tobacconomics
Economic Research Informing
Tobacco Control Policy

The Impact of

Cigarette Price Increases on the Prevalence of Daily Smoking and Initiation in Mexico

INSTITUTE FOR
HEALTH RESEARCH
AND POLICY
UIC

Citation

Franco-Churruarin F and Gonzalez-Rozada M. The impact of cigarette price increases on the prevalence of daily smoking and initiation in Mexico. A Tobacconomics Research Report. Chicago, IL: Tobacconomics, Health Policy Center, Institute for Health Research and Policy, University of Illinois Chicago, 2021. *www.tobacconomics.org*

Authors

This Research Report was written by Martin Gonzalez-Rozada PhD, Professor of Econometrics, Department of Economics, Universidad Torcuato Di Tella and Fiona Franco-Churruarin, Researcher, Department of Economics, Universidad Torcuato Di Tella. It was peer-reviewed by Corné van Walbeek, PhD, Professor of Economics, University of Cape Town, Cape Town, South Africa; and and Guillermo Paraje, PhD, Professor, Business School, Universidad Adolfo Ibáñez, Santiago, Chile.

This Research Report is funded by Bloomberg Philanthropies.

About Tobacconomics

Tobacconomics is a collaboration of leading researchers who have been studying the economics of tobacco control policy for nearly 30 years. The team is dedicated to helping researchers, advocates, and policymakers access the latest and best research about what's working—or not working—to curb tobacco consumption and its economic impacts. As a program of the University of Illinois Chicago, Tobacconomics is not affiliated with any tobacco manufacturer. Visit *www.tobacconomics.org* or follow us on Twitter *www.twitter.com/tobacconomics*.

Key Messages

Increasing cigarette excise taxes that increase cigarette prices reduces the initiation of daily smoking in Mexico.

The largest reduction in daily smoking prevalence due to increasing prices would occur among youth, the poor, and women, since these groups are more sensitive to price changes.

Daily smoking prevalence increases with wealth and is higher for men than for women.

Price increases delay the age of daily smoking initiation, and this impact is higher for women than for men. Delaying the age at which individuals start smoking makes initiation itself less likely, as fewer people initiate as they get older.

Increasing cigarette taxes is an important public policy because the poor, youth, and women would be among the groups to benefit most.

Executive Summary

There is abundant evidence documenting the negative consequences of smoking. In Mexico, smoking usually starts in adolescence. On average average, people start smoking every day at 18 years old, with many starting as young as age ten. Moreover, early smoking initiation predicts long-term nicotine dependence, affecting smoking behavior for the duration of the person's life. In practice, it is not possible to predictively identify potential smokers. Therefore, there is a compelling need to address the issue of onset and prevalence of smoking in young individuals.

This research report analyzes the determinants of daily smoking prevalence and smoking initiation in Mexico with data from the Global Adult Tobacco Survey (GATS) 2015.

Determinants of prevalence of daily smoking are estimated using a probit model, and price elasticity is also derived. Using two alternative price variables the authors find a prevalence price elasticity of around -0.40, suggesting that an increment of ten percent in cigarette prices could reduce the daily smoking prevalence by 4.0 percent.

This study’s results show that the participation demand elasticity (or prevalence price elasticity) is larger, in absolute value, for women than for men: an increase of ten percent in price is associated with a 4.6-percent decrease in the probability of daily smoking among women versus a 3.5-percent decrease in the probability of daily smoking among men. Additionally, increases in prices affect Mexican youth and older people more than middle-aged people. A ten-percent increase in cigarette prices

reduces the probability of daily smoking by 4.3 percent among youth between 15 and 24 years old, 3.9 percent for the middle-aged population, and 4.4 percent for those more than 65 years old.

Smoking onset, or initiation, is modeled using a split-population model, allowing for the possibility that some individuals will never start smoking. At the mean risk age of 18 years, an increase of ten percent in prices is expected to delay the onset of daily smoking by one year and four months.

These findings suggest that a public policy of increasing excise taxes on cigarettes, with the objective of increasing prices, could be very important to reduce daily smoking prevalence and smoking initiation. Moreover, increasing the excise tax would be a progressive public policy because the poor, youth, and women would be among the groups to benefit most.

1. Introduction

Nicotine addiction is the fundamental reason that individuals persist in using tobacco products, and this persistent tobacco use contributes to many diseases (USDHHS, 2010). Global evidence shows that nicotine dependence symptoms can manifest soon after onset in some adolescents, often well before they start smoking daily or even regularly (DiFranza et al., (2000), DiFranza et al., (2007), Gervais et al. (2006), O'Loughlin et al., (2003), O'Loughlin et al., (2009))., and that early onset predicts long-term adult smoking (Chassin et al., 1990). Since it is not possible to identify those individuals who, after first use of tobacco, will adopt the habit of sustained smoking, the need to prevent that first use is compelling (Klein, 2006; Gervais et al., 2006). Moreover, there is evidence that the public policy of increasing excise taxes, leading to a rise in cigarette prices, is less effective at reducing consumption among those with a longer history of addiction compared to those who have been smoking for a shorter period of time (Gonzalez-Rozada & Montamat, 2019). This evidence highlights the importance of addressing the tobacco epidemic through control policies targeted at early ages, since even delaying by a few years the age at which individuals start smoking can have substantial health benefits.

This research report analyzes the determinants of daily smoking initiation, and in particular the impact of increasing the price of cigarettes, via increasing cigarette excise taxes, on prevalence and the onset of cigarette use. In Mexico, cigarettes are subject to an excise tax with two components, one ad-valorem and one specific, per cigarette. To examine the effectiveness of increasing cigarette prices, through taxes, on those who are most likely to become addicted, this study focuses on the determinants of smoking initiation and smoking onset among daily smokers. There is substantial evidence that, among those individuals who have ever tried smoking, about one-third become daily smokers (USDHHS, 1994), and among those smokers who try to quit, less than five percent are successful at any one time (CDC, 2002, 2004). Consequently, any efforts to reduce tobacco initiation must consider the addictive potential of cigarettes.

This report is organized as follows. Section 2 describes the data used in this study's estimations. Section 3 discusses the methodology and presents the split-population model used to obtain the results that are presented in Section 4. A discussion of the study's findings and a conclusion follow in Sections 5 and 6. Finally, appendices provide further details on the estimation procedures and other analyses.

2. Data

The Global Adult Tobacco Survey (GATS) is a nationally representative household survey of adults 15 years of age and older. The survey systematically monitors adult tobacco use and tracks key tobacco control indicators. It is intended to generate comparable data within and across countries. In Mexico GATS was implemented in 2015 using a multistage stratified cluster sample design. The sample size was 17,765 selected households with 14,664 completed individual interviews. Table 1 shows some descriptive statistics (weighted data) for the whole population considered in the Mexico GATS.

The authors of this study define daily smokers as individuals who self-report as smokers and who smoke a positive number of cigarettes each day. GATS defines that "daily" means smoking at least one tobacco product every day or nearly every day over a period of a month or more. Using survey weights, the proportion of individuals who smoke daily is 7.52 percent, with a high preponderance of men over women. This behavior is consistent with evidence from other countries (Guindon, 2014; Vellios & van Walbeek, 2016).

In Mexico, daily smoking typically starts at the ages between 17 and 20 years old, with men starting earlier. On average, smoking behavior starts at 17.7 years of age for men but nearly 20 for women. At the time of the survey in 2015 smokers paid, on average, around 62 Mexican pesos (US$ 4.05, at the average MXN/USD exchange rate of the survey period) for a package of 20 cigarettes. Women paid slightly more than men.

Around 16 percent of the population has not finished any level of formal education, 19 percent finished primary schooling, 51 percent finished some sort of secondary studies, and 14 percent of the population has attained higher education. These proportions are stable between men and women. The main difference between genders is a higher proportion of women with no formal or primary education. Results also show that 10.5 percent of the population identify as students, and that number is slightly biased towards males. More than half of the population in the survey is employed while nearly six percent are unemployed, although there is a considerable gender difference that is explained by many women being out of the labor force. Additionally, 21 percent of the population lives in rural areas.

Due to the high amount of missing data in reported monthly income, the authors build a wealth index using Principal Components Analysis (PCA). Weights for this index are defined with the first principal component. The variables included in the PCA are binary and reflect socioeconomic characteristics of the person surveyed, such as education above secondary school and household possessions. The index ranges from 0 to 1 and is higher for individuals with more characteristics. The authors create four categories of wealth according to the person's position in the index relative to the quartiles of the distribution of this wealth index. Table 1 shows the average value of the index in each quartile, multiplied by 100 for ease in reading.

