UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Economía

SELF-ASSESSED HEALTH AS A KEY DETERMINANT OF LIFESTYLES: AN APPLICATION TO TOBACCO CONSUMPTION IN ARGENTINA

Mariana Conte Grand and Vanesa D´Elia

Julio 2008
Nro. 375

www.cema.edu.ar/publicaciones/doc_trabajo.html
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina,
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding <jae@cema.edu.ar>

Self-assessed health as a key determinant of lifestyles: An application to tobacco consumption in Argentina

MARIANA CONTE GRAND*
Universidad del CEMA

and

VANESA D´ELIA
Universidad del CEMA and ANSES

July 2008

Abstract

The relationship between lifestyle choices and health has been widely studied in the epidemiological and economic literature. In the last years, empirical research was directed towards the use of recursive systems with structural equations for a health production function and reduced form equations for lifestyles. As a result, behaviors toward health are taken to be determined by exogenous socio-economic variables. In this article, we show that health is a key determinant of health habits. When people feel well, they adopt less healthy behaviors. We use maximum simulated likelihood for a multivariate 5 equation probit model. In that model, lifestyles (diet, exercise, alcohol consumption and smoking) are a function of exogenous socio-economic variables and self-reported health. Self-reported health varies with socio-economic characteristics and depends on health indicators that are the consequence of lifestyles undertaken in the past (i.e., overweight, blood pressure, diabetes and cholesterol levels). Data is that of adults in Argentina´s 2005 Risk Factors National Survey. We find that health partial effects on lifestyle are much larger having accounted for health endogeneity. Accounting for unobservable variables that jointly determine all lifestyles does not change much the magnitude of our results. Our findings are robust to different specifications.

JEL codes: I10, I12

Keywords: lifestyles, health, multivariate probit, smoking

* We thank Seminar participants at UCEMA and FIEL, for their comments. The views and opinions expressed in this publication are those of the authors and are not necessarily those of the Universidad del CEMA.

1. Introduction

There is a long lasting theoretical literature that examines the relationship of lifestyle choices and medical care with health status. Empirically, it is recognized that variations in medical care access and use cannot by themselves explain health differences but rather personal lifestyle has a significant impact on health (Auster et al 1969 or Fuchs 1986).

The microeconomic rationale for the link between health, medical care and lifestyles can be found in Grossman (1972) seminal work, based on human capital theory (Becker 1965, 1967). Grossman describes health as a consumption and investment commodity. Persons consume more health to avoid the disutility of being ill and invest in health care or lifestyles to have more "healthy" time available for market and nonmarket activities. Hence, a standard health decision model is one where health enters the utility function, individuals have budget and time constraints, and there is a movement equation for health investment, which is produced by household production functions. First order conditions of such models state that individuals maximize their utility by setting to zero the net benefits of additional consumption of each health input. Hence, the demand for medical care and lifestyles (which are also clearly "health inputs") is a derived demand from the basic demand for health. Grossman's household production function for health is analogous to a firm production function.[1] Then, production function efficiency is determined by individuals´ socio-economic characteristics in the same way as, in a firm production function, efficiency is determined by technological characteristics. Lifestyles, under Grossman's view, would then be inputs used to produce more "healthy time".

There is also an ample empirical health economics literature, originated in part in epidemiological and medical studies. Belloc and Breslow (1972), for example, use data from an

[1] Note that Grossman (1972) paper deals mainly with medical care, but as stated by Grossman (2000), lifestyles can be treated as medical care because they are also inputs to the household health production function.

Amaleda county survey to investigate the relationship between seven health behaviors and health outcomes. Those health behaviors are: sleeping from seven to eight hours daily, eating breakfast almost every day, never or rarely eating between meals, currently being at or near prescribed height adjusted weight, never smoking cigarettes, moderate or no use of alcohol, regular physical activity. They found that good health practices are associated with better health and that this association was independent of age, sex, and economic status. That finding was confirmed in two follow-up studies where the relationship between health habits and longevity was explored by using death records (Belloc 1973 and Breslow and Enstrom 1980).

Many other authors have examined the impact of health inputs on health status. For example, Rosenzweig and Schultz (1983), analyze the effect of working, prenatal medical care, age of the mother, etc., on birth weight. Kenkel (1995) models current health as dependent on previous health, lifestyles and schooling. However, one of the main problems with these empirical estimations is individual unobservable heterogeneity. Individual heterogeneity may come from the fact that there are variables that are not taken into account. Several unobservable factors might influence personal choices toward health (for example, genetic differences, distinct past experiences, discount rates, among others).

As a result of this empirical issue, research was directed towards the use of recursive systems with structural equations for the health production function and reduced form equations for all the health inputs. For example, in Contoyannis and Jones (2004), health depends on lifestyle and exogenous variables while lifestyles depend on exogenous variables only. In Balia and Jones (2008), mortality depends on health status, lifestyles and other exogenous variables, health depends on lifestyles and exogenous variables and lifestyles depend solely on exogenous variables. Lifestyles are assumed to depend solely on exogenous variables. However, this seems to us a non intuitive formulation since health status might influence lifestyles.

In this paper, instead on investigating if lifestyles have impact on health, we analyze if lifestyles do depend on other variables than exogenous socioeconomic characteristics. Our hypothesis is that self-perceived health status has a significant influence on lifestyles. Our empirical strategy consists of estimating a multivariate 5 equation model, where lifestyles (diet, exercise, alcohol consumption and smoking) are a function of exogenous socio-economic variables and self-reported health. Then, self-reported health is an endogenous variable that varies with socio-economic characteristics, but also with health indicators (i.e., overweight, blood pressure, diabetes and cholesterol levels, that are the consequence of lifestyles undertaken in the past). By including all lifestyles, we use the information due to unobservable variables that jointly determine health behaviors, and by including an equation for self-assessed health, we account for its possible endogeneity in the lifestyles equations. In that sense, our analysis differs from univariate probit estimations of determinants of lifestyles (as Janzon et al 2005, among others).

While we consider lifestyles in general, we focus our analysis on tobacco because it is the leading presentable cause of death and disability among adults in the world today and smoking is also a major cause of morbidity and mortality in Argentina, where our data originate.[2] We use cross-section data from the *Risk Factors´ National Survey* (ENFR, standing for *Encuesta Nacional de Factores de Riesgo*).

The main contributions of our approach are: to show the importance of self-rated health as a determinant of lifestyles. Feeling well makes people pursue non healthy behaviors; to quantify the underestimation that occurs when simpler empirical models (instead of a maximum simulated likelihood approach for a multivariate probit estimation) are used to explain the determinants of

[2] Cigarette smoking causes approximately 30,000 cancer, cardiovascular and respiratory premature deaths per year in Argentina (see Conte Grand and Pitarque, 2005). Economic costs due to premature mortality account for approximately $740 million, which have to be added to $ 4,330 millon costs due to associated morbidity (see Bruni, 2004).

lifestyles; and, to perform an investigation of smoking predictors (one of the key ones being self-assessed health) based on the Argentina *Risk Factors National Survey*, which was designed and used for epidemiological surveillance rather than for academic work.

