

Prices, Cigarette Consumption and Smoking Intensity

Jérôme Adda and Francesca Cornaglia*

University College London and Institute for Fiscal Studies.

June 2004.

Abstract

This paper uses novel data sets, which report both the quantity of cigarettes smoked and a measure of nicotine intake obtained from blood or saliva samples for a large and representative population over time for the UK and the US. We show that self-reported quantities, routinely used in the smoking literature, explain only a small fraction of the variance in smoking as measured by nicotine intake. We argue that a large part of the remaining variance reflects heterogeneity in smoking intensity, defined as the amount of nicotine extracted per cigarette. We show that smoking intensity varies with observable characteristics such as gender, age, socio-economic status or race.

This paper makes therefore two main contributions. First, we show that although higher prices decrease the number of cigarette smoked, they have no effect on nicotine consumption, as smokers compensate by smoking each cigarette more intensively. Second, our results imply that the economic literature on smoking has missed an important adjustment behavior, and that previous studies on smoking probably suffer from severe estimation biases.

* j.adda@ucl.ac.uk and f.cornaglia@ucl.ac.uk. We are grateful to Christian Dustmann, Ken Chay, Martin Jarvis, Steve Machin, Adriana Lleras-Muney and Costas Meghir for helpful discussions and comments and to William Evans for supplying us with data on US tobacco prices and taxes.

I. Introduction

The economic literature on smoking has considered cigarettes as a homogenous consumption product. In this setting, individuals decide on how many cigarettes to smoke, conditional on prices, income and possibly past and future consumption. A number of empirical studies, surveyed in Chaloupka and Warner (2000), have analyzed the effect of prices on quantities smoked. They have consistently found a significant price effect on quantities, which differs in the short run and the long run, due to addiction (see for instance Becker et al. (1994)). Evans et al (1999) also show that smoking bans have an effect on smoking. These studies have been influential for the design of public health policies aimed at reducing smoking, by advocating higher prices or more stringent bans on smoking in public spaces.

The reported quantity of cigarettes, either purchased or consumed, is only one particular measure of the actual consumption of tobacco. Cigarettes vary in their tar or nicotine yields and in size. Another dimension is behavioral. There are many ways to consume a cigarette, by smoking it to its end or discarding it half way. The degree of inhalation may also vary from one smoker to the other. What is important for economic studies is in many cases a global measure of smoking, not the number of cigarettes that have been purchased or even smoked. For instance, an important variable for the addiction literature is the amount of nicotine consumed per day. This is also true for health policies. The health consequences of a cigarette depend on the quantity of tar and carcinogens, which increase with the intensity of inhalation and the degree to which a cigarette is consumed. Smoking more intensively a cigarette up to the filter leads the smoker to inhale more dangerous chemicals, and has been shown to cause cancer deeper into the lung (Thun et al (1997)).

With only data on quantities available, the economic literature has been silent about smoking intensity, or has assumed that quantities are an acceptable proxy for overall

consumption.¹ This is not true if heterogeneity in smoking intensities is important, and especially if it varies systematically across population groups and time. An important exception is the work by Evans and Farrelly (1998) who analyze the effect of prices on the type of cigarette smoked. They find that smokers compensate tax increases by buying longer or heavier cigarettes. Harris (1980) presents a theoretical model predicting this compensating behavior, while Grossman et al (1993) speculate that, a priori, the change in intensity will be of a small magnitude compared with the effect of prices on quitting.

In this paper, we document the heterogeneity in smoking intensity directly by exploiting a new source of data, which reports not only the number of cigarettes smoked and their type but also the cotinine concentration (a marker of nicotine intake) in blood or saliva, for a large number of smokers, and over time, for two countries, the UK and the US.

Analyzing the variations in cotinine concentrations in body fluids allows us to take into account compensating behaviors in its entirety. It is influenced not only by the size or the type of the cigarette smoked but also by the depth of inhalation, the number of puffs per cigarette and the possible blocking of ventilation holes in the filter by the smoker.

We show that there is a considerable heterogeneity in smoking intensity. The number of cigarettes smoked explains only about 26% of the variance in cotinine concentrations of smokers in both countries. Moreover, the type of cigarettes smoked such as the nicotine content, the size of the cigarette or the presence of a filter explains little of this heterogeneity. This implies that there is considerable variance across individuals in how a cigarette is smoked.

There can be several explanations, not mutually exclusive, for this heterogeneity in intensity. First, self-reported quantities can be prone to measurement error. We define the intensity of smoking as the concentration of cotinine per cigarette smoked. Measurement errors in the number of cigarettes smoked could certainly contribute to the apparent heterogeneity in intensity. We document one aspect of reporting bias by looking at self-declared non smokers who have a positive (and significant) level of cotinine. We show

¹ The epidemiological literature has paid much more attention to compensatory smoking behavior. The literature includes for instance Kozlowski et al (1980), Wagenknecht et al (1990), McCarthy et al (1992), Caraballo et al (1998) or Jarvis et al (2001).

that this reporting bias is rather small and cannot explain all the heterogeneity in smoking, conditional on the number of cigarettes.

Another and probably more important source of heterogeneity is *behavioral*. Smokers choose not only the number of cigarettes they smoke, but also their effort to extract nicotine. To achieve a particular level of nicotine in the body, there is a trade-off between more cigarettes, which are costly, and more effort. Poor individuals may absorb more nicotine by smoking more intensively and buying fewer cigarettes. More affluent individuals may decide to purchase more cigarettes and smoke less of them to achieve the same level of nicotine intake.

Our findings are consistent with this explanation. Although we find a negative price elasticity for the quantities smoked, we show that prices have no effect on cotinine concentration in smokers. The average cotinine concentration in smokers has barely changed during the period of analysis, despite large price increases. We also find that smoking intensity is determined by income and wealth measures for the UK and the US. For instance, for the UK, individuals in poorer socio-economic groups can extract up to 40% more nicotine per cigarette.

We find that prices of cigarettes affect differentially the smoking intensity across socio-economic groups. For instance, for the UK, for young women in poorer households, an increase in prices leads to a decrease in quantities smoked but leads to an increase in smoking intensity, so that the net effect on cotinine concentrations is zero. For the US, there is strong evidence that African-Americans and especially women increase their smoking intensity as prices increase.

Our results question the measurement of the tobacco epidemic, when using quantities smoked. This is especially true when comparing different socio-economic groups. For instance, for the UK, we show that the intensity of smoking has changed differentially for young men and women. In the beginning of our period of analysis, the early nineties, young women were smoking fewer quantities and with a lower intensity, achieving significantly lower levels of cotinine than men. However, about ten years later, young women have increased their intensity of smoking, even though smoked quantities stagnated or decreased, achieving for some cohorts a higher cotinine level than men.

For the US, although African-Americans smoke a smaller number of cigarettes than whites, their intensity of smoking is much larger resulting in higher levels of cotinine than whites. The gap has also been increasing over the nineties. Differences in smoking intensity help to explain differences in health and mortality across ethnic groups, for instance, the abnormally high lung cancer rates in African-Americans and the apparent good health of Hispanics in the US. Our findings show that the reality of smoking is much more complex than addressed in the economic literature and that using only the self-reported number of cigarettes smoked can be extremely misleading.

Our results have several important implications for the interpretation of previous results in the literature. Given that the self-reported numbers of cigarettes smoked is a poor proxy for actual smoking, it implies that studies using this variable are prone to attenuation bias. They are also prone to selection bias due to the differential intensity across groups and in response to price changes.

A second and important problem occurs in the empirical literature on (rational) addiction, which regress the quantity of cigarettes on past and possibly future consumption (see for instance Becker et al. (1994) or Chaloupka (1991)). The empirical literature usually instruments the right-hand side variables with prices. However, when using quantities instead of a broader definition of smoking, the intensity of smoking is subsumed in the error term. Given that the intensity is correlated with prices, it implies that the instrument is not valid. As a result, the estimates in this literature are subject to bias. Using our data sets, we evaluate the importance of this bias and show that it can be substantial, potentially invalidating the results in the literature.

Third, given that the intensity of smoking is sensitive to prices and income, and that a higher intensity is detrimental to health, price increases, bans on smoking in the workplace or recession, have- a priori- an ambiguous effect on health. On one hand, they may lead smokers to quit, or to reduce the number of cigarettes smoked. But by increasing the intensity of smoking, the smokers may inhale more tar, carcinogens or carbon monoxide than before. This is consistent with findings in the medical literature

that the prevalence of deep lung cancers (adenocarcinoma) has been steadily increasing in the US since the fifties (Thun et al (1997)).

The remainder of the paper is structured as follows. Section 2, contains a description of our data sets. In Section 3, we document the extent of smoking intensity and link it to a number of observed characteristics. Section 4 investigates the relationship between smoking, smoking intensity and prices. Section 5 evaluates the bias in conventional models which use the number of cigarettes smoked instead of a more accurate measure of smoking, such as the cotinine concentration. Finally, Section 6 concludes and discusses the implications of our results.

II. The Data Set

A. Description

We use data for two countries, the UK and the US. For the UK, the data set is the Health Survey for England 1993-2001 (HSE) and for the US, the National Health and Nutrition Examination Survey (NHANES III and NHANES 1999-2000).

Both surveys are large repeated cross-sectional surveys which report individual and family characteristics, smoking habits, including the self-reported number of cigarettes smoked per day.

In addition, in both surveys, the cotinine concentration was measured either in blood or saliva samples. Cotinine is a metabolite of nicotine. When smoking, nicotine is extracted from the cigarettes and passes into the blood. However, nicotine is rapidly transformed into cotinine, and is therefore difficult to measure. Cotinine has a half-life in the body of about 20 hours and is therefore one of several biological markers that are indicators of regular smoking.² It can be measured in, among other things, saliva or serum. Its concentration is an important marker of smoking for regular smokers and has been extensively used in epidemiology studies on smoking.