Table 2 shows the prevalence of daily smoking by age. Mexicans between 45 and 64 years of age show the highest prevalence, followed by the group aged between 25 and 44, with smokers representing more than eight percent of each group. Smokers make up 6.3 percent of the youngest group. The lowest level of prevalence is among individuals over 65 years old. Disaggregation by gender shows, again, that women smoke much less than men. In both cases, the group aged between 45 and 64 has the highest daily smoking prevalence, but the increase in that age group is relatively higher for women than for men.

Table 1

Mean results of GATS Mexico, 2015

Variables	Total	Male	Female
Daily smoker	7.52% (0.38%)	11.78% (0.68%)	3.60% (0.36%)
Age of daily smoking initiation	18.15 (0.28)	17.67 (0.33)	19.58 (0.59)
Price per pack (20 cigarettes) (MXN)	$61.67 (2.84)	$60.41 (3.39)	$65.30 (5.46)
Highest level of education attained			
No formal education	15.90% (0.60%)	14.29% (0.67%)	17.38% (0.74%)
Primary	19.42% (0.57%)	19.11% (0.80%)	19.70% (0.69%)
Secondary	50.60% (0.75%)	52.52% (1.01%)	48.83% (0.89%)
Tertiary and university	14.08% (0.72%)	14.08% (0.85%)	14.09% (0.87%)
Student	10.49% (0.47%)	11.17% (0.65%)	9.86% (0.59%)
Employment status			
Employed	51.33% (0.66%)	73.35% (0.91%)	31.10% (0.87%)
Unemployed	5.70% (0.32%)	8.81% (0.55%)	2.85% (0.33%)
Out of labor force	42.97% (0.64%)	17.85% (0.79%)	66.05% (0.91%)
Rural	21.34% (1.40%)	21.72% (1.49%)	21.00% (1.40%)
Wealth index (x100)			
1st quartile (poorest)	46.88 (0.23)	47.95 (0.29)	46.83 (0.25)
2nd quartile	67.02 (0.09)	67.00 (0.14)	67.02 (0.11)
3rd quartile	83.53 (0.08)	83.55 (0.12)	83.35 (0.09)
4th quartile (richest)	98.66 (0.07)	98.61 (0.10)	98.70 (0.90)
Age at survey	39.01 (0.26)	38.61 (0.34)	39.55 (0.32)

Note: Linearized standard errors in parentheses.
Source: Authors' calculations

Table 2

Prevalence of daily smoking by age groups

Age group	Aggregate	Men	Women
15-24	6.32% (0.69%)	10.26% (1.26%)	2.38% (0.54%)
25-44	8.08% (0.61%)	12.67% (1.13%)	3.86% (0.46%)
45-64	8.63% (0.81%)	12.77% (1.26%)	5.02% (1.05%)
Over 65	5.29% (0.75%)	9.46% (1.49%)	1.80% (0.48%)

Note: Standard errors in parentheses.
Source: Authors' calculations

Table 3 shows daily smoking prevalence by wealth category. Prevalence of daily smoking disaggregated by wealth shows a different trend than in other Latin American countries (Ciapponi, 2011). In Mexico, daily smoking prevalence (from this point on the phrase "smoking prevalence" in this report will refer to daily smoking prevalence) is highest among individuals in the richest half of the population, at around 8.5 percent. Prevalence is higher for men than women. The overall prevalence by income group seems to be driven by men's prevalence. Smoking prevalence for men in the poorest quartile based on wealth is around nine percent, while in the richest quartile it is around 13 percent. These trends are the same when considering reported monthly income. As stated above, the

Table 3

Prevalence of smoking by wealth quartiles

Wealth quartile	Aggregate	Men	Women
1st quartile (poorest)	5.94% (0.63%)	8.96% (0.54%)	3.22% (0.84%)
2nd quartile	7.39% (0.67%)	11.85% (1.22%)	3.45% (0.54%)
3rd quartile	8.55% (0.78%)	13.41% (1.40%)	4.13% (0.77%)
4th quartile (richest)	8.48% (0.85%)	13.28% (1.59%)	3.66% (0.70%)

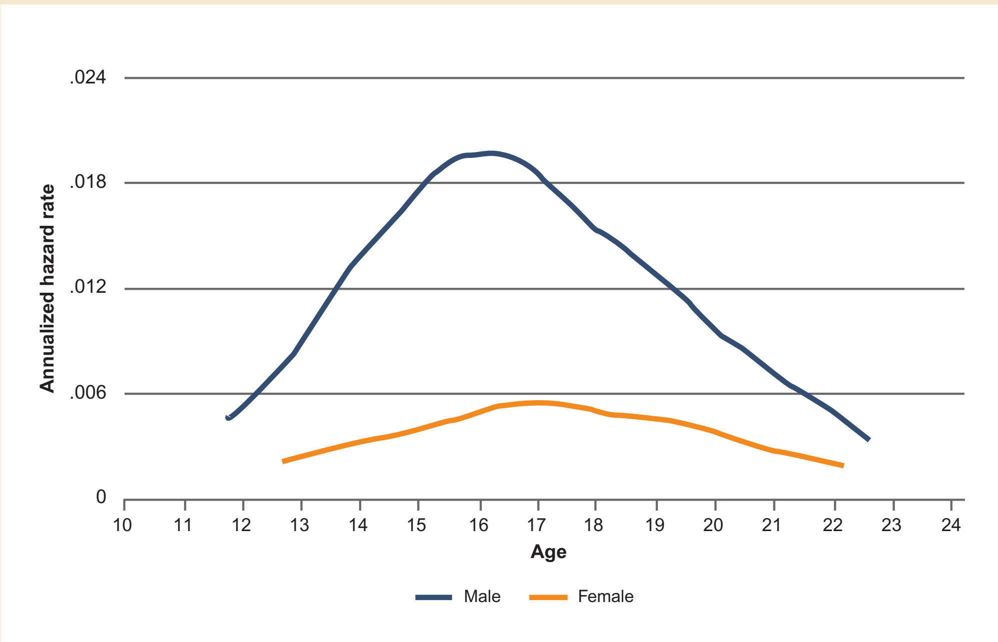
Note: Standard errors in parentheses.
Source: Authors' calculations

authors prefer to report results using the wealth index because the income question has a large non-response[1] that may induce an endogeneity problem in this study's estimations.

As mentioned in the introduction, delaying the age at which individuals start daily smoking can have substantial health benefits. Therefore, this study assesses the impact of increasing cigarette prices, through taxes, on the onset of daily smoking. An "increase in the onset of smoking" means that the age at which individuals start smoking daily is delayed. For a first look at this issue, Figure 1 shows the hazard of initiating the habit of everyday smoking. As shown in the figure, people have a positive risk of initiating smoking daily from around age 12 or 13. Accordingly, in the modeling below, an individual is considered to be at risk of daily smoking at the age of ten. Male teenagers around the age of 16 have the highest risk of picking up a smoking habit, while for women this risk is highest at around 17 years old. The hazard of initiating daily smoking increases sharply around the age of 12 for men and falls after the age of 17.

Figure 2 shows the cumulative hazard function of initiating a daily smoking habit by gender. The cumulative hazard function in Figure 2 describes the total amount of risk of initiating smoking (from this point on "initiating smoking" means initiating daily smoking) that has been accumulated up to each age in the x-axis. As can be seen in the figure, in the case of men, the cumulative hazard of starting smoking begins to increase at 12 years old, while in

Figure 1
Smoothed hazard function



Source: Authors' calculations

[1] 57.7 percent of the sample answered "Don't know" or "No response".

the case of women it seems to start later, around the age of 14. Around 16 years of age, the figure shows that the cumulative hazard of starting smoking among men is at least two times greater than that of women. This relationship is maintained or even accentuated in older ages.