This paper is organized as follows. Section 2 details the main characteristics of our data. We explain our empirical strategy in Section 3. Section 4 details our results and Section 5 concludes.

2. Data

We work with the *Risk Factors´ National Survey* (ENFR) collected for the first (and only) time in 2005.[3] The data is of a cross-section type, with no follow-up. There have been other previous surveys related to lifestyles in Argentina, but none of them follows the same individuals for more than one year. The ENFR is a household survey that includes 41.392 individuals aged 18 years old and over in the whole country. It took place at cities with more than 5,000 inhabitants, representing 96% of the country´s urban population and 82% of the whole population (MSAL, 2008).

This study concentrates on adults because youth´s demand for cigarettes might be influenced by quite different factors (for example, more importance may be given to peer pressure than to health). There are papers dealing with smoking among young people in Argentina. Morello et al (2001) assess the prevalence and correlates of tobacco use among high school students in Buenos Aires. They find that current smoking is associated with having a best friend who smokes. Braun et al (2008), analyze marketing strategies of tobacco companies in Argentina and find that the industry developed strategies focused on youth.

[3] The *Encuesta Nacional de Factores de Riesgo* is a survey undertaken jointly by the Ministry of Public Health of Argentina (MSAL) and the National Institute of Statistics and Censuses (INDEC).

Analyzing only adults reduces the sample to 26,376 individuals, representing 14,150,467 persons. As there were incomplete responses of relevant variables included in our model, we are left with complete observations on 21,544 people, which represent 10,958,435 individuals.[4]

The data derive from individuals´ responses during face-to-face interviews. The questionnaire includes 14 modules and covers the socioeconomic situation of the household (and of the individual surveyed) and health issues. The latter are: self-perceived health, weight, diet, tobacco and alcohol consumption, diabetes, exercise, blood pressure, cholesterol, and preventive actions against accidents (use of seat belts and other preventive activities: mammograms and PAPs in women).

For our analysis, we group variables in the ENFR in four categories: 1) Health status, 2) Lifestyles, 3) Health indicators, and 4) Socio-economic characteristics.

Self-assessed health (SAH) is identified as an important endpoint in the health literature. Several studies (as Miilunpalo et al, 1997, Burström and Fredlund 2001 or Quesnel-Vallée 2007) confirm that self rated health is a valid approximation to health status. Here, *SAH = 1* means that individuals rate their health as good, very good or excellent.

Lifestyles are taken to be those classified by McQueen (1987) as the "holy four", the four key health behaviors of those reported in the "Alameda Seven" study (Belloc and Breslow, 1972): Tobacco Consumption, Alcohol Consumption, Diet, and Exercise. In particular, following international conventions adopted by the Ministry of Public Health, *Smokers* are those individuals who smoked during their lives more than 100 cigarettes and now smoke everyday or some days. *Non smokers* include Former Smokers (individuals who smoked in their lives more than 100 cigarettes but now do not smoke) and Never Smokers (individuals who never consumed tobacco

[4] The variables with incomplete observations were: income (13.5 % of observations), alcohol consumption (1.2 % of observations), overweight (8.2 % of observations), blood pressure (0.5 % of observations), diabetes (3.2 % of observations) and cholesterol (1.4 % of observations).

or smoked less than 100 cigarettes in their lives). The *Alcohol* variable differentiates people who drink "in excess", which means a consumption of more than two drinks per day for men and one drink per day for women (over an average of the last 30 days). Diet is good (*Diet* = 1) if the individuals surveyed have eaten fruits and vegetables 5 days in the last week. And, exercise is adequate (*Exercise* = 1) if individuals have performed moderate or strong physical activities for at least 10 minutes per day within the last week.

We differentiate Lifestyles from health indicators. Lifestyles have to do with voluntary health behaviors at the moment of the survey, while health indicators are the consequence of health habits taking place during months or years before the survey. Those health indicators are: *Overweight*, *Blood Pressure*, *Diabetes* and *Cholesterol*. Except for the first case, these variables take the value of 1 to indicate that some health professional told the individual he had risky levels in blood pressure, diabetes or cholesterol. We believe that people realize by themselves when they suffer from overweight (they do not need a medical visit to discover that). Hence, we consider overweighted individuals those with a Body Mass Index ≥ 25. We include stress as a background health condition (*Anxiety* = 1 if they feel moderate or high anxiety or depression).[5]

Finally, socioeconomic characteristics are: "Physical" characteristics (gender and age), Marital status, Education, Occupational status (if employed full or part time, unemployed, or not active, individual income[6] and unsatisfied basic needs), Household characteristics (number of children or if the person lives alone), and Region or Province of residence. We have also included

[5] As we see below, *Anxiety* is treated in our empirical strategy somehow differently than Health indicators.

[6] For income, we have modified the original variable ranging (ranks of household income) taking the mean value of each of the 18 ranks. For the highest rank of income ($5,001 and more), we have estimated the mean value using the information from the Argentina´s Households Permanent Survey (EPH), individuals´ data 2nd semester 2005 (code ITF weighted by PONDIH).

in our analysis if people smoke around the surveyed individual.[7] Table I details the name of each variable, its description and the codes which were used to build it.

Table I. Variables' description

Name	Description	Codes in NRFS
Health status		
SAH	Self assessed health at least good	CISG01=1, 2 and 3
Lifestyles		
Smokers	Smoke in their lives more than 100 cigarettes and now smoke every day or some day	CITA01=1, CITA03=1 CITA04=1 and 2
Diet	Eats fruits and vegetables at least 5 days within the last week	FRUYVER=1, 2 and 3
Alcohol	Consumption of alcoholic beverages in excess	C_EXC_M=1, C_EXC_V=1
Exercise	Performs physical exercise (moderate or strong) at least 10 minutes within the last week	0< CIAF01< 8 or 0<CIAF03< 8
Health indicators		
Overweight	Having overweight	PC_AGR=2 and 3
Blood Pressure	Having blood pressure above normal levels	CIHA03=2
Diabetes	Having diabetes	CIDI01=1, CIDI02=2
Cholesterol	Having cholesterol	CICO03=1
Anxiety	Feels anxiety or depression	CISG06=2 and 3
Socio-economics characteristics		
Gender	Male	CHCH04=1
Age	Age in years	CHCH05
Widow	Widowed	CHCH07=5
Divorced	Divorced or separated	CHCH07=3 and 4
Married	Married or similar	CHCH07=1 and 2
Single	Single	CHCH07=6
Edu 0-6	No education or primary school incomplete	NIVINSTR=1 and 2
Edu 7-11	Primary school complete and secondary school incomplete	NIVINSTR=3 and 4
Edu 12-16	Secondary school complete and tertiary or university education incomplete	NIVINSTR=5 and 6
Edu 17 +	Tertiary or university education complete	NIVINSTR=7
Employed	Employed	C_ACT=1
EmployedPT	If respondent works 45 hours per week or less.	CISL08=1 ó CISL08=2
EmployedFT	If respondent works more than 45 hours per week.	CISL08=3
Unemployed	Unemployed	C_ACT=2
Noact	No active	C_ACT=3
BasicNeeds	Indicator of unsatisfied basic needs	NBI_TOT=1, 2 3 and 4
Income*	Household income per month in pesos	RANGING
Children	Number of people of 18 years old or less in household	CNTDMMBR-MYRS18
LiveAlone	Living alone	TIPO_H=1
Gran Buenos Aires	If region of residence is Gran Buenos Aires	REGION=1
Pampeana	If region of residence is Pampeana	REGION=2
Noroeste	If region of residence is Noroeste	REGION=3
Noreste	If region of residence is Noreste	REGION=4
Cuyo	If region of residence is Cuyo	REGION=5
Patagónica	If region of residence is Patagonia	REGION=6
SmokeAround	Other people smokes around usually	CITA09=1