² The elimination of cotinine is slow enough to be able to compare measurement done in the morning or afternoon. As NHANES 1999-2000 provides the time of the examination, we present more evidence on this point in section 3.1.

The most health threatening substances in cigarettes are not nicotine but tar and carbon monoxide, causing, for instance, cancer and asphyxiation. However, based on our data sets, which report for some years the nicotine, tar and carbon monoxide yield of each cigarette, the correlations between nicotine and both tar and carbon monoxide are high, 0.97 and 0.89 respectively for the UK and 0.96 and 0.85 for the US. Cotinine is, therefore, a good indicator of health hazards due to smoking.

The HSE is a repeated cross-section covering around 16000 individuals each year, from 1993 to 2001. The data set also reports a number of individual and household characteristics such as age, sex, education levels, occupation, marital status, household income (for the last two years only), and household size. There is no information on wealth directly but the data set reports the ownership of a house, its size and the possession of durable goods such as cars. We use data for the years 1993, 1994, 1996, 1998 and 2001, for which the surveys were representative of the whole population. The other years put more emphasis on particular population groups, such as children or elderly. In total, we observe 43886 individuals both smokers and non smokers with a valid measure of cigarette consumption and cotinine concentration.

NHANES for the US is a nationwide representative sample of the US civilian population. It provides information from 1988 to 1994 and 1999-2000, for approximately 18000 individuals, aged seventeen and above. The data set reports information on the age, sex, race, education and occupation of the individual, as well as information at the household level such as family composition, income, geographical location or welfare eligibility. In 1999-2000, only a limited set of demographic variables is available.

Cotinine is measured in NHANES and in the first two years in the HSE in blood samples.³ After 1997, cotinine was measured in saliva samples in the HSE, but cotinine levels measured in saliva are higher than those measured in blood serum. We converted all the cotinine levels into saliva cotinine using the formula $s = \alpha b$, where b is the

³ In both surveys, response rates are high, although attrition due to the invasive nature of the blood test is present. However, this attrition is not related to the possible use of cotinine as a marker for smoking. The blood sample was used to measure a number of other metabolites apart from cotinine and it is not clear that individuals understand the potential use of cotinine. Moreover, the observable characteristics of the smokers without a valid measure for cotinine are similar to those who gave a sample of blood or saliva (see Table A in the appendix).

cotinine concentration obtained from a sample of serum, and s is the correspondent saliva value⁴. The coefficient α is equal to 1.23 for women and 1.26 for men. All measurements in both countries are expressed as the equivalent saliva concentration in ng/ml. For interpretation of results, note that on average a smoker achieves a concentration of about 24 ng/ml per cigarette, both for the UK and the US.

Both data sets also report in the last years (1999-2000 for the NHANES and 1998-2001 for the HSE) further details on the cigarettes smoked, the nicotine, the tar and the carbon monoxide concentration, as well as the brand name.

For the two countries, we merge information on cigarette prices. For the UK, we use the relative price of tobacco, adjusted for inflation, obtained from the Institute for Fiscal Studies. In the UK, there are no regional differences in taxation of tobacco products, so all the variation comes from time variation. During the period 1993-2001, the price rose by 52%.

For the US, we use information from the Tax Burden on Tobacco, which reports prices and taxes by state and year. We deflated all prices using the consumer price indices. Most of the variation is cross-sectional, where prices can vary by about 30% and taxes by 80%. There is however some differential variations over time across states, that we exploit to identify the effect of prices.

[Table 1]

Table 1 provides a summary statistic of our data sets. Smokers are categorized according to their self-reported smoking status. The average number of cigarettes smoked over the observation window is about 15 per day for the UK and 19 for the US. The average level of cotinine concentration (expressed in ng/ml in saliva for both countries) is somewhat higher in the UK than in the US, about 306 ng/ml and 288 ng/ml. Cotinine levels in non smokers are comparable across countries. In both countries, smokers come proportionally more from low or medium education groups, as well as from unskilled and manual occupations. Men are more likely to smoke as well as young individuals. There is no notable difference in smoking rates across African-Americans and whites for the US or between whites and ethnic minorities for the UK.

⁴ The formula was derived analyzing the data obtained from an experiment mounted on informants taking part in the 1998 Health Survey. This experiment involved taking saliva samples from about 600 of those already giving a blood sample, so that both could be compared for the same set of informants (M. Jarvis et al., 2003).

III. Individual Determinants of Smoking Intensity

A. Heterogeneity in Smoking

Figures 1a and 1b plot the average quantity of cotinine concentration (ng/ml) in saliva as a function of the number of cigarettes consumed, for the UK and the US. Cotinine levels are increasing with the number of cigarettes smoked. The relationship is linear up to 20 cigarettes a day and then levels off. However, the vast majority of smokers in both countries (about 80%) are smoking 20 cigarettes or less. Heavy smokers at 20 cigarettes or more may smoke more frequently but less intensively, to keep the concentration of cotinine at a high level while avoiding an overdose, which induces unpleasant effects such as nausea, headaches and dizziness (see for instance the British National Formulary (2004)).

The graph also plots the 10 and 90% quantiles of cotinine concentration within each level of cigarette consumption. The most striking feature in the graph is the large heterogeneity in cotinine levels. Some smokers reporting 5 to 9 cigarettes per day can absorb more nicotine than most heavy smokers. Conversely, heavy smokers at almost two packs a day can absorb no more nicotine than light smokers under half a pack a day.

[Figures 1a & 1b]

A number of different factors could contribute to explain this heterogeneity. Table 2 provides a simple analysis of variance for the intake of cotinine for smokers by providing the (adjusted) R^2 of a regression of cotinine levels on explanatory variables. The table reports the results for the UK and for the US. Given that the two NHANES data sets contain sometimes different variables we report the R^2 for both parts of the data set.⁵

The number of cigarettes smoked per day (coded with eight dummies) explains only 26% of the variance of cotinine intakes in both countries (slightly less in later periods for the US). This means that the reported number of cigarettes smoked per day is a poor proxy for actual smoking. Yet, the entire economic literature on smoking has used this measure to evaluate the extent of smoking, the effect of prices on smoking, and to evaluate the presence of addiction.

⁵ In particular, NHANES III (1988-94) does not contain information on the nicotine content nor on the brand.

The heterogeneity does not stem from year to year variations as indicated in the second column, which controls for year fixed effects.

Next, we include additional information on the type of cigarette smoked to see to what extent they explain the difference in cotinine concentration in the population.

In column 3, we control for the nicotine content of the cigarette. The nicotine content can vary substantially from one brand to another. In our sample, the lightest cigarette had a yield of 0.05 milligrams of nicotine, whereas the heaviest had a yield of up to 1.07 milligrams (the average content was 0.8 milligrams for the UK and 0.95 milligrams for the US). The information on the brand is only available in 1998 and 2001 for the UK and in 1999-2000 for the US, when individuals were asked to report the brand of cigarettes they smoked.⁶ Including this additional information, the explained variance in cotinine intakes increases from 26% to 30% for the UK and from 22% to 24% for the US, a rather modest improvement. This fact also points to behavioral adjustments in cigarette smoking. It means that smokers who buy “light” cigarettes are probably compensating the low level of nicotine by smoking more intensively (Jarvis et al. (2001)).

Further, we control for the brand and the type of cigarettes smoked directly, rather through their nicotine content only, by introducing brand dummies. The results are in column 4. Interestingly, the type of the cigarette explains part of the variance in nicotine intake. The explained variance rises by 6% for the UK and 9% for the US. Given that the nicotine content was already controlled for in the previous column, the brand indicators may provide information on the presence of filters, but could also pick up heterogeneity in smoking intensity correlated with the brand. It may be that some brands target more intensive smokers, such as men, by promoting a positive image of high intensity smoking. Column 5 controls for a full sets of characteristics including the brand of the cigarette, the nicotine content, the size of the cigarette, the presence of a filter and whether the cigarette contains menthol. With these sets of covariates, we are able to explain 36% of the variance in cotinine concentration.

Finally, NHANES III for the US reports the time of the lab examination, as either morning or afternoon and the day in the week. Given that cotinine is constantly eliminated by the body, one would assume that those measured in the morning would

⁶ For the UK, each brand was then matched with its nicotine content obtained from an independent laboratory. We are grateful to Martin Jarvis for supplying us with this additional data.

have a lower cotinine concentration as part of it may have been eliminated during the night. Moreover, if smoking fluctuates during the week, the variation in the day of the exam could help to explain part of the heterogeneity in cotinine levels. We find that those measured in the morning have indeed a lower cotinine concentration. On average, the concentration is about 12ng/ml lower, controlling also for the number of cigarettes smoked. Although this number is statistically different from zero, it is rather modest, given that the 10-90% percentiles bands are about 400 ng/ml wide. In particular, the time of the lab examination or the day of the exam do not help to improve the explained variance at all.

Given that the number of observations varies from one column to the other, the R^2 measures may be misleading. For the UK, we also did the analysis on the subset of smokers in column 4 for whom we have information on individual characteristics and the nicotine content of the cigarette. The explained variance remained essentially the same for columns 1 to 3 (results not shown).

[Table 2]

The results in Table 2 indicate that a large part of the variance in smoking as measured by cotinine intakes remains unexplained, despite controlling for an extensive set of variables, characterizing the product itself. It appears that it is not so much the characteristics of the product that counts but what the smoker does with it.

B. Misreporting

Part of the heterogeneity in smoking intensity could stem from reporting bias when the individual reports his or her daily consumption of cigarettes. It could be due to difficulties in evaluating the average number of cigarettes, or to the fact that the actual consumption fluctuates and the stated consumption represents a long run average. A third reason could be due to stigma, which leads smokers to under-report their consumption.