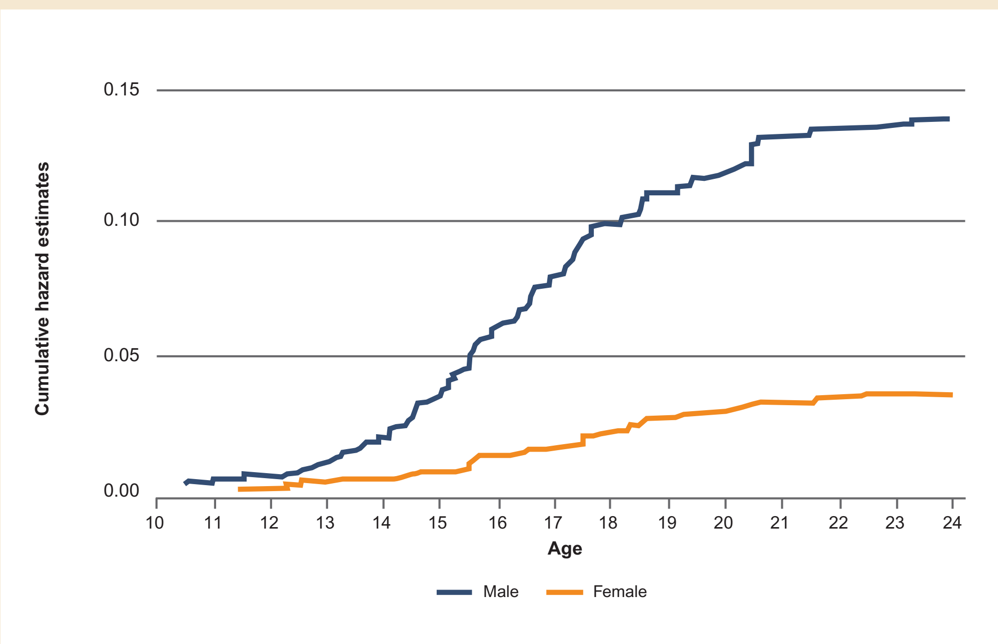
Moreover, the slope of both curves is very different, suggesting that between 13 and 20 years old the risk of initiating smoking for men increases at a faster rate (steeper slope) than for women. For women, the greatest acceleration in the risk of starting to smoke occurs between the ages of 15 and 20, but in a much more attenuated way than for men. After the age of 20, the slope of both cumulative hazards stabilizes, and by the ages of 23 to 24 men have a risk of picking up smoking that is around three times higher than the risk for women. The figure suggests the hazard function of smoking onset shows positive duration dependence for both men and women. Figure 1 shows an annualized hazard rate, whereas Figure 2 shows the cumulative hazard in terms of months.

2.1. The price variable

A common concern when using self-reported prices in the estimation of smoking prevalence with a cross-sectional database such as GATS is its potential endogeneity. To address this potential problem two different price variables are constructed. The first price variable assigns to smokers the self-reported price paid for the last purchase and uses a random regression imputation (sometimes called stochastic regression imputation) to assign a price for those non-smokers in the sample. The second price variable assigns to smokers and non-smokers the average of the self-reported price by primary sampling unit (PSU).

Figure 2

Nelson-Aalen cumulative hazard estimates



Source: Authors' calculations

In the first price variable, the random regression price imputation for those non-smokers is as follows. First, a regression equation is estimated for the smokers in this study's sample using as a dependent variable the self-reported price paid for the last purchase and as explanatory variables: gender (a binary variable equal to one for women), age, labor and education categories, wealth quartiles, binary variables for students and homemakers, and survey strata fixed effects (Table A2 in Appendix A shows the results of this equation). Then, the authors input prices for non-smokers using the predicted price from this regression plus a random draw from a normal distribution with mean and standard deviation equal to the mean and standard deviation of the residuals. The inclusion of a random error adds variability to the deterministic predicted price and is more capable of reproducing the correlation between the self-reported price and the explanatory variables (Little & Rubin, 2020). The average of this random imputation price in the sample is US$ 4.09.

The second price variable is constructed by estimating a regression for the smokers in this study's sample with the self-reported price as the dependent variable and using as explanatory variables dummy variables for each PSU in this study's sample. Then the predicted price from this regression, which is the average self-reported price by PSU, is assigned to both smokers and non-smokers. The average of the self-reported price by PSU is US$ 4.08.

Table 4 summarizes the distribution of these measures of prices by showing the average price by deciles of each variable. The first two columns refer to the random imputation price, whereas the last two columns refer to the average self-reported price. In both cases, the first column shows the price variable measured in logs, and the second column shows the actual price

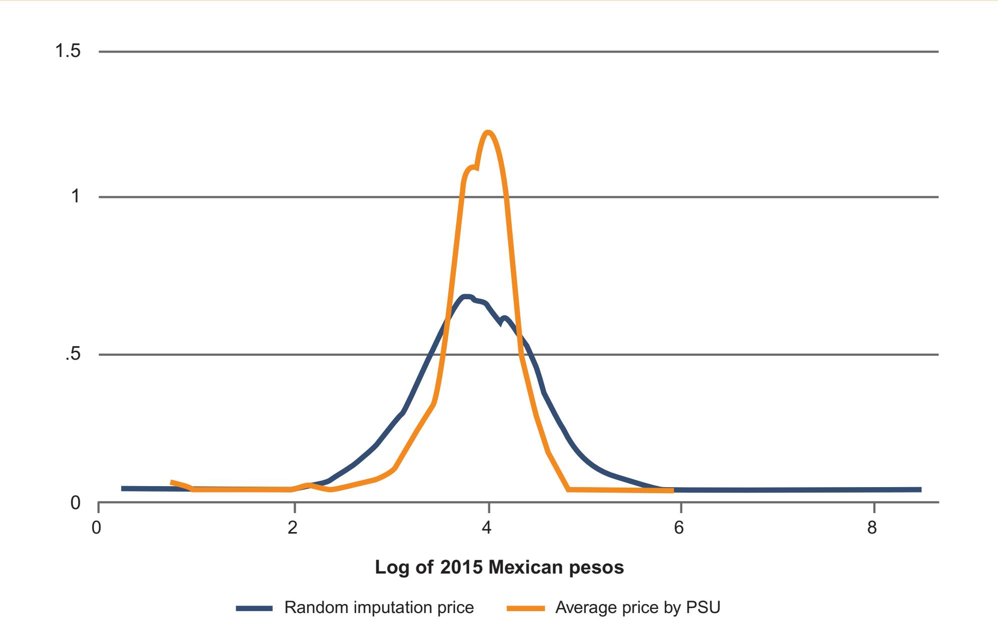
Table 4
Average price by deciles

Deciles	Average random imputation price		Average self-reported price	
	Logged price	Actual price	Logged price	Actual price
1	2.71	15.81	2.99	22.49
2	3.22	25.01	3.56	35.35
3	3.46	31.77	3.73	41.60
4	3.64	38.26	3.83	46.14
5	3.81	45.03	3.92	50.66
6	3.96	52.39	4.05	57.227
7	4.12	61.82	4.13	62.11
8	4.31	74.84	4.15	63.65
9	4.52	91.87	4.26	70.62
10	4.99	160.38	4.49	90.71

Note: Actual prices are in Mexican pesos of 2015.
Source: Authors' calculations

Figure 3

Kernel density estimates of prices



Source: Authors' calculations

(in this case, in Mexican pesos of 2015). From the comparison of both price variables, it seems that the random regression imputation price estimates smaller prices at the lower deciles but larger prices at the higher deciles of prices. This evidence suggests the self-reported price by PSU has a lower variability than the random regression imputation price.

Figure 3 plots the kernel density estimates of the two price variables (measured in logs of Mexican pesos). It is clear from the figure what Table 4 suggests: the variability of the random regression imputation price is larger than the imputed self-reported price by PSU.