Note: We do not report here the names of the 24 Provinces of Argentina due to space reasons.

[7] Note we do not include prices variables because in the ENFR survey there are no questions related to tobacco brands consumed by individuals who smoke. We do not include smoking bans either because, at the time of the survey, no binding smoking ban was in place in Argentina.

Table II. Descriptive´s statistics

Variable	Smokers (N= ,947,198) N	Freq./Mean	Non smokers (N= 8,011,237) N	Freq./Mean
Healh and lifestyles				
SAH	2,277,873	77.3%	5,853,200	73.1%
Diet	1,805,096	61.2%	5,907,481	73.7%
Alcohol	715,870	24.3%	1,304,399	16.3%
Exercise	1,523,098	51.7%	3,761,180	46.9%
Health indicators				
Overweight	1,614,933	54.8%	4,935,867	61.6%
Blood Pressure	658,517	22.3%	2,645,391	33.0%
Diabetes	61,228	2.1%	422,685	5.3%
Cholesterol	523,705	17.8%	2,112,846	26.4%
Anxiety	846,779	28.7%	2,060,987	25.7%
Socio-economic characteristics				
Gender	1,644,136	55.8%	3,634,903	45.4%
Age		49		56
Widow	176,367	6.0%	1,123,293	14.0%
Divorce/separated	358,062	12.1%	659,886	8.2%
Married	2,177,536	73.9%	5,730,025	71.5%
Single	235,233	8.0%	498,033	6.2%
Edu 0-6	351,365	11.9%	1,361,220	17.0%
Edu 7-11	1,368,927	46.4%	3,616,473	45.1%
Edu 12-16	811,117	27.5%	1,872,041	23.4%
Edu 17+	415,789	14.1%	1,161,503	14.5%
Employed	2,272,470	77.1%	4,544,486	56.7%
Unemployed	142,962	4.9%	261,699	3.3%
Noact	531,766	18.0%	3,205,052	40.0%
Basic needs	468,341	15.9%	840,192	10.5%
Income		$ 1,010		$ 970
Children	1,924,867	65.3%	4,219,474	52.7%
Alone	185,800	6.3%	661,705	8.3%
Smoking around	1,996,535	67.7%	2,873,093	35.9%
Gran Buenos Aires	1,065,592	36.2%	3,084,378	38.5%
Pampeana	1,034,890	35.1%	2,726,160	34.0%
Noroeste	290,867	9.9%	729,007	9.1%
Noreste	182,325	6.2%	556,737	6.9%
Cuyo	222,116	7.5%	567,167	7.1%
Patagónica	151,408	5.1%	347,788	4.3%

Looking at our descriptive statistics (see Table II), self-assessed health (*SAH*) is better among smokers than among non smokers (77.3% of smokers have at least good SAH while that percentage is 73.1% among non smokers). It seems as if “smoking is good for health”, when what

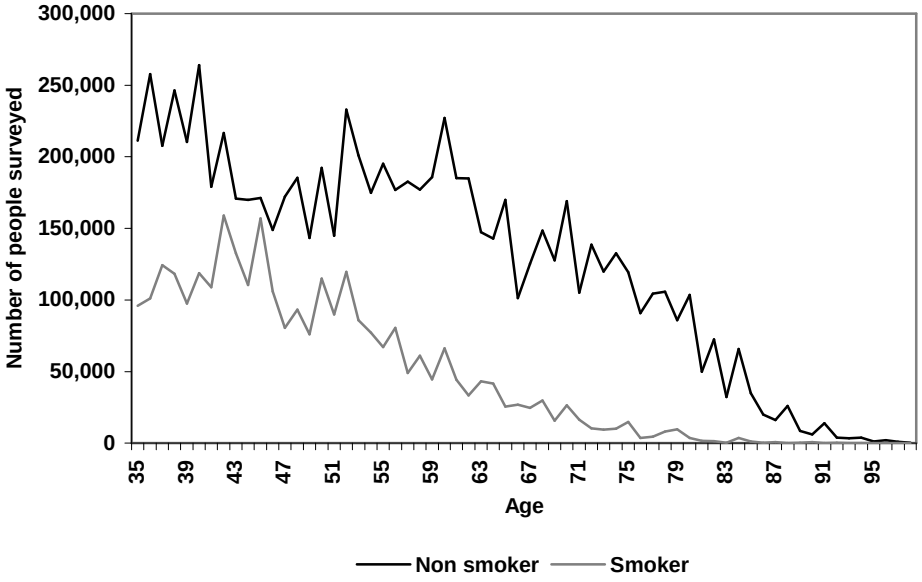
may be happening is that, because smokers feel their health is good, they continue their consumption of tobacco.

Respect to lifestyles, smokers seem to have more tendency to consume alcohol in excess and follow a poorer diet, but more than half of them have exercise as a routine. People with healthy behavior (this is, *Diet* = 1, *Alcohol* = 0 and *Exercise* = 1), can be found in a lower proportion among smokers (25%) than among non smokers (28%). Of those who adopt healthy behaviors, 25% are smokers and 75% are non smokers (this happens while 21% of the population represented by the sample adopts all four health lifestyles). Hence, in general, smokers seem to adopt less healthy behaviors than non smokers.

With respect to health indicators, except for overweight (and stress), smokers seem to have better health indicators. But, part of this may be due to the fact that overweight and anxiety are the only measures that are easy to evaluate without a medical visit. Smokers visit their doctor less and, as a consequence, are less aware that they suffer from risk factors. We find that only 41% of smokers visit their doctor while 50% of non smokers do so (CIAM01_1: medical visits within the last month). So, some of the smokers do not know their health indicators are indicating any risk, and, as a consequence, the impression they have on their health is of poor quality.