It is obviously difficult to evaluate the extent of reporting bias, and even more so with only data on self-reported consumption, which is the norm in the economic literature.

We explore here the importance of stigma by looking at individuals who declare themselves as non smokers but have none the less a significant amount of cotinine in their fluid sample.

In the US, 3.3% of non smokers have a cotinine level in excess of 15 ng/ml, a common threshold to assess the smoking status. However, occasional or very light smokers could categorize them-selves as non smokers. We therefore use also two higher thresholds, equivalent to approximately five and ten cigarettes per day (cotinine concentration of 100ng/ml and 200ng/ml). We find that the misclassified individuals represent respectively about 1% and 0.5% of the population of non smokers for the US. For the UK, we find approximately the same numbers, 3.5% for the lower threshold and 1.8% and 0.9% for the higher ones. In calculating these numbers, we excluded consumers of other types of tobacco such as cigar or pipe smokers and, for the US, consumers of chewing tobacco and snuff. This later behavior is not documented in the Health Survey for England. Hence the numbers for England are probably spuriously larger.

Conversely, the proportion of declared smokers who have a cotinine concentration below the 15 ng/ml threshold is tiny when we consider smokers above half a pack of cigarettes, in both countries.

[Table 3]

We now look at the determinants of misreporting. Table 3 provides the marginal effects of different characteristics on the probability of misreporting (as evaluated with a threshold of cotinine concentration of 15ng/ml). For the US, the biggest effect is race, with African-Americans having a 4.4% higher probability of misreporting their smoking status. Among African-Americans, the smoking prevalence is 30% using self-reported consumption and 34% using actual cotinine intakes. Whites have also a higher probability of misreporting, at around 2.5% (the omitted ethnic group are Hispanics). For the UK, misreporting is more common among men, young individuals (of age between twenty and twenty five), and individuals with low education or in manual occupation and in ethnic minorities.

This misreporting, due perhaps to stigma, could explain only part of the heterogeneity. Some individuals may under-report the quantities of cigarettes smoked, explaining why light smokers have a high cotinine level. But it cannot explain why a substantial fraction of heavy smokers absorb so little nicotine.

A different source of misreporting could be due to heaping, whereby smokers round up their quantities smoked to focal numbers, such as half a pack or a pack of cigarettes. For

instance, this would lead individuals smoking both 18 and 22 cigarettes to report a daily consumption of 20, but the former would tend to have smaller amount of cotinine than the latter. This could contribute to the overall heterogeneity in cotinine levels. The existence of heaping would imply that the heterogeneity at the focal points is larger than for other reported quantities. Thus, a test of the importance of this phenomenon would be to compare the heterogeneity in cotinine levels for those who report 5, 10, 15, 20... cigarettes per day with those who reports different quantities, controlling for the number of cigarettes smoked. The results both for the UK and the US show that the heterogeneity at the focal points is not larger than at other points, indicating that heaping is not a likely source for the observed heterogeneity in cotinine levels.

Coming back to Figures 1a and 1b, the 10 and 90% bands around the average concentration are so large, roughly equivalent to plus or minus ten cigarettes, that it is doubtful that the heterogeneity stems only from short run fluctuations or approximations in the self-reported number of cigarettes. Clearly misreporting cannot be the only explanation to this heterogeneity.

Some studies in epidemiology have tried to better measure the number of cigarettes smoked, by following smaller groups of smokers, counting the number of cigarettes smoked and evaluating their cotinine concentration. It appears that the heterogeneity in smoking is still present (see Clark et al. (1996)).

C. Determinants of Smoking Intensity

We next explore the determinants of smoking and smoking intensity in relation to individual characteristics. Tables 4a and 4b present the determinants of smoking using either the daily quantities of cigarettes or the cotinine concentration as a measure of smoking. The tables also present the determinants of smoking intensity, defined as the cotinine concentration per cigarette smoked. For ease of interpretation, all the dependant variables are expressed in logs. Table 4a, column 1 displays the regression coefficients of the logarithm of the number of cigarettes smoked for the UK. Men smoke on average

15% more cigarettes than women. Education is an important determinant with low educated individuals smoking around 30% more than higher educated ones. Total household income appears to be marginally significant, with individuals at the top of the distribution smoking about 5% less. Smoking is also determined by occupation as individuals in manual unskilled occupation are smoking about 10% more. Next, we include several markers of household wealth. An individual owning a house smoke about 10% less, but the size of the house appears to have no effect on smoking, at least measured by quantities. White individuals smoke about 30% more in the UK. Household size is positively associated with smoking. Finally, there is a strong age effect, which can capture cohort effects as well as age effects as the regression is done on cross sections. We decompose age and cohort effects in section 4.2.

Table 4a column 2 presents the effect of the same variables on smoking as measured by *cotinine concentration*. Overall, we find qualitatively similar effects, but there are some important and noteworthy differences. First, the education differences in smoking is deepened, with low educated individuals smoking 37% more than high educated ones. More importantly, there is much more variation in smoking between income groups and wealth groups. Low income individuals smoke 19% more than better off individuals. Wealth, as measured by the size of the house is also clearly significant, with individuals in one bedroom accommodation smoking 21% more than those in larger houses. The difference in behavior across ethnic groups is negligible when using cotinine as a measure, despite the fact that white individuals appear to consume more cigarettes.

The differences in the results between columns 1 and 2 show that the self-reported number of cigarettes consumed does not fully capture all the aspects of smoking behavior, especially when one considers the economic environment. Given that quantities and concentration differs across groups, it means that the amount of cotinine per cigarette reported to be consumed, namely the intensity of smoking, will vary across groups. This is shown in Table 4a, column 3. The intensity of smoking is higher for low income people who extract 14% more cotinine per cigarette. Individuals with less wealth, as measured by the size of the house, also smoke more intensively. Finally, there is a marked difference in smoking intensity between whites and ethnic minorities, the former extracting 23% less cotinine.

In total, individuals in the bottom of the income distribution, with a low education, who are renting a small house, can extract up to 40% more nicotine, a quite substantial difference.

[Tables 4a & 4b]

Table 4b displays the results for the US. As for the UK, the results using the number of cigarettes smoked differ from those using the cotinine concentration. The effect of gender, education, income and the size of the house are stronger when using a more direct measure of smoking. However, the most notable difference for the US is the effect of race. When looking at the self-reported number of cigarettes, African-Americans appear to smoke less than whites (approximately 40% less), but similar quantities to Hispanics. However, when using the cotinine concentration, African-Americans absorb by far the highest amount of nicotine. Unsurprisingly, these differences are picked up in the intensity equation, where the biggest effect is race and indicators of wealth (as captured by the size of the house).

The race difference in cotinine has been pointed out in the epidemiological literature, for instance in McCarthy et al. (1992) or Caraballo et al. (1998) (and the references within). There is a debate whether the race difference is due to differences in smoking behavior or to differences in nicotine metabolism. McCarthy et al. (1992) argue that the difference is entirely behavioral. Clark et al. (1996) found that African-Americans smoke more of each cigarette. However, Benowitz et al. (1999) find some differences in nicotine metabolism between African-Americans and whites, although they do not attribute all the difference in cotinine levels to this fact alone.

For the US, the difference in the intensity of smoking between whites and African-Americans is of the same order for men and women. Interestingly, for the UK, although black men have a higher intensity of smoking than whites, there is not much race difference in smoking intensity in women. This also suggests that part of the race difference is behavioral.

A final argument in favor of the behavioral origin in differences in cotinine concentration across race is its effect on health and mortality. For instance, African-American men have

by far the highest incidence of lung cancer at about 120 per 100000, while the rate is at 79.4 for whites and only 46.1 for Hispanics. (National Cancer Institute, 2004). This occurs despite the fact that whites smoke more cigarettes than both African-Americans and Hispanics. It is also interesting to note the differences in both lung cancer rates and smoking intensity in African-Americans and Hispanics, two population groups which are relatively economically deprived. From data on quantities smoked alone, the apparent good health of Hispanic appears somehow as a puzzle, which can be partly explained by differences in intensity.

The lung cancer rates for the UK vary also with ethnicity, reflecting in part the differences in the smoking prevalence, the quantities smoked and the effort of smoking per cigarette (Davey Smith et al (2000)). The difference is that for the UK, whites have higher standardized mortality ratios for lung cancer than ethnic minorities (although not always significantly higher), especially women.

D. Smoking Intensity in Young Individuals

The results in the previous sections were obtained on individuals of age 20 or more. Given the problems of misreporting for younger individuals, we excluded children and adolescents from the analysis so far. We focus in this subsection on younger individuals. We are only able to use the data for the UK, as in NHANES, young individuals were not asked about their smoking behavior.

In the HSE, individuals eighteen or under were asked to fill in a separate questionnaire, in private, which recorded the number of cigarettes smoked. The minimum age required to fill up the questionnaire varied from one year to another, from 12 to 16. Our data set also reports cotinine measures in children.

We pool all years together given the relatively small sample of children who smoke. We report the average smoking intensity (cotinine concentration per cigarettes smoked) in Figure 2, by age and sex. We use information on children aged 14 and above for which the sample is large enough.

[Figure 2]

The graph shows that the intensity of smoking is, at young ages, low for both boys and girls, about half of the intensity of adults. This is probably due to the fact that at initiation, children do not inhale as much and as deeply as adults. However, for boys aged

sixteen onwards, the intensity jumps up to adult levels and appears to be stable thereon. The behavior of girls is different and shows a gradual increase in the intensity of smoking, catching up with the boys at around age 25. Given the small sample size (and possible measurement error in the number of cigarettes), we do not attempt to look at the heterogeneity in smoking as we have done for adults.