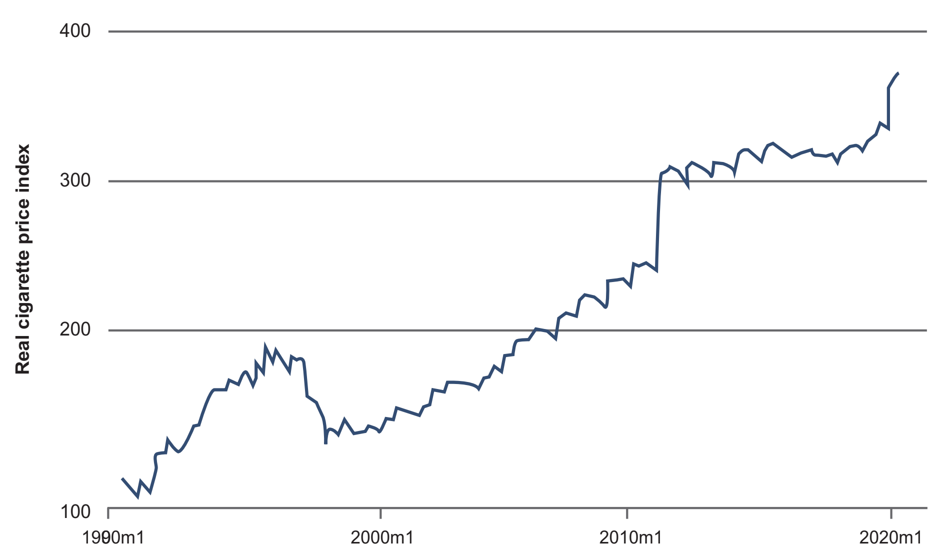
For the estimation of the impact of cigarette prices on smoking onset, the data first needs to be transformed into a pseudo panel in order to assign to each smoker the cigarette price at the smoking initiation date. For this task a monthly index is used for the real price of cigarettes. This index is calculated with data from the National Institute of Statistics and Geography (INEGI) for the period January 1990 to May 2015. INEGI elaborates an index that aggregates cigarette prices in different package types and cities. This study's index is calculated as the ratio of INEGI's index to the consumer price index (which is also produced by INEGI). Figure 4 shows the evolution of the index used in this study.

3. Methodology

This study uses survival analysis estimation, focusing not only on smoking probability but also the onset of cigarette use. For smoking prevalence the authors estimate a probit model, and for smoking onset a split-population model is used (Schmidt & Witte, 1989).

Figure 4

Evolution of real price of cigarettes



Source: Authors' calculations

Since GATS has a single record per individual for their starting age of smoking, the authors construct a pseudo panel. Based on the reported age of initiation, the authors create for each individual a duration spell. Duration refers to the time that elapses between the risk age of smoking onset (the age at which individuals begin to be at risk of starting to smoke) and the age of starting. Therefore, a spell begins at the risk age (which the authors assume to be ten years old) and either ends in the period the individual reported to have started smoking or at the survey date if they never started.

The main idea behind the use of a split-population model is to account for the fact that not all individuals who have an incomplete spell will eventually start smoking, as opposed to the traditional assumption of standard duration models that they all will. The duration process applies then only to those individuals who are predicted to eventually "fail." The likelihood of each observation is weighted with the probability that the individual will ever start smoking. Formally expressed, the log-likelihood function to be maximized is:

$$ln\,(L) = \sum w_i\{c_i ln\,[\Phi(\alpha' z_i) f(t/s = 1, x_i(t))] + (1 - c_i) ln[1 - \Phi(\alpha' z_i) + \Phi(\alpha' z_i) S(t/s=1, x_i(t))]\} \quad (1)$$

where c_i is a dummy variable equal to 1 if individual i ever smoked and 0 otherwise, s_i is another dummy equal to 1 if the individual will eventually start smoking and 0 if they never do. Φ is the standard normal cumulative function and z_i are time-invariant covariates. f refers to the chosen conditional density function to model duration, S is the respective survival function, and w_i is a survey weight. $x_i(t)$ are time varying covariates including the price of cigarettes.

The contribution to the log likelihood (1) for individual i observed smoker in the sample (c_i = 1, uncensored observations) is simply the natural logarithm of the probability of daily smoking, $\Phi(\alpha' z_i)$, multiplied by the probability density function of starting smoking at the observed starting age $f(t/s=1,x_i(t))$. For those i observed not starting smoking (c_i = 0, censored observations) the contribution is the natural logarithm of the probability of no daily smoking, $1-\Phi(\alpha' z_i)$, plus the probability of starting after the age observed in the survey, $\Phi(\alpha' z_i)S(t/s=1, x_i(t))$ (Forster & Jones, 2001).

Notice that in the traditional split-population model the probability to start smoking is constant for all individuals, $\Phi(\alpha' z_i) = k$, while here with a more general setup not all individuals have the same probability of starting to smoke. Smoking prevalence depends on the socioeconomic characteristics of the individuals. That is,

$$Pr(y_i = 1 \mid z_i) = \Phi(\alpha' z_i) \qquad (2)$$

where y_i=1 indicates that individual i smokes and z_i is a vector of explanatory variables including the log of the imputed self-reported cigarette price; the wealth index; a dummy for female, rural residence, being a student; and labor and age categories.

Using (2) as part of the log likelihood (1) means that instead of estimating a single coefficient k for smoking prevalence, as in the traditional split-population model, the authors need to estimate the coefficients of a nonlinear function. This makes the log likelihood (1) to be maximized highly nonlinear and difficult to fit because the convergence to a maximum is more likely to fail (Jenkins, 2001). To avoid this problem, the strategy adopted here is first using a probit model to estimate equation (2), $\Phi(\alpha' z_i)$ and then introduce this estimation into equation (1) to estimate the duration coefficients. This procedure has the advantage of allowing the authors to compute the prevalence elasticity directly from equation (2) using,

$$\epsilon_i = \frac{\partial \Phi(\alpha' z_i)}{\partial \ln(cp_i)} \times \frac{1}{\Phi(\alpha' z_i)} \qquad (3)$$

where ln (cp_i) is the log of the imputed self-reported cigarette price. Equation (3) is a function that gives a different elasticity for each i. Therefore, when reporting the estimated elasticity, the average prevalence price elasticity is presented over the relevant group of people.

This study follows Forster and Jones (2001), who also use a split-population model to study the effect of tobacco taxes on smoking initiation, choosing the distribution of duration time to be log-logistic. This means that the density function in (1) is

$$f\left(t|s=1, x_i(t)\right) = \frac{1}{\gamma} \frac{\psi^{1/\gamma} t^{1/\gamma - 1}}{\left[1 + (\psi t)^{1/\gamma}\right]^2} \qquad (4)$$

where $\psi = \exp\left(-\beta' x_i(t)\right)$. The authors refer to γ as the "shape parameter" because it governs the shape of the density and the hazard. The hazard function of the log-logistic model is

$$\lambda\left(t|s=1, x_i(t)\right) = \frac{1}{\gamma} \frac{\psi^{1/\gamma} t^{1/\gamma - 1}}{\left[1 + (\psi t)^{1/\gamma}\right]} \qquad (5)$$

The log-logistic model belongs to the continuous time accelerated failure time (AFT) class of models. Since this study uses monthly data, and the event of interest happens years after starting to be at risk, the assumption of continuous time is a reasonable one. The AFT class of models leads to an intuitive interpretation of coefficients because they are interpreted as the proportional change in survival time for a unit change in the regressor (Jenkins, 2005). In the case of regressors measured in logarithms, the coefficient accompanying it is an elasticity. The authors seek to estimate the price elasticity of daily smoking onset η_p, which is

$$\eta_p = \frac{\partial \ln (T)}{\partial \ln (p)} = \beta_1 \qquad (6)$$

and so this study's results can be interpreted as "a 1 percent increase in prices (in real terms) leads to a β_1% increase in daily smoking onset." As mentioned previously, an increment in smoking onset suggests a delay in the age at which individuals start smoking. The delay is calculated in "months after the risk age of 10," which is the (dependent) time variable in the model. Thus the delay in months at a given age *a* and risk age *r* (both in years) after a given price change of Δ_p is

$$D(\beta_1, \Delta_{p,} a , r = \beta_1 \cdot \Delta_p \cdot 12(a - r) \qquad (7)$$

where $\Delta_p = (p_1 - p_o) / p_o$. After calculating this, it is easy to recover the delay in years. It is important to acknowledge that the delay cannot be compared to the results of studies in which the individuals are assumed to be at risk at other starting ages (Guindon, 2014).

The standard errors of the estimated duration coefficients in this approach cannot be computed from the inverse of the Fisher information matrix due to the presence of coefficients estimated using the probit model in the first step. To be able to use the usual standard errors, the probit coefficient estimates in the first stage would need to converge to their asymptotic distribution at a faster rate than the convergence of the estimated duration coefficients. However, this does not occur since the rate of convergence of both sets of coefficients is the same: the square root of the number of observations. Therefore, this study uses bootstrapping to compute them.

In order to estimate the split-population model with time-varying covariates the authors expand the survey data from the risk age of smoking onset up to the date of the survey. For those individuals that started smoking daily the cigarette price is linked to the calendar month-year in which they started to smoke daily. That is, if the person is 25 years old at the date of the survey and began smoking daily at 15 years old, the authors assign the cigarette price the person faced when he/she was 15 years old.