Some of the variation in smoking seems to be related to socioeconomic characteristics. The proportion of men is higher among smokers than among non smokers. Mean age is lower in smokers. This may indicate that as people become old more health problems induce them to stop smoking, or that smokers die younger than non smokers. Figure 1 illustrates the relationship between age and being a smoker or not. At age 45 there are almost the same number of people who smoke than those who do not. But, as people get older, the gap increases.

Figure 1. Smokers and non smokers by age



According to Table II, the percentage of people with low education is higher among smokers (58.3% versus 62.1%) There are relatively more unemployed people among smokers than among non smokers. Not satisfying basic needs is more common among smokers despite of the fact that the mean income is higher. Smokers are more likely to have children aged 18 years old or younger (that may be in part due to their lower age). Finally, there is a substantial difference in that 68% of smokers who deal with other people smoking around them, while that fraction is only 36% for non smokers.[8]

3. Methods

The discussion in the previous section is based on differences in frequencies between smokers and non smokers. An econometric analysis of our data should shed light over relationships among tobacco consumption, lifestyles, health and characteristics of the population. The first temptation

[8] Of those who smoke, 48% have people smoking in their home and 39% have people smoking around at work (those percentages are 22 and 15% for non smokers).

when trying to explain the likelihood to be a smoker is to estimate a univariate probit with a dependent variable indicating smoking (1 if individuals do smoke and 0 otherwise) and health behaviors variables, self-assessed health and socio-economic characteristics as explanatory factors.

However, this procedure would not account for two issues. First, it would not consider potential unobservable factors: genetic factors, individuals´ family and peers influence attitudes toward risk, or rate of time preferences, etc. For example, an individual who values the future less (or/and is risk averse), will be less prone to undertake "healthy" practices in order to avoid illnesses or death (see Barsky et al 1997). Having less or more educated parents may also affect lifestyles.

Beyond what is their origin, if there are unobservable determinants that impact simultaneously on lifestyles´ decisions, estimates would be inconsistent. To take into account that fact, we estimate the 4 lifestyle equations (diet, exercise, alcohol consumption and smoking) as a system of equations. Moreover, an univariate probit estimation would ignore the potential endogeneity of SAH. To include that possibility, we add an extra equation of self-assessed health as a function of exogenous characteristics and health indicators.

Hence, our model consists of a system of simultaneous equations for the 4 lifestyles and SAH:

$$\begin{aligned} y_{il}^{*} &= \alpha_l \cdot \mathrm{X}_i + \delta_l \cdot y_{ih} + \varepsilon_{il} \\ y_{ih}^{*} &= \alpha_h \cdot \mathrm{X}_i + \beta_h \cdot \mathrm{Z}_i + \varepsilon_{ih} \end{aligned} \quad , i = 1,\ldots,n; \quad l = 1,\ldots,4; \quad h = SAH. \tag{1}$$

where y_{il}^{*} is a vector of the underlying latent variables of the lifestyles and y_{ih}^{*} is the latent variable for self-assessed health.

For the latent dependent variables, we assume that:

$$y_{il,h} = \begin{cases} 1 & \text{if } y^{*}_{il,h} > 0 \\ 0 & \textbf{otherwise} \end{cases} \qquad (2)$$

Moreover, $\mathbf{X}_i$ is a matrix of exogenous variables and $\mathbf{Z}_i$ is a matrix of the exogenous regressors included only in the SAH equation (overweight, blood pressure, diabetes and cholesterol).[9]

The estimation of this 5-equations model ("Full Model") is performed using Maximum Simulated Likelihood (MSL) for a multivariate probit (MVP) with STATA. These types of estimations are cumbersome because unobservable factors are assumed to be jointly normally distributed.[10]

The error term of the latent equations have a multivariate normal distribution with mean zero and covariance matrix Σ, that is, $\varepsilon_i \sim \text{MVP}(0,\Sigma)$ where $\Sigma = \{\rho_{jk}\}$. It is assumed that the variance-covariance matrix Σ of the cross equation error terms has values of 1 on the leading diagonal, while the off diagonal elements have to be estimated. The parameter ρ_{jk} measure how the unobserved factors influenced health relevant behavior and self assessed health.

As it was mentioned, all the equations in the system can be estimated separately as single univariate probit models, but this procedure does not account for the correlation between the error terms. Maddala (1983), finds that only in the case of independent error terms (ρ not significantly different from zero), the separate ML estimation of univariate probit gives consistent estimates of the parameters. Using a bivariate probit model, Knapp and Seaks (1998) show that the difference between the joint estimation of both equations and the separate estimation of two individual probit models is controlled by the parameter ρ. Then, the estimation of a bivariate probit model

[9] $\mathbf{X}_i$ includes the same variables in all the equations, except for *Smoking Around*, only present in the Smoker equation.

[10] Because the probabilities that enter the log-likelihood function are high dimensional multivariate normal distributions, they are simulated using the Geweke-Hajivassiliou-Keane (GHK) algorithm (see Greene, 2003).

provides an estimate of the asymptotic standard error of $\hat{\rho}$. Therefore, as an alternative to the Hausman test for the exogeneity of a dummy variable, they proposed to compute the statistic $z = \frac{\hat{\rho}}{S.E(\hat{\rho})}$ to test the null hypothesis $H_0 : \rho = 0$. If the error terms are independent (the null is not rejected), the MVP estimation is equivalent to the univariate probit estimations.

The estimation of recursive multivariate probit model requires some consideration for the identification of the parameters. Schmidt (1981) shows that simultaneous probit models suffer from identification problems. Given model in (1), Maddala (1983) shows that, as the number of parameters is larger than the number of probabilities, the parameters in the structural equations are not identified (type 6 model in Maddala). He proposes that at least one of the exogenous variables is not included in the structural equations as regressors.[11] However, Wilde (2000) argues that Maddala concentrates on the special case of constant exogenous regressors and that his statement is valid only for that case and shows that the parameters of the model are identified if there is a varying exogenous regressor. He concludes that for the standard case with varying exogenous variables, the full rank of regressors´ matrix is sufficient for the identification of the parameters.

Hence, here, we perform seven alternative specifications to the full system where specifications differ according to the inclusion of different exogenous variables ($\mathbf{X}_i$) in equation (1). In specification 1, occupation status is modeled as employed and unemployed while not active is the base category. In specification 2, Income is excluded and only the Basic Needs variable is left in the model. In specification 3, Basic Needs is excluded considering it is already taken into account by Income. In specification 4, the variable Living Alone is excluded based on the fact that it may be already captured by marital status. In specification 5, we differentiate those

[11] On the contrary, the structural equations may contain regressors not included in the reduced form equations.

who are employed depending on working hours (part-time/full-time). In specification 6, we replace regions by provinces. In Specification 7, we take together specifications 2, 4, 5 and 6.

To define the specification of the system, we consider the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The AIC is defined as $AIC = -2 \cdot \log L + 2 \cdot K$, and the BIC is calculated, as $BIC = -2 \cdot \log L + \log(N) \cdot K$, where K are the degrees of freedom and N is the sample size (Scott Long, 1997). These criteria represent a trade-off between the goodness of the estimation and the parsimony of the specification. The model with the lowest value of AIC or BIC is chosen as the best.