IV. The Effect of Prices on Quantities and Intensity

A. Price Elasticities

The economic literature has mostly focused on prices as a mean to curb the consumption of cigarettes. In this section, we evaluate the effect of prices on consumption using as a measure of smoking, either the quantities of cigarette smoked, or the cotinine levels. The objective is twofold. First, we explore how the price effects usually estimated in the literature compare to the one obtained using a direct measure of smoking intake instead, the cotinine levels. Second, we explore how prices affect changes in behavior related to the intensity of smoking.

For both countries, we report the price elasticity of the number of cigarettes, cotinine levels and smoking intensity. These are estimated using an ordinary least square procedure on repeated cross-sections, controlling for a number of individual characteristics. In addition, we include a time trend for the UK to capture other unobserved effects, which may vary with time. For the US, as we have both time and geographical variation in prices, we include both state and year dummies.

The OLS results are only valid if there is no selection out of smoking.⁸ The constant rise in prices during the nineties for the UK and the US may have changed the composition of the pool of smokers. If low intensity smokers are more likely to quit following a price increase, the OLS estimates would spuriously find a positive correlation between prices and smoking intensity. For the UK, over the period 1993-2001, about 13% of those who

⁸ The selection into smoking is minor when looking at an adult population, as the vast majority of smokers start before their twenties. Moreover, we condition the analysis on individuals who were smoking at the start of the period of analysis, using information on the duration of the smoking spell. This lead to exclude a small fraction of the youngest smokers in our sample.

smoked in 1993 quit. A similar proportion is observed for the US between 1988 and 1994.

Controlling for selection is not an easy task, especially in the present case. To fully address the selection issue, we would need to find a variable that affects smoking participation but not smoking intensity. There is no obvious candidate for such a variable. Prices or individual characteristics could affect both the intensive and the extensive margin. We address this problem in two ways. First, we present results for a Heckman selection model (Heckman (1979)). In the absence of any clear exclusion restrictions, the model is identified from the functional forms, which requires a rather strong set of assumptions. Second, we present worst case bounds on the effect of prices, following Manski (1994), a procedure which, on the contrary, imposes no restriction on the model. We refer the reader to the appendix for a short description of the estimation procedure. To calculate the worst case bounds, we use information on the intensity of smoking in two periods, as well as information on the duration since quitting, which allows us to evaluate the probability of quitting between two given dates.

We first examine the UK evidence. In Table 5a, we report the regression results where we regress the log of the number of cigarettes smoked (column 1 and 4), the log of cotinine (column 2 and 5) and the log of smoking intensity (column 3 and 6) on the relative price of cigarettes and individual characteristics. The estimations pool the years 1993, 1994, 1996, 1998, and 2001. Price variation is annual. The first three columns display results using ordinary least squares on a sample of current smokers. The last three columns display the results using a Heckman selection model, using the sample of all current and former smokers.

[Table 5a]

Table 5a, Panel A displays the average price elasticity of smoking. At -0.81, the price elasticity of the number of cigarettes is in line with those reported in the literature. The second column reports the price elasticity of cotinine. The elasticity is not statistically different from zero and is in fact positive, at 0.22. This reflects the fact that the average cotinine concentration in smokers has remained largely constant over that period, despite large price increases. The rather large standard error reflects the heterogeneity in changes in cotinine levels across different groups over time, which we document below.

The price elasticity of the intensity of smoking is positive at about 0.43, although it is also statistically insignificant. Columns 4 to 6 displays the corresponding elasticities, corrected for selection due to non participation. The elasticity of cotinine and smoking intensity are closer to zero, but not different from the OLS results.

Next, we interacted prices with a number of different observed characteristics. The results are presented in Table 5a, Panel B. The reference individual is a low income, young, non white female. The price elasticity of the reported number of cigarettes is clearly negative at -1.78 for this group. There is a clear race effect, with whites having lower price elasticities, at around -0.80. The second column displays the elasticity of cotinine for various groups. It varies from 0.89 for white women over 40 to -0.37 for young rich men. Column 3 displays the elasticity of the intensity of smoking. It ranges from 1.55 (statistically significant at the conventional 5% level) for non white, poor, young women to 0.14 for elder, rich, white men. We get similar results when attempting to correct for selection.

[Figure 3a]

We now turn to the results using worst case bounds to check whether these results are attributable only to selection. In Figure 3a, we display the cumulative distribution function of smoking intensity, for the year 1993 and the worst case bounds for the distribution in 1994. These bounds represent the maximum and minimum effect if all those who quit belonged to either the highest or the lowest intensity group. Because few smokers have quit during that interval of time, the bounds are quite tight.⁹ The cumulative distribution in 1993 is above the upper bound in 1994 for most of the range of intensity. This indicates that the intensity of smoking has significantly increased, a fact that cannot be attributed to selection alone. For the lowest values of the intensity the test is inconclusive as the 1993 cumulative distribution is between the two worst case bounds. We get similar results when comparing other adjacent years. When we compare the 1993 CDF with later ones, the bounds become too large and we cannot conclude as to the effect of prices in one way to the other. This is because a larger fraction of smokers who smoked in 1993 have quit.

⁹ For quitting, we use information on the duration since quitting which is recorded in the data set.

Overall, our results show that prices have an effect on the intensity of smoking, independently of any selection out of smoking.

[Table 5b]

Table 5b presents similar results for the US. Panel A displays the average elasticity of smoking with respect to taxes (federal and state level taxes). We use taxes instead of prices as the latter are potentially endogenous. Moreover, from a policy point of view, the relevant variable is the excise tax. All the regressions control for age, sex, race, education levels, occupation, income level, household size, as well as year and state dummies. As in Table 5a, the first three columns report OLS results while the last three report the results from a Heckman selection model.

The elasticity of the number of cigarettes is equal to -0.65. The elasticity of cotinine is positive (0.79), but not statistically different from zero at the 5% level. Here again, the large standard errors reflect the heterogeneity in behavior. The elasticity of the intensity of smoking is clearly positive at 1.18 and different from zero at the 1% level. Controlling for selection does not change the results significantly.

Panel B allows for heterogeneity in the tax elasticities. The elasticity of the number of cigarettes varies from -1.02 for white men to -0.34 African-American women. The elasticity of cotinine varies from 0.36 for elder white men, up to 1.26 for young African-American women. As a result, the elasticity of the intensity of smoking varies from 1.05 for elder men to 1.37 for younger women. Again, correcting for selection does not change the results significantly. These results are therefore similar to the UK, and show that nicotine intakes are not sensitive to prices (or taxes), contrary to the quantity of cigarettes smoked.

[Figure 3b]

We next present the results using worst case bounds. For the US, we have two limitations. First, the sample is smaller than for the UK, which means that the computation of the bounds is more difficult. Second, when comparing two periods of time, not all states would have seen an increase in the price of tobacco. We therefore separate the sample into two periods, whether the interview took place in 1988-1990 or 1990-1994. We only consider those who live in a state in which a price increase has occurred between these two periods. Figure 3b presents the cumulative distribution of smoking intensity in the first period (labeled 1989) together with the bounds for the

second period. For an intensity of smoking of 16 ng/ml and above, there is clear evidence that the intensity increased between these two periods, as the cumulative distribution function is above the upper bound. Below that level, the test is inconclusive.

The evidence for the US is therefore in line with the one for the UK, suggesting that the increase in smoking intensity is not entirely due to selection but driven by higher prices.

Finally, we investigate the effect of state specific tobacco control policies. We use information on smoke free air laws collected by the ImpactTeen project¹⁰. The data provides a detailed account of different laws at the state level, between 1991 and 2001. Given that our data set starts in 1989, and the small variation in state laws, we used only cross-section variations in state laws (smoke free air laws and smoke-free air preemption laws) to identify a possible effect on intensity. We use the average number of restriction over the period 1991-1993, and construct a dummy of high regulation if the state has 20 or more regulations in place. Only three states in our data set had a high level of regulation, Pennsylvania, Florida and Illinois. We found mixed results. The effect of restrictions is mostly insignificant, apart for African-Americans of working age, for whom the smoking intensity is about 13% higher (but significant at the 10% level only). Given the lack of a clear pattern, we do not report the results. To identify an effect of smoking bans, we would need better data, reporting for instance workplace restriction on smoking.

B. Differential Cohort Effects

In this subsection, we analyze the change of behavior over time in both countries for some particular socio-economic groups.

For the UK, we focus in particular on the gender effect. While the gender effect in smoking intensity appeared to be limited when pooling all cross-sections, we show that behavior changed over time. For the US, we present results on behavior by race, as the gender differences appear to be more stable over the period of observation. Note also that we cannot investigate the effect of race for the UK, given the small proportion of non whites in our sample.

¹⁰ <http://www.impactteen.org>

Using repeated cross-sections, we construct a pseudo-panel following over time groups of fixed membership, according to their year of birth and sex (or race for the US). We measure the average number of cigarettes smoked, the average cotinine concentration and smoking intensity (cotinine per cigarette smoked).

1. Gender Differences in Behavior

Figure 4a plots the average number of cigarettes as a function of age, sex and birth cohorts for the UK. Each line follows a given birth cohort as they age.

As expected, there is a clear effect of gender, with men smoking more than women. The graph shows also a clear cohort effect, with members of older cohorts smoking more.

[Figures 4a, 4b, 4c]

Following different cohorts through time shows that quantities are declining slightly, at least for the younger cohorts. This reflects mainly the large rise in cigarette prices over that period. For older cohorts, there may be a change in the composition in the pool of smokers where lighter smokers are more likely to quit, which may result in either an apparent increase of smoking through time or, at least, a lesser decrease. This is why we are only presenting the results for younger cohorts. Apart from a difference in levels, the change in cigarette consumption appears to be similar across gender groups.

Figure 4b plots the average cotinine concentration as a function of age, sex and birth cohort. Here, the dynamic of smoking appears to be different across gender groups. Men see the average cotinine concentration clearly declining with time, reflecting the decrease in the number of cigarettes smoked. On the contrary, for women, the cotinine level is generally *increasing*, despite the moderate decrease in quantities.