Assignment of months is randomized due to the fact that people report the age in years at which they began. For this example, in the case of the GATS 2015 Mexico, this corresponds to the year 2005. Since the survey asks only about the age at which individuals started smoking, the authors input the price of a month of that year at random. This cannot be done for those individuals who had not started smoking at the time of the survey. The solution to this problem adopted here is to attribute these individuals the cigarette price at the time of the survey. This procedure is constrained by the availability of cigarette price data. Prices are available from January 1990. For those observations whose age of starting smoking corresponds to a calendar month-year before January 1990 the authors do not have any price to assign.

The other explanatory variables in the duration part of the model are time-invariant. The authors assign the value of the covariate at the date of the survey for each individual *i* in the new database. Thus, covariates vary between individuals but are fixed in time.

4. Results

4.1. Daily smoking prevalence

Table 5 shows the estimation results of the prevalence price elasticity using a probit model. Although the introduction of this report argues for the importance of assessing the impact of a policy of increasing excise taxes on the prevalence of daily smokers, Table 5 also shows the estimation for the prevalence of current smokers who smoke daily and less than daily. This last estimation allows the authors to give a more complete interpretation of this study's results. Columns (1) and (2) in Table 5 show the estimation for daily smokers, while columns (3) and (4) present the estimation for daily and less than daily smokers. Columns (1) and (3) in the table show the results using the random imputation price variable and columns (2) and (4) using the average self-reported price by PSU. The rest of the explanatory variables include the wealth index; a dummy for female gender, rural residence, and being a student; and labor and age categories. All coefficients on the price variable are negative as expected.

This evidence suggests that, irrespective of the construction of the price variable and the definition of prevalence, an increase in prices is associated with a fall in smoking probability. However, the magnitude of the effect is different depending on the prevalence measure adopted. For the daily prevalence of smoking the price variable coefficients are statistically significant, while for the daily and less than daily prevalence they are not. The authors interpret this result as indicating that some smokers could go from daily to less than daily smokers rather than quitting smoking.[2]

This interpretation suggests that increasing excise taxes on the consumption of cigarettes discourages daily smoking more than inducing smokers to quit. To give a measure of this impact on smoking prevalence, the last row in the table shows the implied average prevalence price elasticity. This elasticity is around -0.40 in the first two columns of the table, suggesting that an increment of ten percent in cigarette prices is expected to reduce, on average, daily smoking prevalence by four percent. The average price prevalence elasticity for daily and less than daily smokers is smaller in absolute value, ranging from -0.005 to -0.15, and it is not statistically significant.

Results in Table 5 suggest that women have a lower daily smoking prevalence than men (the average marginal effect is about -8 percent) and that people living in rural areas have a lower daily smoking prevalence compared with people living in urban areas. The wealth index has a positive and statistically significant coefficient, suggesting those with high socioeconomic status have a larger probability to smoke daily than those with low values in the wealth index. For the random price imputation column an increase of one unit in the wealth index induces, on average, an increment of about five percent in daily smoking prevalence.

Prevalence price elasticity[3] for daily smokers, shown in Table 6, is larger, in absolute value, for women than for men. Women are more responsive to price increases than men. This difference is statistically significant at the usual levels. An increase of ten percent in price is associated with a decline in smoking probability of 4.6 percent for women and 3.5 percent for men. As mentioned in the methodology section, these are prevalence price elasticities computed using equation (3), averaging over men and women, respectively.

[2] The authors thank Guillermo Paraje for pointing this out.

[3] The results reported here are based on column (1) of Table 4. Qualitatively these results are similar for the other columns in the table, and they are reported in Appendix B of this policy report.

Table 5

Prevalence price elasticity estimation

Deciles	Daily smoking		Daily and less than daily smoking	
	(1)	(2)	(3)	(4)
Price of cigarettes (in logs)	-0.2129 (0.042)*** [-0.0322]	-0.2097 (0.047)*** [-0.0277]	-0.0031 (0.032) [-0.0008]	-0.0963 (0.066) [-0.0216]
Gender (Female=1)	-0.5511 (0.075)*** [-0.0814]	-0.5738 (0.069)*** [-0.0740]	-0.6509 (0.058)*** [-0.1643]	-0.6241 (0.054)*** [-0.1415]
Wealth index	0.3309 (0.144)** [0.0492]	0.2230 (0.137) [0.0294]	0.2577 (0.117)** [0.0642]	0.1552 (0.109) [0.0349]
Residence (Rural=1)	-0.4858 (0.057)*** [-0.0609]	-0.4527 (0.050)*** [-0.0491]	-0.4686 (0.050)*** [-0.1048]	-0.4214 (0.047)*** [-0.0842]
Student	-0.1833 (0.152) [-0.0234]	-0.2081 (0.140) [-0.0245]	-0.1336 (0.099) [-0.0320]	-0.1226 (0.102) [-0.0264]
Age category				
25-44 years old	0.0393 (0.082) [0.0056]	0.0368 (0.079) [0.0048]	-0.0487 (0.065) [-0.0130]	-0.0606 (0.062) [-0.0147]
45-64 years old	0.0734 (0.098) [0.0112]	0.0852 (0.088) [0.0115]	-0.1977 (0.069)*** [-0.0499]	-0.2111 (0.068)*** [-0.0481]
more than 64 years old	-0.2088 (0.106)* [-0.0271]	-0.1388 (0.100) [-0.0161]	-0.5129 (0.085)*** [-0.1136]	-0.4768 (0.091)*** [-0.0960]
Labor category				
Unemployed	0.2492 (0.136)* [0.0457]	0.2086 (0.125)* [0.0327]	0.1733 (0.106) [0.0493]	0.0901 (0.099) [0.0230]
Out of labor force	-0.1142 (0.087) [-0.0156]	-0.1193 (0.073) [-0.0150]	-0.2403 (0.066)*** [-0.0587]	-0.2604 (0.062)*** [-0.0570]
Intercept	-0.4352 (0.215)*	-0.4632 (0.217)**	-0.4770 (0.166)***	-0.1628 (0.288)
Prevalence price elasticity	-0.4070 (0.080)***	-0.4180 (0.138)***	-0.0046 (0.048)	-0.1547 (0.105)

*Note: Statistical significance * 10%, ** 5% and *** 1%. Standard errors in parentheses. Marginal effects in brackets.*

Table 6 also shows that increases in prices affect mostly Mexican youth and older people. A ten-percent increase in cigarette prices reduces the probability of daily smoking by 4.3 percent among youth between 15 and 24 years old, 3.9 percent for the middle aged, and 4.4 percent for those older than 65 years.

The daily prevalence price elasticity decreases, in absolute value, with wealth quartiles (see Table 6). This result indicates that an increase in prices—while it would induce a reduction in prevalence among all wealth groups—would have a greater effect on those with lower levels of socioeconomic status than those in the highest wealth quartile. All these figures are statistically significant at the usual levels, as shown in Table 6.

4.2. Smoking onset

Table 7 presents the duration estimation results. The participation column reproduces column (1) of Table 4, and the duration columns shows the estimates for the time that elapses between the risk age of daily smoking onset and the age of starting.