To better assess if there are advantages from using this more sophisticated empirical strategy, we first test the correlations across our 5 equations. If they are significant, we conclude the system was the correct way to proceed. However, we also carry out simpler estimations to quantify the difference between our estimation and what we would have obtained if we had used a simpler empirical strategy. Hence, after testing correlations (parameter ρ_{jk} , for j = 1,…,4, h = *SAH*), we estimate three alternatives. First, we try a restricted multivariate probit model of 4 equations (one for each lifestyle), but with SAH taken as an exogenous variable. Then, we estimate a bivariate probit model of smoking equation and self-assessed health equation, including *SAH* as a regressor in the smoking equation. Finally we estimate a simple univariate probit for smoking decisions taking *SAH* as an exogenous variable and not considering the rest of the lifestyles. Hence, the alternative models have the following forms:[12]

[12] We have also estimated one alternative for models 3 and 4, which includes the three lifestyles different than smoking as exogenous variables in the smoking equation. No significant differences appear. The exogenous lifestyles have significant coefficients of the expected signs.

Model 2: MVP of Lifestyles (SAH as an exogenous variable)

$$y_{il}^{*} = \alpha_{l}^{'} \cdot \mathrm{X}_{i} + \delta_{l}^{'} \cdot y_{ih} + \varepsilon_{il} \quad , i = 1,\ldots,n; \quad l = 1,\ldots,4; \quad h=SAH \tag{3}$$

Model 3: Bivariate probit Smoking and SAH (as an endogenous variable)

$$\begin{aligned} y_{is}^{*} &= \alpha_{s}^{''} \cdot \mathrm{X}_{i} + \delta_{s}^{''} \cdot y_{is} + \varepsilon_{is} \\ y_{ih}^{*} &= \alpha_{h}^{''} \cdot \mathrm{X}_{i} + \beta_{h}^{''} \cdot \mathrm{Z}_{i} + \varepsilon_{ih} \end{aligned} , i=1,\ldots,n; \quad s=Smoker; \quad h=SAH \tag{4}$$

Model 4: Univariate probit Smoking (SAH as an exogenous variable)

$$y_{is}^{*} = \alpha_{s}^{'''} \cdot \mathrm{X}_{is} + \delta_{s}^{'''} \cdot y_{ih} + \varepsilon_{is} \quad , i=1,\ldots,n; \quad s = Smoker; \quad h=SAH \tag{5}$$

Finally, to complete the analysis, we estimate partial effects to assess the quantitative influence of all variables in the decision of being a smoker, and evaluate if estimation methods yield different results.

4. Results

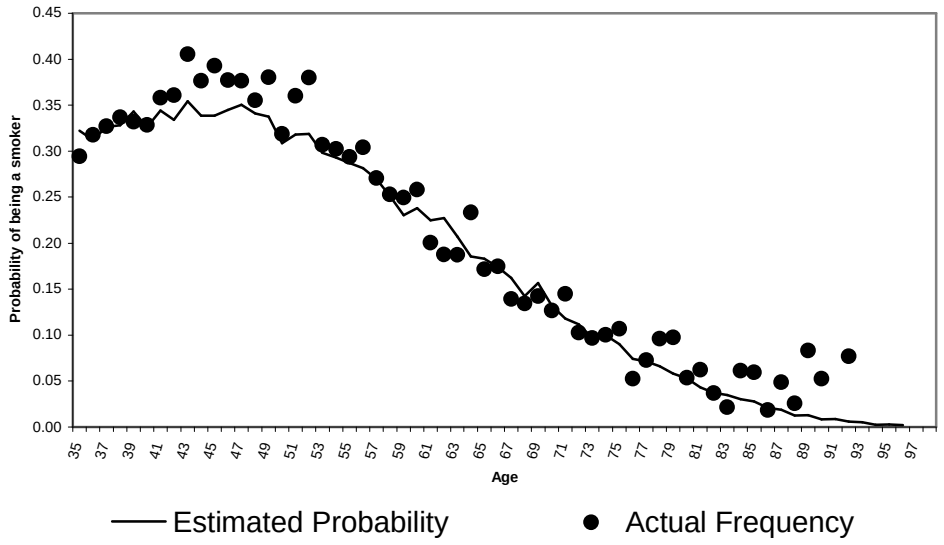
Table III reports the coefficients estimates for the smoking equation under the seven full system alternative specifications. Most of the signs are robust to all specifications. Being a man has a positive effect on the decision to smoke. More age is positively related to smoking habits, which may reflect the fact that older people belong to a generation with high smoking prevalence. But, that relationship has a downward quadratic shape. Figure 2 shows the relationship between Age and the predicted probability of being a smoker (evaluated at the mean of the rest of the predictors) and actual frequencies in our data.

Table III. Smoking equation coefficients estimates under alternative specifications of the Full model

