in Duration Modeling: The Split Population Duration Model*.
Available: http://polisci.osu.edu/sites/polisci.osu.edu/files/AdvancesinDuration.pdf [2012, July 30]

Blecher, E. 2012. A mountain or a molehill: is the illicit trade in cigarettes undermining tobacco control policy in South Africa. *Trends in Organized Crime.* 13(4): 299 - 315.

Clark, T., Bradburn, M., Love, S. & Altman, D. 2003. Survival analysis part I: basic concepts and first analyses. *British Journal of Cancer.* 89: 232 - 238.

Clark, T., Bradburn, M., Love, S. & Altman, D. 2003. Survival analysis part II: multivariate data analysis - an introduction to concepts and methods. *British Journal of Cancer.* 90: 431 - 436.

Clark, T., Bradburn, M., Love, S. & Altman, D. 2003. Survival analysis part III: multivariate data analysis - choosing a model and assessing its adequacy and fit. *British Journal of Cancer.* 89: 605 - 611.

Clark, T., Bradburn, M., Love, S. & Altman, D. 2003. Survival analysis part IV: further concepts and methods in survival analysis. *British Journal of Cancer.* 89: 781 - 786.

Douglas, S. & Hariharan, G. 1994. The hazard of starting smoking: estimates from a split population duration model. *Journal of Health Economics.,* 12: 213 - 230.

Eriksen, M., Mackay, J. & Ross, H. 2012. *The Tobacco Atlas.* Atlanta: American Cancer Society; New York: World Lung Foundation.

Forster, M. & Jones, A. 2001. The role of tobacco taxes in starting and quitting smoking: duration analysis using British data. *Journal of the Royal Statistical Society.* 164 (3): 517 - 547.

Forster, M. & Jones, A. 2003. Corrigendum: The role of tobacco taxes in starting and quitting smoking: duration analysis of British data. *Journal of the Royal Statistical Society,* 166 (3): 441 - 442.

Grignon, M. 2009. An empirical investigation of heterogeneity in time preferences and smoking behaviors. *The Journal of Socio-Economics.* 739 - 751.

Guindon, G. 2012. The impact of tobacco prices on smoking onset: a methodological review. Under review by *Tobacco Control* (personal communication with author).

Guindon, G. 2009. The impact of tobacco prices on smoking onset in Vietnam: duration analysis of retrospective data. Centre for Health Economics and Policy Analysis, McMaster University.

International Agency for Research on Cancer (IARC). 2011. Volume 14. *Effectiveness of Tobacco Tax and Price: Policies for Tobacco Control*. France: World Health Organization.

Jenkins, S. 2001. *Help for 'spsurv'*. Available: http://fmwww.bc.edu/RePEc/bocode/s/spsurv.hlp

Kidd, M. & Hopkins, S. 2004. The hazards of starting and quitting smoking: some Australian evidence. *The Economic Record.* 80: 177 - 198.

Kleinbaum, D. & Klein, M. 2005. *Survival Analysis: A Self-Learning Text.* 2nd ed. New York City: Springer.

Leibbrandt, M., Woolard, I. & De Villiers, L. 2009. *Methodology: Report on NIDS Wave 1. Technical paper no. 1,* South Africa: National Income Dynamics Study. Available: http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/technical-papers.html

López, N. 2002. How important are tobacco prices in the propensity to start and quit smoking? An analysis of smoking histories from the Spanish National Health Survey. *Econometrics and Health Economic.* 11: 521 - 535.

Madden, D. 2002. Do tobacco taxes influence starting and quitting smoking? Duration analysis approach using evidence from a sample of Irish women. Working Paper Series 02/06. Centre for Economic Research, University College Dublin.

National Library of Science.*The Reports of the Surgeon General: The 1964 Report on Smoking and Health*. Available: http://profiles.nlm.nih.gov/ps/retrieve/Narrative/NN/p-nid/60
[2013, January 17].

Schmidt, P. & Witte, A. 1989. Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics.* 40: 141 - 159.

Singer, J. & Willett, J. 2003. *Applied Longitudinal Data Analysis.* New York: Oxford University Press.

Southern Africa Labour and Development Research Unit. 2002. National Income Dynamics Income Study 2008, Wave 1 [dataset]. Version 4.1. Cape Town: Southern Africa Labour and Development Research Unit [producer], Cape Town: Datafirst [distributor].

Statistics South Africa. 2012. *Latest key indicators.* Available:
http://www.statssa.gov.za/keyindicators/keyindicators.asp [2012, November 5] .

Tauras, J. & Chaloupka, F. 1999. Determinants of smoking cessation: An analysis of young adult men and women. Working Paper 7262: National Bureau of Economic Research.

Terblanche, T. 2011. *The Opportunity Cost of Tobacco Use.* University of Cape Town: Unpublished Honours paper.

Van Walbeek, C. 2005. *The Economics of Tobacco Contol in South Africa.* South Africa: University of Cape Town. PhD thesis.

World Bank Report. 1999. *Curbing the epidemic: Governments and the Economics of Tobacco Control.* Washington D.C.: World Bank.

World Health Organization. 2012. *Statistics on South Africa.*
Available: http://www.who.int/countries/zaf/en/ [2012, August 30]

Wheaton, B., Roszell, P. & Hall, K. 1997. The impact of twenty childhood and adult traumatic stressors on psychiatric disorder in *Stress and Adversity over the life course.* I. Gotlib, & B. Wheaton, Eds. Cambridge: Cambridge University Press. 50 – 72.