Figure 4c plots the average smoking intensity, as a function of age, sex and birth cohort. The graph shows also a clear cohort effect in levels, with members of older cohorts smoking more intensively. The intensity of men is decreasing over time whereas it is increasing for women, at least up to age 40. Looking at quantities alone, one would

conclude, spuriously, that price increases, which have taken place during the period of analysis have been successful, even for women, when in fact this is not the case.

Moreover, the observed differential trends in intensity seem not to support the hypothesis that the sample changed in composition, because of the price increases, with heavier smokers representing an increasing proportion of the pool of smokers.

Conventional studies using only reported quantities miss the heterogeneity in smoking behavior across gender, education and income groups. But more importantly, they also miss time varying differences in smoking behavior across groups. Given that the intensity has an impact on health, Figure 4c shows that, although cigarette consumption decreased for women, the health impact of smoking, for those who did not quit, may have worsened.

2. Race Differences

We follow the same methodology as for the UK. We use data from both NHANES III and NHANES 1999-2000.

To save on space, we do not report the profiles of cigarette consumption or cotinine concentration, but directly the cotinine concentration per cigarette, our measure of the intensity of smoking. The results are reported in Figure 5.

[Figure 5]

There is a clear difference in smoking intensity by race, with African-Americans smoking more intensively than whites. On average the first group has a cotinine concentration of about 30ng/ml per cigarette smoked, while the latter is at about 20ng/ml. Moreover, there is a differential trend over time. For whites, the intensity appears to be slightly decreasing over time, while for African-Americans, it is actually increasing for some cohorts, especially those who are born in the 1950s or 1960s.

For these young cohorts, the intensity of smoking have increased by about 5 ng/ml over that period, while for whites it has decreased by up to 5 ng/ml. Further analysis shows

that this widening race difference is present for both males and females, but is more marked for young males.¹¹

V. Are the Estimates in the Economic Literature on Smoking Biased?

The economic literature on smoking behavior uses the number of cigarettes consumed as a measure of smoking. In this section, we argue that this leads to misleading conclusions, especially for the estimation of models of addiction.

In the previous section, we have shown that self-reported quantities explain only a small proportion of the heterogeneity in smoking, as measured –more accurately- by the concentration of cotinine in body fluids. In other words, the traditional measure of smoking is subject to a large measurement error, which can lead to a bias when smoking is used as an explanatory variable. Moreover, when using quantities of cigarettes as a proxy for smoking, the intensity of smoking is left as an omitted variable. As we have shown, the intensity is correlated with prices, which are usually used as instruments in the estimation of addiction models. The implication is that these instruments are not valid, as they are correlated with the error term, which also leads to biased estimates. In this section, we quantify some of these biases using our data sets for the UK and the US.

Suppose that actual smoking, noted S , is proportional to the product of the number of cigarettes, C , and the effort or intensity of smoking, E .

$$S = C.E$$

Denote by lower case letters the log of the variable, e.g. $s = \log(S)$. In many studies, the important variable to model is actual smoking, captured by S , not the number of cigarettes smoked per se. This is certainly the case when studying addiction, as the nicotine level in blood should be more important than the reported cigarette consumption. A second example is the effect of tobacco on health and mortality, where the quantity of nicotine, tar and carcinogens inhaled determines the health impact of smoking.

¹¹ These differential trends are also an indication that part of the racial difference in cotinine levels is behavioural.

In previous studies, only C is observed and used as a proxy for S . In the following, we assume that smoking is fully captured by cotinine measures, but not by the number of cigarettes because the smoker can adjust the effort. We characterize the bias in two different contexts. First, we look at a simple model, which relates an outcome variable to smoking, estimated simply by OLS. Second, we look at a model, which relates smoking to past smoking, estimated by instrumental variables, and where the instruments are prices.

A. Bias When Smoking is an Explanatory Variable

Suppose that an outcome variable Y is related to smoking such as:

$$Y = \beta s + u \quad (1a)$$

where Y can be, for instance, wages, health or other risky behavior. A number of studies have compared wages for smokers and non smokers (see for instance Hersch and Viscusi (1990) or Levine et al (1997)). An imposing literature, mainly in the epidemiological and medical field, has related health outcomes or mortality to the quantities of cigarette smoked. One example is for instance Doll et al (1994) who evaluate the duration to death in relation to use of tobacco.

If actual smoking is not observed, (1a) is estimated by replacing the log of actual smoking, s by c , the log of the number of cigarettes smoked. The equation which is estimated is then:

$$Y = b c + v \quad (1b)$$

The OLS coefficient of such a regression would be:

$$E\hat{b} = \frac{\text{cov}(c, Y)}{\text{Var}(c)} = \beta \frac{\text{cov}(c, s)}{\text{Var}(c)}$$

In computing the expression for the OLS coefficient, we have assumed that $\text{cov}(c, u)$ is zero. We can evaluate the bias, using our data sets, by computing the ratio $\text{cov}(s, c)/\text{Var}(c)$. We use the cross-sectional covariance between cotinine concentration and the number of cigarettes for all active smokers.

The bias is similar for the US and for the UK, at about 0.40. In other words an OLS regression would yield an estimated coefficient, $E\hat{b}$, 60% lower than its actual value.

In the case of the effect of tobacco on health, this represents a serious underestimate. Moreover, given that the intensity of smoking varies across groups, the bias is not uniform. It is larger for those who smoke more intensively. This is the case, for instance, for individuals with a low income for the UK, where the bias is equal to 0.33.

B. Bias When Estimating Addiction Models

In the last two decades, the economic literature has devoted a lot of effort to model the addictiveness of tobacco and to evaluate its magnitude. It is an important issue as it allows measuring the short and long run effect of public policies, especially the response of smoking to a change in prices.

To illustrate our point, we use a model of rational addiction as used in Becker et al (1994) or Chaloupka (1991), where current smoking depends on past and future smoking, because of the addictive nature of nicotine and the forward looking behavior of the agent. Suppose we are interested in a model such as:

$$s_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 s_{t+1} + \beta_3 p_t + u_t \quad (2a)$$

This equation can be derived from the first order condition of the program of such a forward looking agent. Hence the coefficients β_1 , β_2 and β_3 are linked to underlying preference parameters. In particular, the theory imposes that $\beta_1 > 0$ and $\beta_2 > 0$.

The economic literature does not observe smoking (nicotine) directly, s_t , but the number of cigarettes consumed, c_t . The equation estimated instead of (2a) is therefore

$$c_t = b_0 + b_1 c_{t-1} + b_2 c_{t+1} + b_3 p_t + v_t \quad (2b)$$

In doing this, the intensity of smoking, e_t , is an omitted variable and is subsumed into the error term. Note that the error term contains both the effort in period t , $t-1$ and $t+1$, as smoking appears on both sides of the equation. Suppose that equation (2b) is estimated by IV, using leads and lags of prices (or taxes), p_t , as an instrument. This is a standard practice, because the model incorporates a lagged dependent variable which may be correlated with the error term due to unobserved fixed characteristics.

Given that prices are correlated with the intensity, e_t , the IV will not yield an unbiased estimate.

For simplicity of notation, let's denote $\mathbf{X}=[1 \ c_{t-1} \ c_{t+1} \ p_t]$ and $\mathbf{Y}=[1 \ s_{t-1} \ s_{t+1} \ p_t]$ two $T \times 4$ matrices with the explanatory variables in each equations. Both equations can be written more compactly in matrix notations:

$$\mathbf{s} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{u} \quad (3a)$$

$$\mathbf{c} = \mathbf{X}\mathbf{b} + \mathbf{v} \quad (3b)$$

where $\boldsymbol{\beta}$ and \mathbf{b} are two 4×1 matrices of coefficients. Let \mathbf{Z} be a $T \times 4$ matrix of instruments. The instrumental variable estimator of \mathbf{b} can be written as:

$$E\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{c} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y}\boldsymbol{\beta} - (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{e} \quad (4)$$

where we have used the definition $\mathbf{s}=\mathbf{c}+\mathbf{e}$.

The IV estimation of equation (2b) leads to a biased estimates of the original parameters $\boldsymbol{\beta}$. The bias is composed of two terms, one multiplicative, and a second, additive which depends on the covariance matrix between the instruments (prices or taxes) and the effort to smoke a cigarette. Note that if the number of cigarettes correctly approximates smoking, the first term in (4) would be equal to the identity matrix, while the second term would be zero. In this case, there would be no bias introduced by approximating (2a) with (2b).

We compute these biases for both countries using data on the number of cigarettes, cotinine levels and prices or taxes. The results are displayed in Table 6a. We augment models (2a) and (2b) to control for individual and area characteristics. The control variables are age, sex, race and education for both countries. For the UK, we also control for the region of living and a year trend, while for the US, we control for year and state specific effects. For the US, we distinguish the effect of using either prices or excise taxes as an instrument. For both the UK and the US, the instruments are current prices as well as one lag and lead of prices.

[Tables 6a & 6b]

In all cases, the first term is different from the identity matrix, while the second term is different from zero, which implies that the rational addiction literature suffers from biases. From Table 6a, it is difficult to get a sense of the importance of the bias. We therefore compute the vector of coefficients β implied by a set of estimated coefficients chosen to be $b_1=0.5$, $b_2=0.5$ and $b_3=-0.5$. These correspond to a model estimated with the number of cigarettes as variables, appearing to display addiction and forward looking behavior. These numbers are not at odds with those estimated by Becker et al (1994) or Chaloupka (1991). The implied coefficients for model (2a) are displayed in Table 6b.