	Specification 1		Specification 2		Specification 3		Specification 4		Specification 5		Specification 6		Specification 7	
	Employed		No income Only basic needs		No basic needs Only income		Without Live alone		With Part and full time		Provinces		Specification 2, 4, 5 and 6 taked together	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Gender	0.2785 ***	0.0211	0.2782 ***	0.0211	0.2820 ***	0.0211	0.2830 ***	0.0210	0.2712 ***	0.0217	0.2781 ***	0.0212	0.2767 ***	0.0216
Age	0.0708 ***	0.0072	0.0708 ***	0.0071	0.0702 ***	0.0072	0.0702 ***	0.0072	0.0721 ***	0.0072	0.0708 ***	0.0072	0.0711 ***	0.0071
Age2	-0.0008 ***	0.0001	-0.0008 ***	0.0001	-0.0008 ***	0.0001	-0.0008 ***	0.0001	-0.0008 ***	0.0001	-0.0008 ***	0.0001	-0.0008 ***	0.0001
Married	-0.1423 ***	0.0335	-0.1427 ***	0.0333	-0.1465 ***	0.0335	-0.1629 ***	0.0317	-0.1481 ***	0.0335	-0.1413 ***	0.0336	-0.1707 ***	0.0315
Divorced	0.1203 ***	0.0372	0.1204 ***	0.0372	0.1191 ***	0.0372	0.1206 ***	0.0372	0.1116 ***	0.0372	0.1201 ***	0.0373	0.1104 ***	0.0373
Widow	-0.0249	0.0442	-0.0253	0.0442	-0.0283	0.0442	-0.0193	0.0441	-0.0346	0.0442	-0.0248	0.0442	-0.0307	0.0442
Edu7-11	0.0595 *	0.0303	0.0590 *	0.0304	0.0465	0.0300	0.0592 *	0.0303	0.0510 *	0.0304	0.0588 *	0.0304	0.0506 *	0.0305
Edu12-16	-0.0209	0.0355	-0.0228	0.0352	-0.0426	0.0348	-0.0201	0.0355	-0.0364	0.0356	-0.0210	0.0356	-0.0363	0.0354
Edu17+	-0.1066 **	0.0416	-0.1103 ***	0.0398	-0.1287 ***	0.0410	-0.1032 **	0.0416	-0.1258 ***	0.0417	-0.1068 **	0.0418	-0.1273 ***	0.0401
Employed	0.0129	0.0268	0.0127	0.0269	0.0132	0.0268	0.0132	0.0268			0.0141	0.0269		
Part-time									-0.0083	0.0300			-0.0046	0.0300
Full-time									0.0180	0.0295			0.0181	0.0296
Unemployed	0.1850 ***	0.0549	0.1857 ***	0.0548	0.1900 ***	0.0549	0.1841 ***	0.0549	0.1817 ***	0.0549	0.1907 ***	0.0550	0.1876 ***	0.0549
Income	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
BasicNeeds	0.0902 ***	0.0311	0.0909 ***	0.0311			0.0926 ***	0.0311	0.0978 ***	0.0312	0.0926 ***	0.0312	0.1064 ***	0.0312
Children	0.0378	0.0252	0.0376	0.0252	0.0428 *	0.0252	0.0215	0.0237	0.0410	0.0253	0.0398	0.0253	0.0283	0.0238
LiveAlone	0.0649 *	0.0343	0.0656 *	0.0343	0.0689 **	0.0343			0.0599 *	0.0344	0.0660 *	0.0344		
Anxiety	0.1702 ***	0.0257	0.1704 ***	0.0258	0.1697 ***	0.0257	0.1691 ***	0.0257	0.1914 ***	0.0257	0.1759 ***	0.0257	0.1907 ***	0.0259
SmokeAround	0.5619 ***	0.0195	0.5623 ***	0.0195	0.5625 ***	0.0195	0.5619 ***	0.0195	0.5589 ***	0.0195	0.5589 ***	0.0196	0.5567 ***	0.0196
SAH	0.2461 ***	0.0498	0.2461 ***	0.0502	0.2434 ***	0.0498	0.2457 ***	0.0498	0.3315 ***	0.0499	0.2612 ***	0.0497	0.3245 ***	0.0502
Cons	-2.63945 ***	0.2082	-2.64065 ***	0.20816	-2.58792 ***	0.2071	-2.59138 ***	0.20695	-2.73469 ***	0.20771	-2.67359 ***	0.2111	-2.69637 ***	0.20935
logL		-55,941		-56,118		-55,967		-55,969		-55,601		-55,400		-54,971
AIC		112,140		112,484		112,182		112,186		111,471		111,157		110,369
BIC		113,169		113,473		113,171		113,175		112,539		112,586		112,076
n		21,544		21,544		21,544		21,544		21,432		21,544		21,432

*Note: ***, ** and * denote significance at 1%, 5% and 10% respectively.*

Figure 2. Estimated Probabilities and Actual Frequencies for Smokers



Being married decreases the probability of being a smoker, but being divorced increases it. This is in line with the literature that links marriage to health (see Duncan et al, 2006, and Khwaja et al 2006 for smoking). Marriage leads to healthier behaviors in some cases (reduced heavy drinking) but leads to poor healthy behaviors in others (sedentary life and weight gain).[13]

Related to socioeconomic status (SES), studies for developed countries find a negative relationship between SES (i.e., income, education, occupational status) and smoking (see Pampel 2004, among others). Income is expected to impact on smoking habits. In particular, higher income may imply purchasing power that can be used to buy cigarettes, but may also mean better health coverage to quit smoking. *Basic needs* captures more broadly SES because it refers to living conditions. *Being unemployed* stands for occupational status. But, of all the variables

[13] Some researchers also point out that part of the link between health and marriage comes from the selection of healthier people into marriage (i.e., healthier people are more likely to marry). See, for example, Clark and Etilé (2006) that, using nine waves of British data find that the correlation between partners' smoking is a consequence of matching in marriage over smoking, rather than bargaining for healthier behavior within the couple.

related to SES, *Education* may be one of the key ones. Education does not change during adulthood and is stably related to continuing to smoke, more than present income or occupational status. The highly positive correlation between education and health has been well documented in the literature.[14] Less educated people are generally less aware of the health risks posed by smoking. Even if people were aware of risks, education brings the ability that helps people to confront that problem and undertake active actions against smoking. Education may also aid in resisting the pressure from others to smoke, view smoking advertisements with skepticism, etc.

We find here no significant income effect on smoking habits. One reason may be the bias always present in that kind of variable. But, we find the expected signs for Basic Needs and Being unemployed. In effect, when basic needs are not satisfied, smoking is higher. Being unemployed (another proxy to low occupational status) goes in the same direction as smoking. For education (the variable we think more reliable), we find that having less than secondary school education is positively linked to be a smoker and, having tertiary or university education is negatively related to the decision to be a smoker.

Beyond socio-economic variables, there are environmental factors that seem to have some role in lifestyles decisions. People who have others smoking around are more likely to be smokers. Feeling anxiety is also positively linked to smoking, which is reasonable since smoking may serve as a coping mechanism for people suffering from anxiety.[15]

The *SAH* coefficient is highly significant (and has a positive sign). Self-perceived health good or better increases the likelihood of being a smoker. So, what seems to be happening is that

[14] Another view is that there is a reverse causality and that in fact health results in more education because healthier students may be more efficient in studying (Currie and Hyson, 1999).

[15] In what refers to regions, Regions North East and Patagonia are the only significant ones, but they have a different sign. Living in the North East appears to decrease the likelihood of being a smoker, while living in la Patagonia increases it. We do not have a good explanation for that fact. The North East region of Argentina is mainly where tobacco is produced, and Patagonia is a region of low population density.

people who perceive their health is good have a higher probability of being smokers rather than non smokers.

In terms of signs and significance, results almost do not differ among specifications. Our results are robust to specification changes. The information criteria (AIC and BIC) favor specification 7, which we then take as our "base model".

Table IV shows the correlation coefficients among equations and the significance of each of them based on the full model. Most of them are significant and have the expected signs. Unobservables which affect the propensity to smoke, are positively related to those which affect the propensity to consume alcohol in excess. That complementarity between decisions to smoke and drink has been documented for other countries. For example, Zhao and Harris (2004), using a multivariate probit model and information for Australia found significant and positive correlations across marijuana, alcohol and tobacco consumption. We also find that unobservables which affect the propensity to smoke are negatively correlated to those which affect the frequency toward a healthy diet and exercising.[16]

Table IV. Correlations between equations for the Full model

	Exercise		Diet		Alcohol		Smoke		SAH
Exercise									
Diet	0.0629	***							
	0.0000								
Alcohol	0.0450	***	-0.0339	**					
	0.0010		0.0150						
Smoke	-0.0261	**	-0.1000	***	0.1360	***			
	0.0280		0.0000		0.0000				
SAH	-0.0102		0.0775	***	-0.0567	**	-0.1580	***	
	0.7460		0.0090		0.0280		0.0000		

*Note: ***, ** and * denote significance at 1%, 5% and 10% respectively.*

[16] The sole correlation coefficient with a sign different than expected (it is positive and not negative) is the one between alcohol and exercise.