The duration component of the model is presented in accelerated failure time format and, therefore, the estimated coefficients can be interpreted as regression coefficients for the logarithm of time elapsed between the risk age of daily smoking onset and the age of starting. Therefore, if an explanatory variable is expressed in natural logarithm its coefficient can be interpreted as an elasticity (Forster & Jones, 2001). In

Table 6

Prevalence price elasticity for daily smokers by categories

Categories	Prevalence price elasticity	Clustered standard error	p-value
Men	-0.3487	0.0522	0.0000
Women	-0.4636	0.0581	0.0000
Age			
15-24 years old	-0.4279	0.0780	0.0000
25-44 years old	-0.3989	0.0785	0.0000
45-64 years old	-0.3875	0.0772	0.0000
over 65 years old	-0.4388	0.0821	0.0000
Wealth quartiles			
Q1 (poorest)	-0.4374	0.0843	0.0000
Q2	-0.4108	0.0770	0.0000
Q3	-0.3939	0.0754	0.0000
Q4 (richest)	-0.3742	0.0679	0.0000

Note: Prevalence price elasticities are estimated using equation (3), averaging over the relevant group. Clustered standard errors by location (basic geostatistical areas).
Source: Authors' calculations

Table 7
Split-population estimates

	Split-Population model		
	Participation	Duration	Duration
Price of cigarettes (in logs)	-0.2129 (0.042)***	1.5982 [0.211]***	
Price-Female interaction			1.8186 [0.266]***
Price-Male interaction			1.4316 [0.212]***
Gender (Female=1)	-0.5511 (0.075)***	0.3445 [0.161]**	
Wealth index	0.3309 (0.144)**		
Wealth index quartiles			
1st quartile (poorest)		0.0304 [0.195]	0.0254 [0.194]
2nd quartile		0.0085 [0.191]	0.0190 [0.189]
3rd quartile		0.0653 [0.188]	0.0639 [0.184]
Residence (Rural=1)	-0.4858 (0.057)***	0.0244 [0.122]	0.0290 [0.127]
Age categories	YES	NO	NO
Labor categories	YES	YES	YES
Education categories	NO	YES	YES
Intercept	-0.4352 (0.215)*	3.1038 [0.768]***	3.2032 [0.795]***
Shape		0.3275 [0.018]***	0.3244 [0.019]***
Prevalence price elasticity	-0.4070 (0.002)***		
Observations	11,979	5,988	5,988

*Note: Statistical significance * 10%, ** 5% and *** 1%. Bootstrapped standard errors in brackets.*

particular, a positive coefficient indicates that higher values of the explanatory variable delay the initiation in daily smoking.

Column (2) shows the estimation of the duration component using the natural logarithm of the real price, and column (3) uses the logarithm of this price multiplied by the gender dummy variable. Column (2) shows a 1.5982 estimated onset price elasticity. This positive and statistically significant elasticity suggests that increasing prices by ten percent delays the age of daily smoking initiation by almost 16 percent. The onset price elasticity implies that, at the mean risk age of 18 years, an increase of ten percent in prices is expected to delay the onset of daily smoking by one year and four months.

The positive coefficient on the gender variable indicates that women initiate daily smoking at an older age than men. The estimation of the shape parameter of the hazard rate is 0.3275, positive, and statistically less than one, implying that the smoking hazard rate first rises with time and then falls monotonically as suggested in Figure 1 above. Column (3) suggests that increasing prices by ten percent delays daily smoking initiation of women by 18.2 percent, while this figure is 14.3 percent for men, which translates to a delay of one year and nine months for women and one year and one month for men. This finding is consistent with the prevalence elasticity found in the participation component of the split-population model.

There is no effect of wealth index or residence on smoking onset. Age ranges are not included in the duration estimation because this study's data set includes individuals 35 years old and younger.

5. Discussion

Results of the prevalence and duration analyses suggest that cigarette prices in Mexico have a statistically significant effect on the probability of smoking daily and on the age of starting smoking. Increments in cigarette prices are associated with declines in smoking prevalence and could also delay daily smoking initiation. This is important because it implies that higher cigarette excise taxes, leading to higher prices, would reduce daily smoking prevalence. This reduction in daily smoking prevalence seems to be explained by smokers going from daily to less than daily smoking rather than by encouraging daily smokers to quit. This study finds that increasing cigarette prices by ten percent induces a reduction in daily smoking probability of 4.6 percent among females and 3.5 percent among men.

This study also finds a substantial effect on the age of daily smoking initiation. The authors estimate the price elasticity of smoking onset at 1.5982, indicating that a ten-percent increase in real retail price would delay the age of starting smoking daily by almost one year and four months. This finding is of particular importance for Mexican youth, because it shows that they are very sensitive to increasing cigarette prices.

In contrast to studies (Ciapponi, 2011) in many other countries, in Mexico prevalence of daily smoking is positively associated with wealth. Individuals in the richest quartile of wealth have a daily smoking prevalence of 8.5 percent, while those in the lowest wealth quartile have a daily prevalence of only six percent. As expected, the prevalence price elasticity indicates that the impact of increasing cigarette prices is larger for those in the poorest wealth quartile compared to those in the richest quartile. The authors estimate that a ten-percent increase in prices is associated with a reduction of 4.4 percent in the smoking prevalence for individuals in the poorest wealth quartile.

Daily smoking prevalence is higher for men, almost 12 percent, than for women, around four percent. On average, women begin smoking at a later age than men. This study's estimations suggest that increments in cigarette prices reduce the probability of smoking initiation for women by a greater percentage than for men. This evidence indicates that using excise taxes to induce increments in prices would be more effective to deter smoking initiation for women. Moreover, increasing excise tax is an effective public policy, as the poor, youth, and women would be among the groups to benefit most.

This study does have some limitations worth noting. First, even though the study uses a sample of young people aged 35 years old or less in 2015, there could be a recall error since individuals have to remember when they started smoking daily. Second, the study does not account for price variation across brands. Therefore, any potential substitution between cheaper or illegal brands and more expensive brands when there is a rise in cigarette prices is not captured. A potential implication of not capturing this substitution is that the estimated prevalence elasticity could be viewed as an upper-bound estimation.

6. Conclusion

In this research report the authors estimate the impact of increasing cigarette prices on daily smoking prevalence and on the age of starting smoking. A split-population model is used to specify daily smoking participation and smoking onset equations. The authors estimate these equations to obtain prevalence and onset price elasticities and to evaluate the importance of other determinants of smoking probability and factors affecting the starting age of smoking.

The empirical evidence presented suggests that prices, gender, residence, age, and wealth are important determinants of daily smoking prevalence in Mexico. Wealthier men living in urban areas have a higher smoking prevalence. Increasing prices have a larger effect on the probability of smoking for young women in the poorest wealth quartile.

The introduction of this report highlights that the addictive nature of tobacco products is at the center of many health problems, and adolescence is a key phase in which addiction might develop (DiFranza et al., 2000, 2003, 2007; Gervais et al., 2006, 2009). The evidence presented in this report suggests that increases in cigarette prices are, on average, linked to less frequent smoking and a delay in the development of the habit of daily smoking. Delaying or reducing smoking at young ages is expected to improve health outcomes over the life course. As delaying initiation makes initiation itself less likely, fewer people will initiate as they get older. Hence, a public policy of increasing excise taxes with the objective of increasing cigarette prices could be very effective to reduce or delay smoking initiation.

References

CDC. (2004). The health consequences of smoking: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, CDC. Available at http://www.cdc.gov/tobacco/data_statistics/sgr/sgr_2004/index.htm

CDC. (2002). Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Economic Costs --- United States, 1995--1999. Morbidity and Mortality Weekly Report (MMWR). Available at https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5114a2.htm

Chassin, L., Presson, C. C., Sherman, S. J., & Edwards, D. A. (1990). The natural history of cigarette smoking: predicting young-adult smoking outcomes from adolescent smoking patterns. *Health Psychology, 9*(6), 701–716. http://dx.doi.org/10.1037/0278-6133.9.6.701

Ciapponi, A. (2011). Systematic review of the link between tobacco and poverty. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/44453

DiFranza, J. R., Rigotti, N. A., McNeill, A. D., Ockene, J. K., Savageu, J. A., St Cyr, D., Coleman, M. (2000). Initial symptoms of nicotine dependence in adolescents. *Tobacco Control, 9*(3), 313–319. http://dx.doi.org/10.1136/tc.9.3.313

DiFranza, J.R., Savageau, J.A., Fletcher, K.E., O'Loughlin, J., Pbert, L., Ockene, J. K., McNeill, A. D., Hazelton, J., Friedman, K., Dussault, G., Wood, C., Wellman, R. J. (2007). Symptoms of tobacco dependence after brief intermittent use: the Development and Assessment of Nicotine Dependence in Youth study II. *Archives of Pediatrics & Adolescent Medicine, 161*(7), 704–710. http://dx.doi.org/10.1001/archpedi.161.7.704

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

Gervais, A., O'Loughlin, J., Meshefedjian, G., Bancej, C., & Tremblay, M. (2006). Milestones in the natural course of onset of cigarette use among adolescents. *Canadian Medical Association Journal, 175*(3), 255–261. http://dx.doi.org/10.1503/cmaj.051235

Gonzalez-Rozada, M., & Montamat, G. (2019). How increasing tobacco prices affects the decision to start and quit smoking: evidence from Argentina. *International Journal of Environmental Research and Public Health, 16*(19), 3622. https://doi.org/10.3390/ijerph16193622

Guindon, GE. (2014). The impact of tobacco prices on smoking onset in Vietnam: duration analysis of retrospective data. *European Journal of Health Economics*, 15, 19–39.

Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271. doi:10.2307/1913827

Jenkins, S. (2001). SPSURV: Stata module to fit split population survival ('cure') model. Statistical Software Components S418601, Boston College Department of Economics.

Jenkins, S. (2005). Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Klein, J. (2006). Adolescents and smoking: the first puff may be the worst. *Canadian Medical Association Journal, 175*(3), 262–263.

Little, R. J. A., & Rubin, D. B. (2020). Statistical analysis with missing data. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, third edition.

O'Loughlin, J., DiFranza, J. R., Tyndale, R. F., Meshefedjian, G., McMillan-Davey, E., Clarke, P. B. S., Hanley, J, Paradis, G. (2003). Nicotine-dependence symptoms are associated with smoking frequency in adolescents. *American Journal of Preventive Medicine, 25*(3), 219–225. http://dx.doi.org/10.1016/S0749-3797(03)00198-3

O'Loughlin, J., Gervais, A., Dugas, E., & Meshefedjian, G. (2009). Milestones in the process of cessation among novice adolescent smokers. *American Journal of Public Health, 99*(3), 499–504. http://dx.doi.org/10.2105/AJPH.2007.128629

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39*(3), 347-366.

Schmidt, P., & Witte, A. D. (1989). Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics, 40*(1), 141-159.

U.S. Department of Health and Human Services. (2004). Preventing Tobacco Use Among Young People: A Report of the Surgeon General. Atlanta (GA): US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. (2010) How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Vellios, N., & van Walbeek, C. (2016). Determinants of regular smoking onset in South Africa using duration analysis. *BMJ Open*, 6, e011076. doi:10.1136/bmjopen-2016-011076.

Wellman, R., Dugas, E. N., Dutczak, H., O'Loughlin, E. K., Datta, G. D., Lauzon, B., & O'Loughlin, J. (2016). Predictors of the onset of cigarette smoking. *American Journal of Preventive Medicine, 51*(5), 767–778.

World Health Organization. (2010). Economics of tobacco toolkit: economic analysis of demand using data from the Global Adult Tobacco Survey (GATS).

Appendix A

Table A1
Complete split-population model estimation

	Split-Population model		
	Participation	Duration	Duration
Price of cigarettes (in logs)	-0.2129 (0.042)***	1.5982 [0.211]***	
Price-Female interaction			1.8186 [0.266]***
Price-Male interaction			1.4316 [0.212]***
Gender (Female=1)	-0.5511 (0.075)***	0.3445 [0.161]**	
Wealth index	0.3309 (0.144)**		
Wealth index quartiles			
Q2		0.0653 [0.188]	0.0639 [0.184]
Q3		0.0085 [0.191]	0.0190 [0.189]
Q4		0.0304 [0.195]	0.0254 [0.194]
Residence (Rural=1)	-0.4858 (0.057)***	0.0244 [0.122]	0.0290 [0.127]
Student	-0.1833 (0.152)		
Age category			
25-44 years old	0.0393 (0.082)		
45-64 years old	0.0734 (0.098)		
more than 64 years old	-0.2088 (0.106)*		
	(1)	(3)	(2)

Table A1

Complete split-population model estimation (Cont.)

	Split-Population model		
	Participation	Duration	Duration
Labor category			
Unemployed	0.2492 (0.136)*	0.2317 [0.392]	0.2155 [0.386]
Out of labor force	-0.1142 (0.087)	-0.2143 [0.140]	-0.1698 [0.142]
Education categories			
Primary		0.0702 [0.751]	0.0888 [0.762]
High school		-0.0323 [0.746]	-0.0009 [0.752]
University		0.2537 [0.796]	0.2754 [0.810]
Intercept	-0.4352 (0.215)*	3.1038 [0.768]***	3.2032 [0.795]***
Shape		0.3275 [0.018]***	0.3244 [0.019]***
Prevalence price elasticity	-0.4070 (0.002)***		
Observations	11,979	5,988	5,988
	(1)	(3)	(2)

*Note: Statistical significance * 10%, ** 5% and *** 1%. Bootstrapped standard errors in brackets.*

Table A2

First step random regression imputation for cigarette prices

	Price of 20-cigarette pack (in logs)
Gender (Female=1)	0.0599 (0.0533)
Age category	
25-44 years old	-0.127 (0.0450)**
45-64 years old	-0.239 (0.0527)***
more than 64 years old	-0.448 (0.115)***
Education categories	
Primary	0.0593 (0.0801)
High school	0.150 (0.0756)*
University	0.0783 (0.0873)
Student	-0.0410 (0.158)
Homemaker	0.0163 (0.160)
Labor category	
Unemployed	-0.128 (0.0811)
Out of labor force	-0.0211 (0.140)
Wealth index quartiles	
Q2	0.0179 (0.0531)
Q3	0.0130 (0.0505)
Q4	0.0565 (0.0559)

Table A2

First step random regression imputation for prices. (Cont.)

	Price of 20-cigarette pack (in logs)
Sample strata	
12	-0.0776 (0.0762)
13	0.0412 (0.0746)
14	0.0110 (0.0792)
15	0.369 (0.103)***
21	0.0194 (0.0966)
22	-0.0482 (0.0793)
23	0.146 (0.0942)
24	-0.00597 (0.0899)
25	0.347 (0.108)**
31	0.122 (0.138)
32	-0.160 (0.0949)
33	0.0508 (0.135)
34	0.0193 (0.0891)
35	0.119 (0.118)
41	0.575 (0.152)***
42	-0.139 (0.189)
43	-0.355 (0.0822)***
Intercept	3.900 (0.122)***
Observations	1785

*Note: Statistical significance * 10%, ** 5% and *** 1%. Bootstrapped standard errors in brackets.*

Appendix B

Introduction

This supplemental material complements the policy report "Impact of the Increase in Cigarette Prices on Prevalence of Daily Smoking and Initiation in Mexico," describing in greater detail the treatment of the potential endogeneity of the self-reported price in the GATS data. The authors also analyze the estimation of the prevalence price elasticity using two alternative price variables that are not discussed in the policy report.

Are self-reported cigarette prices endogenous?

A common concern when using self-reported prices in the estimation of smoking prevalence with a cross-sectional database such as GATS is their potential endogeneity. In the policy report, the authors first check for endogeneity of the self-reported price using the Rivers-Vuong (1988) test statistic. The Rivers-Vuong procedure is similar to the Heckman (1978) test for endogeneity in the linear model but applied to the probit prevalence estimation.

The prevalence model is first written in latent variable form,

$$y_1^* = z_1 \ \delta_1 + \alpha_1 \log p + u_1 \qquad (1)$$

$$\log p = z_2 \ \delta_2 + v_2 \qquad (2)$$

$$y_1 = 1[y_1^* > 0] \qquad (3)$$

Equation (1) states that smoking prevalence, y_1^*, is a function of the natural logarithm of the self-reported price (*log p*) and other exogenous variables included in the vector z_1 such as age, gender, and socioeconomic status. Equation (2) is the reduced form of the natural logarithm of the self-reported price. As suggested in the *Economics of Tobacco Toolkit: Economic analysis of demand using data from the Global Adult Tobacco Survey (GATS)* (WHO, 2010), in equation (2):

> ...use the cigarette price data from the GATS and derive an average consumption-weighted price for cigarettes among all smokers in different locations (examples of locations are towns, counties, provinces, regions, or other geographic areas). The location-specific cigarette price will be assigned to all individuals (both smokers and non-smokers) that reside in that location.

In the case of Mexico, the authors create a price variable that assigns to smokers and non-smokers the average of the self-reported price by primary sampling unit. Finally, equation (3) is an indicator variable that reports if the authors observe a smoker ($y_1 = 1$) or a non-smoker ($y_1 = 0$).