For the UK, the effect of past smoking, β_1 , is higher than b_1 , while the effect of future smoking, β_2 , is very close to zero. Note that the implied price effect is positive. In this light, once corrected for the bias, one would conclude that agents are myopic as future consumption does not appear to play a significant role.

For the US, the bias is even worse. When prices are used, the implied effect of future smoking is negative, which is in contradiction with the theory. Similarly, when taxes are used, the effect of past smoking is large and negative, which is not consistent with the theory either.

In all cases, IV can lead to a severe bias, which potentially casts some doubts on the validity of previous results, especially in the addiction literature. With only data on the number of cigarettes, it is difficult to find a convincing set of instruments, which would affect the number of cigarettes smoked, without changing the intensity of smoking. A better way to test the rational addiction model would be to use panel data on cotinine measures but that type of data is not easy to come across.

VI. Conclusion

Our paper shows that smoking behavior is much more complex than what the empirical economic literature has suggested. There are different ways to smoke a cigarette, and smoking behavior is influenced by economic variables such as income or prices in complex ways. Smokers adjust not only the quantities of cigarettes, but also the intensity with which they smoke them.

In this paper, we define intensity as the amount of cotinine extracted per cigarette. This is a broad definition, which encompasses many compensating behaviors, such as smoking heavier or longer cigarettes but also deeper inhalation, more frequent puffs per cigarette and smoking more of a cigarette. Hence our characterization of compensating behavior is broader than the one considered by Evans and Farrelly (1998) and we show that a large heterogeneity in smoking remains even when controlling for detailed characteristics of the cigarette smoked.

Moreover, we show that the intensity varies with observed characteristics such as income, wealth, age and ethnicity. We also show that higher prices lead some socio-economic groups to smoke more intensively. This is the case, for instance, for women in low socio-economic groups and African-Americans. Smoking more intensively is not without health consequences and the disparity in smoking intensity shows up in lung cancer rates. Miller et al (1996) and Campbell et al (2000) show that African-Americans suffer disproportionately more from lung cancers, despite the fact that they smoke less cigarettes than whites. This disparity can be partly explained by the differences in intensity of smoking.

Given that the intensity of smoking affects health, factors that change the intensity may lead to more severe health threats. Consider for example an increase in the relative price of cigarettes. This would lead smokers to decrease the quantities consumed, but it would also lead to an increase in the intensity, at least for some groups of individuals. We also found that the intensity is significantly related to income levels. This suggests that recessions may lead to fewer quantities but to an increase in the intensity. Ruhm (2000) argues that the lower death rates during recession may occur because of improved health behaviors. Our results show that the apparent health behavior improvement during recessions may be in fact spurious in some smokers belonging to poorer socio-economic groups.

The discrepancy between the number of cigarettes smoked and cotinine concentration implies that the evaluation of economic models of smoking behavior is subject to considerable bias. Our results call for more analysis using similar data to understand better the complexity of smoking, and its implication for health. The economic

literature on smoking has much to gain from exploiting this new source of data. Cotinine concentration can be used to better understand the process of addiction in which nicotine plays an important role. It should also be helpful in understanding differences in quitting across ethnic or socio-economic groups, with different smoking intensities. Similarly, cotinine measures should better capture health effects of smoking. A natural extension of the present work is the analysis of passive smoking, whose dangers have been stressed by a large medical and epidemiological literature. These issues are addressed in Adda and Cornaglia (2004) who estimate the price elasticity of second hand smoke using the same data sets used in this paper.

REFERENCES

- Adda, J. and F. Cornaglia (2004) “The Price Elasticity of Passive Smoking”, mimeo University College London.
- Becker, G, Grossman M and K Murphy (1994) “An Empirical Analysis of Cigarette Addiction”, *American Economic Review*; 84; 3; 396-418.
- Becker, G. and K. Murphy (1988) “A Theory of Rational Addiction”, *Journal of Political Economy*, 96, 4, 675-699.
- Benowitz, N, Perez-Stable, E, Fong I, Modin, G, Herrera, B and P Jacob, III (1999) “Ethnic Differences in *N*-Glucuronidation of Nicotine and Cotinine”, *Journal of Pharmacology and Experimental Therapeutics*, 291, 1196-1203.
- British National Formulary (BNF) Number 47 (2004), section 4.10. Pharmaceutical Pr; 1st edition. British Medical Association, Royal Pharmaceutical Society of Great Britain, London.
- Campbell D. E., Greenberg E. R., Polednak A. P., DeShazer C., Bach P. B., Cramer L. D., Begg C. B., Warren J. L. (2000) “Racial Differences in the Treatment of Early-Stage Lung Cancer”, *New England Journal of Medicine* ; 342:517-519, Feb 17.
- Caraballo, R, Giovino G, Pechacek T, Mowery P, Richter P, Strauss W, Sharp D, Eriksen M, Pirkle J and K Maurer (1998) “Racial and Ethnic Differences in Serum Cotinine Levels of Cigarette Smokers”, *JAMA*, 280:135-139.
- Caraballo R, Giovino G, Pechacek T and P Mowery (2001) “Factors Associated with Discrepancies between Self-Reports on Cigarette Smoking and Measured Serum Cotinine Levels among Persons Aged 17 Years or Older Third National Health and Nutrition Examination Survey, 1988–1994“, *American Journal of Epidemiology*, 153, 8, 807-814.
- Chaloupka, F (1991), “Rational Addictive Behavior and Cigarette Smoking”, *Journal of Political Economy*, 99, 4, 722-742.
- Chaloupka, F. and Warner K., (2000), “The Economics of Smoking”, Handbook of Health Economics, J Newhouse and A Cuyler editors.
- Clark, P, Gautam S, Hlaing W and L Gerson (1996) “Response error in self-reported current smoking frequency by black and white established smokers”, *Annals of Epidemiology*, Volume 6, Issue 6, 483-489.
- Davey Smith, G Chaturvedi N, Harding S, Nazroo J, Williams R (2000), “Ethnic Inequalities in Health: a Review of UK Epidemiological Evidence”, *Critical Public Health*, 10,4, 375-408.

Doll, R, Peto R, Wheatley K, Gray R, and I Sutherland (1994) "Mortality in Relation to Smoking: 40 years' observation on Male British Doctors", *British Medical Journal*, 309, 6959, 901-911.

Evans W and M Farrelly (1998) "The Compensating Behavior of Smokers: Taxes, Tar, and Nicotine", *RAND Journal of Economics*, 29, 3,578-595.

Evans W, Farrelly M and E Montgomery (1999) "Do Workplace Smoking Bans Reduce Smoking?" *American Economic Review*; 89; 4; 728-747.

Grossman M, Sindelar, J Mullahy, J and R Anderson (1993) "Policy Watch: Alcohol and Cigarette Taxes", *The Journal of Economic Perspectives*, 7, 4, 211-222.

Gruber J. (2001), "Risky Behavior Among Youth: An Economic Analysis, Introduction" in Jonathan Gruber, ed., *Risky Behavior Among Youth: An Economic Analysis*. Chicago: University of Chicago Press, p. 1-28.

Gruber J and J Zinman (2001) "Youth Smoking in the U.S.: Evidence and Implications," in Jonathan Gruber, ed., *Risky Behavior Among Youth: An Economic Analysis*. Chicago: University of Chicago Press, 2001, p. 69-120.

Gruber J and B Koszegi (2001) "Is Addiction 'Rational'? Theory and Evidence", *Quarterly Journal of Economics*, 116, 4,1261-1303 .

Harris, J (1980) "Taxing Tar and Nicotine", *American Economic Review*, 70, 300-311.

Heckman, J (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47,1, 153-160.

Hersch,J and K Viscusi (1990) "Cigarette Smoking, Seatbelt Use, and Differences in Wage-Risk Tradeoffs" *Journal of Human Resources*, Spring; 25(2): 202-27.

Jarvis, M J, Boreham R, Primatesta, P, Feyerabend, C and A Bryant (2001), "Nicotine Yield from Machine-Smoked Cigarettes and Nicotine Intakes in Smokers: Evidence from a Representative Population Survey", *Journal of the National Cancer Institute*, 93, 2, 134-138.

Jarvis, M.J., Primatesta P., Erens B., Feyerabend, Bryant A., (2003), "Measuring nicotine intake in population surveys: comparability of saliva cotinine and plasma cotinine estimates", *Nicotine and Tobacco Research*, 5, pp 349-355

Kozlowski, L.T., Frecker, R.C., Khouw, V., Pope, M.A. "The Misuse of "Less Hazardous" Cigarettes and its Detection: Hole-blocking of Ventilated Filters". *American Journal of Public Health* 70(11): 1202-1203, November 1980.

Levine, P Gustafson, T, Velenchick (1997) "More Bad News for Smokers? The Effects of Cigarette Smoking on Wages", *Industrial and Labor Relations Review*. April, 50(3): 493-509.

McCarthy WJ, Caskey NH, Jarvik, ME (1992) "Ethnic Differences in Nicotine Exposure", *American Journal of Public Health*, 82, 1171-1172.

Manski, C (1994) "The Selection Problem" in *Advances in Econometrics, Sixth World Congress* Vol 1. Editor C Sims, Cambridge University Press.

Miller BA, Kolonel LN, Bernstein L, et al. (1996) "*Racial/Ethnic Patterns of Cancer in the United States, 1988-1992*". Bethesda, Md: National Cancer Institute;. NIH publication 96-4104.

National Cancer Institute (2004) <http://www.cancer.gov/newscenter/healthdisparities>.

Perez-Stable, E. J, Herrera B, Jacob III P, and N. L. Benowitz (1998) "Nicotine Metabolism and Intake in Black and White Smokers", *JAMA*, 280(2): 152 - 156.

Ruhm C. (2000) "Are Recessions Good for Your Health?", *Quarterly Journal of Economics*, 115, 2, 617-650.