Hence, there are no grounds to exclude any of the four lifestyles. We confirm that observable characteristics are unable to completely explain the smoking decision, and that there are unobserved variables that jointly influence lifestyle choices. Significant correlation coefficients between SAH and lifestyle decision equations (including of course the one of smoking) also have the expected signs. In particular, issues that impact positively on self-reported health, also have a positive effect on having a "healthy" diet. On the other side, what increases the probability of being in good health also decreases "unhealthy" behaviors. The correlation between SAH and exercise (a "healthy" behavior) is negative but not significant. Hence, we confirm that self-assessed health has to be modeled as an endogenous variable.

We have shown that we obtain reasonable and robust results when estimating a full 5-equations multivariate probit model (Model 1). Table V shows our results for simpler models.

Table V. Smoking decision coefficients in alternative Models

	Full Model			Multivariate(SAH exo)			Bivariate (SAH endo)			Univariate (SAH exo)		
	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.
Gender	0.2767	***	0.0216	0.2841	***	0.0216	0.2814	***	0.0217	0.2856	***	0.0216
Age	0.0711	***	0.0071	0.0679	***	0.0072	0.0702	***	0.0072	0.0683	***	0.0072
Age2	-0.0008	***	0.0001	-0.0008	***	0.0001	-0.0008	***	0.0001	-0.0008	***	0.0001
Married	-0.1707	***	0.0315	-0.1698	***	0.0316	-0.1704	***	0.0316	-0.1697	***	0.0316
Divorced	0.1104	***	0.0373	0.1158	***	0.0374	0.1146	***	0.0374	0.1179	***	0.0374
Widow	-0.0307		0.0442	-0.0224		0.0443	-0.0285		0.0444	-0.0224		0.0444
Edu7-11	0.0506	*	0.0305	0.0763	**	0.0302	0.0622	**	0.0306	0.0781	**	0.0303
Edu12-16	-0.0363		0.0354	0.0153		0.0343	-0.0145		0.0357	0.0170		0.0344
Edu17+	-0.1273	***	0.0401	-0.0616		0.0385	-0.0988	**	0.0405	-0.0591		0.0386
EmployedPT	-0.0046		0.0300	0.0131		0.0299	0.0002		0.0301	0.0112		0.0300
EmployedFT	0.0181		0.0296	0.0485	*	0.0292	0.0260		0.0298	0.0447		0.0292
Unemployed	0.1876	***	0.0549	0.1956	***	0.0550	0.1914	***	0.0550	0.1957	***	0.0550
BasicNeeds	0.1064	***	0.0312	0.0923	***	0.0312	0.0991	***	0.0314	0.0898	***	0.0313
Children	0.0283		0.0238	0.0247		0.0238	0.0252		0.0238	0.0234		0.0238
Anxiety	0.1907	***	0.0259	0.1214	***	0.0230	0.1653	***	0.0265	0.1230	***	0.0230
SmokeAround	0.5567	***	0.0196	0.5588	***	0.0196	0.5722	***	0.0197	0.5729	***	0.0197
SAH	0.3245	***	0.0502	0.0691	***	0.0244	0.2256	***	0.0539	0.0696	***	0.0244
log L			-54,971			-44,802			-21,407			-11,224
n			21,432			21,432			21,432			21,432

*Note: ***, ** and * denote significance at 1%, 5% and 10% respectively.*

Model 2 has four equations (one for each lifestyle), but SAH is taken to be exogenous. Model 3 is a bivariate probit, where lifestyles, others than smoking, are not considered, but SAH is modeled as endogenous. And, Model 4 is the simplest univariate probit (with SAH taken as exogenous).

The results almost do not change in terms of signs and significance. Gender, age, being divorced, having low education, having unsatisfied basic needs, being unemployed, being anxious, having people smoking around and living alone is positively linked with being a smoker. Being married and having more than secondary school education decreases the chances of being a smoker. The correlation coefficients among equations for models 2 and 3 have similar signs and significance to those in Table IV.

We conclude that there are no major changes in signs and significance of coefficients of smoking decision neither in alternative specifications of the full model, nor in simpler models. If that is the case, there would not be enormous gains from using such complex tools. However, when looking at the magnitude of the coefficients (marginal effects estimated at the means) we do find differences.

On one side, we can see that the major predictors to the probability of being a smoker are having people smoking around and self-assessed health. In particular, as shown in Table VI, for the full model, having people smoking around increases the probability of being a smoker by 18%, while those who perceive their health is fair or better have 10% higher probabilities of being smokers than individuals who believe their health is regular or bad. On the other side, marginal effects are similar across models for the *Smoke Around* variable, but nor for *SAH*. It seems that what changes substantially the result of the impact of self-reported health on the probability of being a smoker is when a specific equation for SAH is included (Model 1 and Model 3). When SAH is taken to be an exogenous variable, its impact on smoking is underestimated. More specifically, an

increase in well-being (SAH good or better) increases the probability of being a smoker in 2% and not 7% or 10% as is the case in the models which consider SAH as a variable explained by socio-economic characteristics as well as by health indicators. This may explain why self-assessed health is not usually considered as an important factor of the probability of being a smoker.

Table VI. Marginal effects in alternative models

Variables	**Model 1** Full model *5 equations*		**Model 2** MVP SAH exo *4 equations*		**Model 3** Bivariate SAH endo *2 equations*		**Model 4** Univariate SAH exo *1 equation*	
Gender	0.0871	***	0.0892	***	0.0884	***	0.0896	***
Age	0.0224	***	0.0213	***	0.0221	***	0.0214	***
Age2	-0.0003	***	-0.0002	***	-0.0003	***	-0.0002	***
Married	-0.0537	***	-0.0533	***	-0.0535	***	-0.0532	***
Divorced	0.0348	***	0.0364	***	0.0360	***	0.0370	***
Widow	-0.0097		-0.0070		-0.0090		-0.0070	
Edu7-11	0.0159	*	0.0240	**	0.0195	**	0.0245	**
Edu12-16	-0.0114		0.0048		-0.0046		0.0053	
Edu17+	-0.0401	***	-0.0193		-0.0310	**	-0.0186	
EmployedPT	-0.0015		0.0041		0.0001		0.0035	
EmployedFT	0.0057		0.0152	*	0.0082		0.0140	
Unemployed	0.0591	***	0.0614	***	0.0601	***	0.0614	***
BasicNeeds	0.0335	***	0.0290	***	0.0311	***	0.0282	***
Children	0.0089		0.0078		0.0079		0.0073	
Anxiety	0.0600	***	0.0381	***	0.0519	***	0.0386	***
SmokeAround	0.1753	***	0.1754	***	0.1797	***	0.1798	***
SAH	0.1022	***	0.0217	***	0.0708	***	0.0218	***
log L	-54,971		-44,802		-21,407		-11,224	
n	21,432		21,432		21,432		21,432	