In this prevalence model, the self-reported price is endogenous if u_1 and v_2 are correlated. In order to estimate the prevalence equation using a probit model the authors need to assume bivariate normality for (u_1, v_2) and $Var(u_1)=1$. Therefore, the authors can write[4]

$$u_1 = \frac{Cov(u_1, v_2)}{Var(v_2)} v_2 + e_1 = \theta_1 v_2 + e_1 \qquad (4)$$

Note that, from (4), Var(e_1) = 1 - $p1^2$. Replacing (4) into (1)

$$y_1^* = z_1 \delta_1 + \alpha_1 \log p + \theta_1 v_2 + e_1 \qquad (5)$$

The following probit model can now be estimated,

$$Pr(y = 1 \mid z_1, \log p, v_2) = \Phi\left(\frac{z_1 \delta_1 + \alpha_1 \log p + \theta_1 v_2}{\sqrt{1 - \rho_1^2}}\right) = \Phi(z_1 \delta_1^* + \alpha_1^* \log p + \theta_1^* v_2)$$

If the natural logarithm of the self-reported price is exogenous, then $p1^2 = 0$ and θ_1^*=0. The Rivers-Vuong procedure consists of two steps. First, a regression is run of the natural logarithm of the self-reported price on a price variable that

[4] Here "Var" and "Cov" indicate variance and covariance, and $\rho 1^2$ is the correlation coefficient between u_1 and v_2.

assigns to smokers and non-smokers the average of the self-reported price by primary sampling unit and the residuals, $\hat{\nu}_2$ are obtained. Second, a probit model of y_1 on z_1, $log\ p$, and $\hat{\nu}_2$ is estimated. Then, a check is performed for exogeneity of the self-reported price using a standard t-test on the coefficient of $\hat{\nu}_2$. The rejection of the null hypothesis of this test (H0: $\theta^*_1=0$) in favor of the alternative (H1: $\theta^*_1\neq0$) would indicate endogeneity of the self-reported price.

Table B1 shows the results of the Rivers-Vuong test statistic. The table reports the probit model estimation of the smoking prevalence (y_1) on several exogenous variables (z_1), the natural logarithm of the self-reported price ($log\ p$) and the residuals of the first stage (Residuals=$\hat{\nu}_2$). The estimated coefficient on this last variable $\hat{\theta}1^*$ is equal to 0.192 and is not statistically different from zero at any usual significance level (p-value=0.262) suggesting that the self-reported price is an exogenous variable.

Table B1
Rivers-Vuong exogeneity test

Dependent variable: y1	Estimated coefficient	Linear-ized std. err.	t-statistic	p-value
Price of cigarettes (in logs)	-0.370	0.165	-2.240	0.026
Female=1	-0.535	0.077	-6.990	0.000
Wealth index	0.325	0.144	2.250	0.025
Residence (Rural=1)	-0.476	0.058	-8.250	0.000
Student=1	-0.197	0.151	-1.300	0.193
Residuals	**0.192**	**0.171**	**1.120**	**0.262**
Age category				
25-44 years old	-0.002	0.084	-0.030	0.977
45-64 years old	0.012	0.108	0.110	0.910
more than 64 years old	-0.307	0.146	-2.110	0.036
Labor category				
Unemployed				
Out of labor force	0.213	0.140	1.530	0.128
Intercept	0.217	0.677	0.320	0.748

Source: Authors' calculations

3. Smoking prevalence under alternative cigarette prices

Even when the self-reported price in the GATS data is exogenous, in the policy report the authors take a conservative approach and estimate smoking prevalence using two different cigarette prices.

The first price variable, called "random imputation price," is as follows. First, a regression equation is estimated for the smokers in this study's sample using as a dependent variable the self-reported price paid for the last purchase and as explanatory variables: gender (female=1), age, labor and education categories, wealth quartiles, binary variables for students and homemakers, and survey strata fixed effects. Then the authors input prices for non-smokers using the predicted price from this regression plus a random draw from a normal distribution with mean and standard deviation equal to the mean and standard deviation of the residuals.

A second price variable is constructed called "self-reported price by PSU," estimating a regression for the smokers in this study's sample with the self-reported price as the dependent variable and using as explanatory variables dummy variables for each PSU in this study's sample. Then, the predicted price from this regression, which is the average self-reported price by PSU, is assigned to both smokers and non-smokers.

Table B2 reproduces the first two columns of the smoking prevalence estimation presented in the policy report.

Column (1) in the table shows the results using the random imputation price variable and column (2) using the average self-reported price by PSU.

In the policy report the authors analyze the prevalence elasticity for the self-reported price while in this appendix the authors first show the results for the average self-reported price by PSU, which is the price suggested in the *Economics of Tobacco Toolkit: Economic analysis of demand using data from the Global Adult Tobacco Survey (GATS)* (WHO, 2010).

Column (2) of Table B2 shows that the prevalence price elasticity is -0.418, and it is statistically significant at the usual levels. This implies that a ten-percent increase in cigarette prices would induce a decline in smoking prevalence of around 4.2 percent. Prevalence price elasticity is larger, in absolute value, for women than for men. Women are more responsive to price increases than men. An increase of ten percent in price induces a decline in smoking probability of 4.7 percent for women and 3.6 percent for men (see Table B3).

Increases in prices affect mostly Mexican youth and older people. A ten-percent increase in cigarette prices reduces the probability of smoking by 4.4 percent among youth between 15 and 24 years old, around 4.0 percent for middle age ranges, and 4.5 percent for those more than 65 years old. These differences are statistically significant at the usual levels.

The prevalence price elasticity decreases with the wealth quartiles. This result indicates that an increase in prices—while it would induce a reduction in prevalence among all wealth groups—it would have a larger impact on those individuals in the lower wealth quartiles. In particular, an increase of ten percent in cigarette prices would reduce smoking prevalence by 4.4 percent for those in the lowest wealth quartile compared to only a 3.9 percent decrease for those in the highest quartile.

Overall, the results are qualitatively similar for both price specifications in terms of the prevalence price elasticity estimation. This is important because it adds some robustness to this study's analysis. That is, independently of how prices are constructed, the evidence presented here and in the policy report suggests that a public policy of increasing excise taxes on the consumption of cigarettes with the objective of increasing prices could be very effective to reduce daily smoking prevalence. In particular, the increment in cigarette prices would affect the smoking prevalence of women, youth, and the poorest individuals.

Table B3

Prevalence price elasticity by categories using average self-reported price by PSU

Categories	Prevalence price elasticity	Clustered standard error	p-value
Male	-0.3597	0.0462	0.0000
Female	-0.4736	0.0495	0.0000
Age			
15-24 years old	-0.4374	0.0749	0.0000
25-44 years old	-0.4098	0.0717	0.0000
45-64 years old	-0.4029	0.0706	0.0000
more than 65 years old	-0.4535	0.0736	0.0000
Wealth quartiles			
Q1 (poorest)	-0.4449	0.0785	0.0000
Q2	-0.4217	0.0728	0.0000
Q3	-0.4093	0.0699	0.0000
Q4 (richest)	-0.3929	0.0645	0.0000

Note: Prevalence price elasticities are estimated using equation (3), averaging over the relevant group. Clustered standard errors by location (basic geostatistical areas).
Source: Authors' calculations

Table B2

Prevalence price elasticity estimation

	Daily smoking	
	(1)	(2)
Price of cigarettes (in logs)	-0.2129 (0.042)*** [-0.0322]	-0.2097 (0.047)*** [-0.0277]
Gender (Female=1)	-0.5511 (0.075)*** [-0.0814]	-0.5738 (0.069)*** [-0.0740]
Wealth index	0.3309 (0.144)** [0.0492]	0.2230 (0.137) [0.0294]
Residence (Rural=1)	-0.4858 (0.057)*** [-0.0609]	-0.4527 (0.050)*** [-0.0491]
Student	-0.1833 (0.152) [-0.0234]	-0.2081 (0.140) [-0.0245]
Age category		
25-44 years old	0.0393 (0.082) [0.0056]	0.0368 (0.079) [0.0048]
45-64 years old	0.0734 (0.098) [0.0112]	0.0852 (0.088) [0.0115]
more than 64 years old	-0.2088 (0.106)* [-0.0271]	-0.1388 (0.100) [-0.0161]
Labor category		
Unemployed	0.2492 (0.136)* [0.0457]	0.2086 (0.125)* [0.0327]
Out of labor force	-0.1142 (0.087) [-0.0156]	-0.1193 (0.073) [-0.0150]
Intercept	-0.4352 (0.215)*	-0.4632 (0.217)**
Prevalence price elasticity	-0.4070 (0.080)***	-0.4180 (0.138)***

*Note: Statistical significance * 10%, ** 5% and *** 1%. Standard errors in parentheses. Marginal effects in brackets.*

www.tobacconomics.org

INSTITUTE FOR
HEALTH RESEARCH
AND POLICY
UIC