MJ Thun, CA Lally, JT Flannery, EE Calle, WD Flanders, and CW Heath, Jr (1997) "Cigarette smoking and changes in the histopathology of lung cancer", *Journal of the National Cancer Institute*, 89: 1580-1586.

Wagenknecht, GR Cutter, NJ Haley, S Sidney, TA Manolio, GH Hughes, and DR Jacobs (1990) "Racial differences in serum cotinine levels among smokers in the Coronary Artery Risk Development in (Young) Adults study" *American Journal of Public Health*, 80, 1053-1056

Appendix A: Worst Case Bounds

We summarize here the methodology in Manski (1994). Let e denote the effort exerted to smoke a cigarette and X a set of conditioning variables. Let I be an indicator which takes the value of 1 if e is observed and 0 otherwise. In our case, $I=1$ indicates that the individual is still smoking, whereas $I=0$ indicates that the person has quit since the start of the sample. The probability of $I=1$ given X is written as $P(X)$, the probability of not quitting. We write the conditional cumulative distribution of e given X in period t as $F_t(e|X)$.

We are interested to compare $F_t(e|X)$ and $F_{t+1}(e|X)$, assuming that a price increase has taken place between period t and $t+1$. While in our data we can characterize the cumulative distribution $F_t(e|X)$, as we observe every smoker in that period, it is not the case for period $t+1$. Some smokers may have quit, and if quitting is not random the cumulative distribution in period $t+1$ reflects both the effect of prices on smoking intensity and the change in the composition of the sample. However we can write

$$F_{t+1}(e|X) = F_{t+1}(e|X, I=1)P(X) + F_{t+1}(e|X, I=0)[1 - P(X)] \quad (A1)$$

The data in period $t+1$ identify the distribution $F_{t+1}(e|X, I=1)$ and $P(X)$, but not the distribution of the intensity of those who happened to quit between t and $t+1$, $F_{t+1}(e|X, I=0)$.

However, the inequality

$$0 \leq F_t(e|X, I=0) \leq 1$$

is always true. Using it in (A1) produces bounds

$$F_{t+1}(e|X, I=1)P(X) \leq F_{t+1}(e|X) \leq F_{t+1}(e|X, I=1)P(X) + [1 - P(X)] \quad (A2)$$

or

$$F_{t+1}^{low}(e|X) \leq F_{t+1}(e|X) \leq F_{t+1}^{high}(e|X)$$

Clearly, if $P(X)$ is close to one, i.e. very few smokers quit, then the bounds will be tight. If, however, the probability of quitting is large, then the bounds become wider and less informative.

We implement this procedure using two adjacent dates in the survey to minimize the effect of quitting. We control for observable characteristics such as age, sex, education

level, race, as well as state fixed effects for the US. The probability $P(X)$ is computed as the proportion of individuals who have not quitted from period t to $t+1$. We use information on the duration since quitting which is documented in both surveys.

Higher prices increases smoking intensity if

$$F_{t+1}(e | X) \leq F_t(e | X) \quad (\text{A3})$$

a sufficient condition is

$$F_{t+1}^{high}(e | X) \leq F_t(e | X) \quad (\text{A4})$$

although there could be cases where (A3) is satisfied even if (A4) is not.

Table 1: Descriptive Statistics

	UK			US		
	All	Smokers	Non Smokers	All	Smokers	Non Smokers
# of observations	43886	11208	32658	18162	4464	13698
average # of cigarettes	3.8	15.0	0	5.2	18.8	0
	(8.0)	(8.9)		(10.6)	(11.6)	
average level of cotinine (ng/ml)	87.4	305.9	7.4	96.1	288	7.5
	(157)	(175.6)	(63.7)	(174)	(190)	(49)
% education high	16.6	10.7	18.7	40	27	44
% education medium	43.7	48.9	43.2	49	63	44
% education low	37.5	40.3	37.6	11	10	12
average age	48.1	43.0	49.8	43.2	39	44.8
sex (% male)	47.6	48.9	47.1	47.7	53.8	45
% Professional and managerial	32.2	23.3	35.2	17	12.2	19
% skilled manual	44.6	45.9	44.1	27	27	27
% unskilled manual	23.2	30.7	20.6	24	34	21
% white	95.7	96.5	95	84.1	84	84.1
% income 1 st quartile		37.6	29.4		18	14
% income 4 th quartile		18.1	27.0		27	38

Standard deviations in parenthesis.

Table 2: Percentage of Explained Variance in Cotinine Intake for Smokers

	UK	26%	26%	30%	36%	-
		[11208]	[11208]	[3272]	[904]	
% explained variance in cotinine intake	US 1988-94	26%	26%	-	-	-
		[4122]	[4122]			
	US 1999-00	22%	-	24%	35%	36%
		[840]		[776]	[776]	[776]
Controls:						
	Number of cigarettes smoked	Yes	Yes	Yes	Yes	Yes
	Year effects	No	Yes	Yes	Yes	Yes
	Nicotine content of cigarette	No	No	Yes	Yes	Yes
	Brand of Cigarette	No	No	No	Yes	Yes
	Filter, Size, Menthol characteristics	No	No	No	No	Yes

Numbers in bracket indicate the number of observations.

Table 3: Probability of claiming not to smoke having a cotinine level >15ng/ml

	UK	US
Sex: men	0.01** (0.002)	0.005 (0.004)
Age < 25	0.05** (0.007)	0.008 (0.006)
Education low	0.006* (0.003)	-0.011* (0.006)
Education medium	0.008** (0.003)	0.0004 (0.005)
Skilled manual	0.008** (0.003)	0.02** (0.008)
Unskilled manual	0.014** (0.004)	0.037** (0.009)
White	-0.024** (0.006)	0.025** (0.01)
African-American	-	0.044** (0.02)
Observations	30887	7052
R-squared	0.02	0.02

Robust standard errors in parentheses. * significant at 10% level; ** significant at 5% level

Table 4a: UK: Determinants of Smoking as measured by log of cigarettes, log of cotinine concentration and log of cotinine concentration per cigarette smoked.

	Log(Cig)	Log(Cot)	Log(Cot/Cig)
Men	0.15** (0.02)	0.18** (0.04)	0.04 (0.03)
Low education	0.31** (0.05)	0.37** (0.08)	0.07 (0.05)
Medium education	0.25** (0.04)	0.31 (0.07)	0.08 (0.06)
Income [0%-33%]	0.05* (0.03)	0.19** (0.06)	0.14** (0.05)
Income [33%-66%]	0.06** (0.03)	0.10** (0.05)	0.06 (0.04)
Unskilled manual	0.10** (0.03)	0.12** (0.06)	0.03 (0.04)
Skilled manual	0.05* (0.03)	0.03 (0.05)	-0.01 (0.04)
Own house	-0.10** (0.02)	-0.09** (0.04)	-0.02 (0.04)
Size of house (number of bedrooms)	0.01 (0.01)	-0.04** (0.02)	-0.05** (0.02)
White	0.31** (0.05)	0.06 (0.1)	-0.23** (0.08)
Household size	0.03** (0.01)	0.08** (0.02)	0.05** (0.01)
Urban living	0.02 (0.02)	0.08* (0.05)	0.03 (0.04)
Age	0.047** (0.004)	0.08** (0.008)	0.03** (0.007)
Age squared	-0.0005** (0.00004)	-0.0007** (0.00008)	-0.0003** (0.00007)
Controls for region of living	Yes	Yes	Yes
Controls for year effects	Yes	Yes	Yes

Estimation done on 3810 individuals for years 1998 and 2001. Robust standard errors in parenthesis. **, *: significant at 5%, 10% level.

Table 4b: US: Determinants of Smoking as measured by log of cigarettes, log of cotinine concentration and log of cotinine concentration per cigarette smoked.

	Log(Cig)	Log(Cot)	Log(Cot/Cig)
Men	0.058* (0.034)	0.011 (0.045)	-0.064** (0.034)
Low education	-0.033 (0.063)	-0.011 (0.084)	0.028 (0.063)
Medium education	0.246** (0.039)	0.298** (0.052)	0.048 (0.039)
Income [0%-33%]	-0.017 (0.049)	-0.073 (0.065)	-0.052 (0.049)
Income [33%-66%]	0.047 (0.038)	0.052 (0.051)	0.003 (0.038)
Unskilled manual	0.143** (0.055)	0.153** (0.072)	0.020 (0.055)
Skilled manual	0.001 (0.054)	-0.002 (0.072)	0.041 (0.054)
Size of house (number of bedrooms)	0.124** (0.061)	0.153** (0.081)	0.061 (0.061)
White	-0.040** (0.009)	-0.077** (0.012)	-0.037** (0.009)
African-American	0.458** (0.092)	0.375** (0.121)	-0.100 (0.091)
Household size	-0.023 (0.100)	0.530** (0.133)	0.527** (0.100)
Urban living	0.016 (0.010)	0.040** (0.014)	0.027** (0.010)
Age	-0.123** (0.033)	-0.101** (0.044)	0.018 (0.033)
Age squared	0.067** (0.006)	0.051** (0.008)	-0.015** (0.006)
Age squared	-0.001** (0.000)	-0.000** (0.000)	0.000** (0.000)
Controls for region of living	Yes	Yes	Yes
Controls for year effects	Yes	Yes	Yes

Estimation done on 4036 individuals for years 1988-1994. Robust standard errors in parenthesis.