*Note: ***, ** and * denote significance at 1%, 5% and 10% respectively.*

If self-rate health does affect the likelihood of being a smoker versus that of being non smoker, impacts should be larger for the probability of being a smoker versus a former smoker. Hence, to confirm that, we run the same full model including only the observations corresponding

to smokers and former smokers. We find that, as expected, the marginal effect of SAH is larger: 20% instead of 10%.[17]

5. Conclusions

Being a Smoker is directly linked to feeling well (i.e., have a good or better self-assessed health). As we derive from our data, feeling in good health increases the probability of being a smoker by 10 percentage points. That partial effect is only higher for smoking around (having people smoking around increases 18% the probability of being a smoker). But, what is more important is that the impact of self-perceived health is underestimated when other lifestyles are not considered in the model estimation, but mostly when self-assessed health is considered to be an exogenous variable. In those cases (models 2 and 4), improvements in own health only increases 2% the probability of being a smoker.

We also confirm that there are significant links between tobacco consumption and other risk factors (i.e., the correlation coefficients between lifestyles equations are significant). In particular, unobservable variables that incline individuals to smoke also tend to increase excessive alcohol consumption and decrease exercise and health diet habits.

Finally, our findings are that being a man, older, with low education, divorced, unemployed, having unsatisfied basic needs, living alone, feeling anxiety and having people smoking around are predictors of being a smoker. Having more education and being married significantly decreases that chance. Income and having children under aged are apparently not significant. In that sense, anti-smoking interventions in Argentina should focus on people with

[17] We have also estimated a model with Smokers and Former Smokers including a variable to reflect smoking initiation age. We do so because in the literature (see, for example, Khuder et al, 1999) smoking initiation at an earlier age is accepted to be a strong predictor of smoking behavior later in life and continuation of smoking for a longer period of time. However, here, we found that the higher the age at smoking onset, the higher is the probability to be a smoker at the time of the survey. One explanation to that fact may be that we already control for self-reported health. Hence, individuals who began to smoke before, may have already quit for health reasons.

lower education level and unsatisfied basic needs, and should be directed to avoid indoor smoking.

References

Auster R, Leveson I, Sarachek D. 1969. The production of health: an exploratory study. *Journal of Human Resources* **4**: 412-436.

Balia S, Jones A. 2008. Mortality, lifestyle and socio-economic status. *Journal of Health Economics*, Elsevier **27(1):**1-26.

Barsky R, Juster F, Kimball M, Shapiro M 1997. Preference parameters and behavioural heterogeneity: an experimental approach in the health and retirement study. *Quarterly Journal of Economics* **112(2)**: 537-579.

Becker G. 1965. A theory of the allocation of time. Economic Journal **75**: 493-517.

Becker G. 1967. Human Capital and the Personal Distribution of Income: An Analytical Approach. University of Michigan.

Belloc N, Breslow L. 1972. Relationship of physical health status and health practices. *Preventive Medicine, 1*, 409-421.

Belloc N. 1973. Relationship of health practices and mortality. *Preventive Medicine **3***:125-35.

Braun S, Mejía R, Ling P, Pérez-Stable E. 2008. Tobacco industry targeting youth in Argentina. *Tobacco Control* **17 (2)**:111-117.

Breslow L, Enstrom J. 1980. Persistence of health habits and their relationship to mortality. *Preventive Medicine* **9(4)**:469–483.

Bruni J. 2004. Costos directos de la atención médica de las enfermedades atribuibles al tabaco en Argentina. Proyecto VIGIA.

Clark A, Etile F. 2006. Don't Give Up on Me Baby: Spousal Correlation in Smoking Behaviour. *Journal of Health Economics* **25(5)**: 958-78.

Contoyannis P, Jones A. 2004. Socio-economic status, health and lifestyle. *Journal of Health Economics*. Elsevier **23(5)**: 965-995.

Currie J, Rosemary H. 1999. Is the Impact of Health Shocks Cushioned by Socioeconomic Status: The Case of Birth Weight. *American Economic Review* **89(2)**: 245-250.

Duncan G, Wilkerson B, England P. 2006. Cleaning Up Their Act: The Effects of Marriage and Cohabitation on Licit and Illicit Drug Use. *Demography* **43(4)**: 691-710.

MSAL (2008), Encuesta Nacional de Factores de Riesgo, www.msal.gov.ar.

Fuchs V. 1986. The Health Economy, Harvard University Press, Cambridge.

Grossman M. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* **80(2)**: 223-255.

Grene W. 2003. *Econometric Analysis*, 5th edition. Prentice Hall.

Janzon E, Engström G, Lindström M, Berglund G, Hedblad B, Janzon L. 2005. Who are the "quitters"? a cross-sectional study of circumstances associated with women giving up smoking. *Scand J Public Health* **33(3):**175-82.

Khuder S, Dayal H, Mutgi A. 1999. Age at smoking onset and its effect on smoking cessation. *Addictive Behavior* **24**: 673-677.

Khwaja A, Sloan F, Chung S. 2006. The Effects of Spousal Health on the Decision to Smoke: Evidence on Consumption Externalities, Altruism and Learning within the Household. *Journal of Risk and Uncertainty* **32(1)**: 17-35.

Knapp L, Seaks T. 1998. A Hausman. Test for a Dummy Variable in Probit. *Applied Economics Letters* **5**: 321-323.

Maddala G. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

McQueen D. 1987. A research programme in lifestyle and health: methodological and theoretical considerations. *Revue d´ Epidemiologie et de Santé Publique* **35(1)**:28-35.

Miilunpalo S, Vuori I, Oja P, Pasanen M, Urponen H. 1997. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology* **50(5)**:517-28.

Morello P, Duggan A, Adger H Jr, Anthony J, Joffe A. 2001. Tobacco use among high school students in Buenos Aires, Argentina. *American Journal of Public Health* **91**:1-6.

Pampel F. 2004. Diffusion, Cohort Change, and Social Patters of Smoking. *Social Science Research* **34**:117-39.

Quesnel-Vallée A. 2007. Self-rated health: caught in the crossfire of the quest for 'true' health?. *Int J Epidemiol* **36(6)**:1161-4.

Rosenzweig M, Schultz T. 1983. Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *Journal of Political Economy* **91(5):** 723-746.

Schmidt P. 1981. Constraints on the parameters in simultaneous tobit and probit models. *Structural Analysis of Discrete Data and Econometrics Applications*. MIT Press: Cambridge.

Scott Long J. 1997. *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publications.

Wilde J. 2000. Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors. *Economics Letters* **69**: 309-312.

Zhao X, Harris M. 2004. Demand for marijuana, alcohol and tobacco: participation, levels of consumption and cross-equation correlation. *The Economic Record* **80**: 394–410.