Table 5a: UK: Prices, cigarette consumption and smoking intensity (if smoker)

	OLS			Heckman Selection Model		
	Log Cig	Log Cot	Log Inten	Log Cig	Log Cot	Log Inten
Panel A: Average Price Effects						
Average price effect	-0.81** (0.40)	0.22 (0.71)	0.43 (0.56)	-0.85** (0.40)	0.05 (0.71)	0.26 (0.56)
Panel B: Heterogeneity in Price Effects						
Relative price	-1.78** (0.42)	0.41 (0.75)	1.55** (0.58)	-1.82** (0.42)	0.24 (0.75)	1.42** (0.59)
Rel. price * men	-0.10* (0.06)	-0.44** (0.11)	-0.26** (0.09)	-0.10* (0.06)	-0.40** (0.12)	-0.26** (0.09)
Rel. price * high income	0.07 (0.11)	-0.34 (0.22)	-0.44** (0.17)	0.07 (0.12)	-0.33 (0.22)	-0.43** (0.17)
Rel price * med income	0.04 (0.08)	-0.04 (0.17)	-0.18 (0.14)	0.05 (0.09)	-0.03 (0.17)	-0.18 (0.14)
Rel price*white	0.98** (0.09)	0.32* (0.18)	-0.64** (0.15)	0.97** (0.09)	0.32* (0.18)	-0.64** (0.15)
Rel. price * (age>40)	0.24** (0.06)	0.16 (0.10)	-0.05 (0.08)	0.24** (0.06)	0.16 (0.10)	-0.06 (0.08)
Observations	17714	10318	10245	43306	35838	35838

* significant at 10% level; ** significant at 5% level. Regression controls for age, sex, race, education, occupation, car ownership, size of house, region and year trend. Robust standard errors in parentheses. The reference individual in Panel B is a young women with low income and from an ethnic minority background.

Table 5b: US: Prices, cigarette consumption and smoking intensity (if smoker)

	OLS			Heckman Selection Model		
	Log Cig	Log Cot	Log Inten	Log Cig	Log Cot	Log Inten
Panel A: Average Price Effects						
Average Price effect	-0.65*	0.79	1.18**	-0.59*	0.83	1.18**
	(0.39)	(0.61)	(0.45)	(0.39)	(0.60)	(0.45)
Panel B: Heterogeneity in Price Effects						
Relative price	-0.35	1.26**	1.24**	-0.29	1.30**	1.25**
	(0.42)	(0.65)	(0.49)	(0.42)	(0.65)	(0.49)
Rel. price * white	-0.67**	-0.56*	0.13	-0.68**	-0.57*	0.13
	(0.24)	(0.32)	(0.28)	(0.24)	(0.31)	(0.28)
Rel. price * men	0.001	-0.24	-0.14	0.01	-0.23	-0.14
	(0.22)	(0.29)	(0.24)	(0.21)	(0.29)	(0.23)
Rel. price* (age>40)	-0.001	-0.11	-0.06	-0.01	-0.11	-0.06
	(0.12)	(0.19)	(0.13)	(0.12)	(0.19)	(0.13)
Observations	2041	1876	1861	4005	3840	3825

* significant at 10% level; ** significant at 5% level. Regression controls for age, sex, race, education, occupation, income level, household size, year and state effects. Robust standard errors in parentheses. The reference individual in Panel B is a young African American women.

Table 6a: Biases in the Rational Addiction Model.

	$(Z'X)^{-1}Z'Y$	$(Z'X)^{-1}Z'E$
UK, instrument: prices		
-3.0	-2.62	-0.06
2.41	0.27	0.05
-1.88	-0.89	0.56
US, instruments: prices		
-0.02	1.22	0.004
-0.64	-0.42	-0.002
0.41	0.59	0.93
US instruments: taxes		
-0.05	-0.31	0.002
0.59	0.31	0.02
0.59	0.47	0.94

Notes: Model controls for age, sex, education level, race, a trend for the UK and for state and fixed effects for the US. A constant was included in the regression. Lags and leads of prices or taxes were used as instruments.

Table 6b: Implied Values for β in Rational Addiction Model.

If $b_1=0.5$, $b_2=0.5$, $b_3=-0.5$

	UK instrument: prices	US instrument: prices	US Instrument: taxes
β_1 (past smoking)	0.75	0.14	-4.75
β_2 (future smoking)	0.03	-0.96	2.70
β_3 (prices)	1.97	1.60	-0.48

Table A: Individual Characteristics for those with and without a Valid Cotinine Measure

	UK			US		
	No Cotinine	Cotinine	t-stat	No Cotinine	Cotinine	t-stat
Number of Cigarettes	3.71	3.60	0.01	4.10	5.32	-0.08
Education, low	0.15	0.16	-0.02	0.31	0.40	-0.13
Education, medium	0.43	0.45	-0.03	0.51	0.49	0.02
Education, high	0.41	0.37	0.049	0.17	0.11	0.14
Age	42.67	42.50	0.005	48.6	42.8	0.20
Sex	0.43	0.48	-0.06	0.38	0.48	-0.14
Professional and managerial	0.29	0.32	-0.04	0.13	0.17	-0.09
skilled manual	0.45	0.45	-0.001	0.21	0.27	-0.09
unskilled manual	0.26	0.23	0.045	0.15	0.25	-0.18
Low income	0.30	0.23	0.11	0.24	0.15	0.16
High Income	0.22	0.25	-0.06	0.28	0.36	-0.12
White	0.89	0.94	-0.12	0.77	0.84	-0.14
African-American				0.18	0.11	0.14

The t-stat column provides a t-test of equality of the columns with and without a valid measure of cotinine.

Figure 1a: Cotinine levels as a function of the number of cigarettes smoked: UK

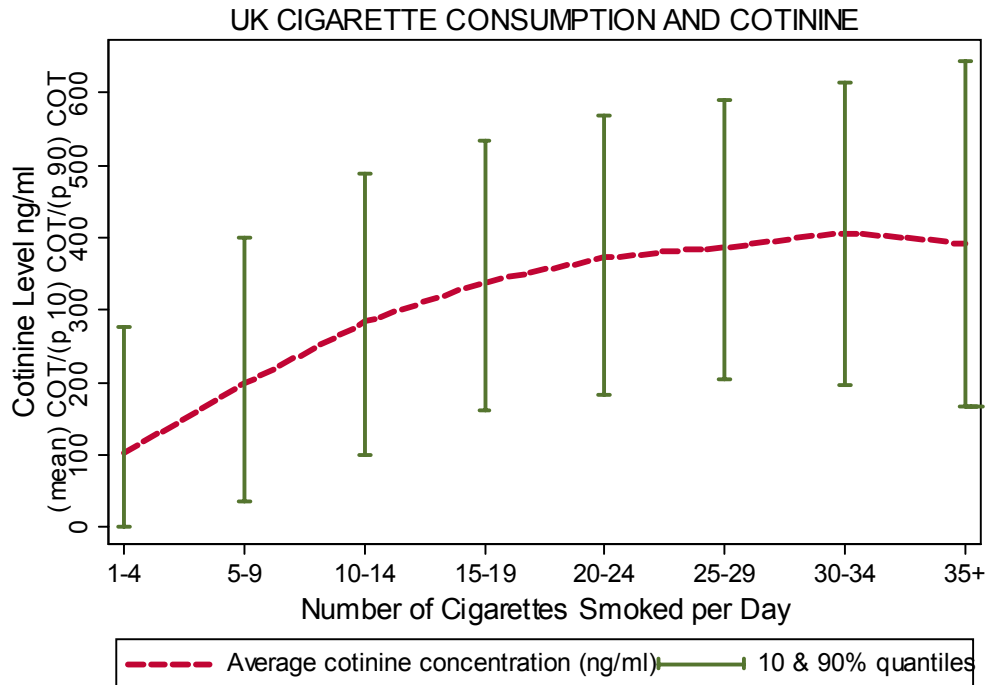


Figure 1b: Cotinine levels as a function of the number of cigarettes smoked: US

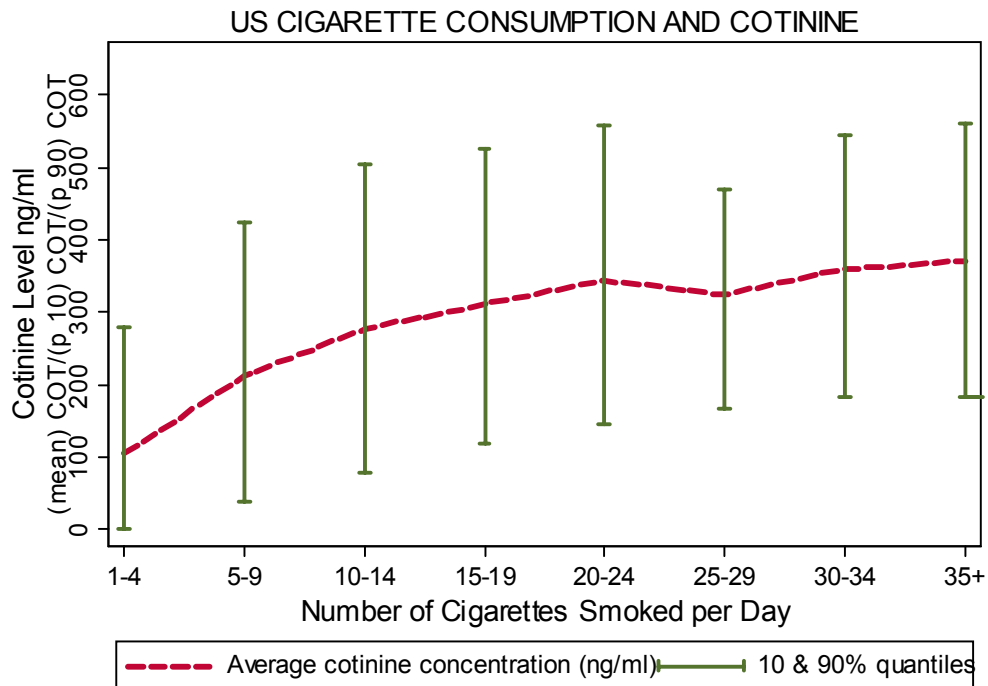


Figure 2: Average Smoking Intensity in Young Individuals, UK

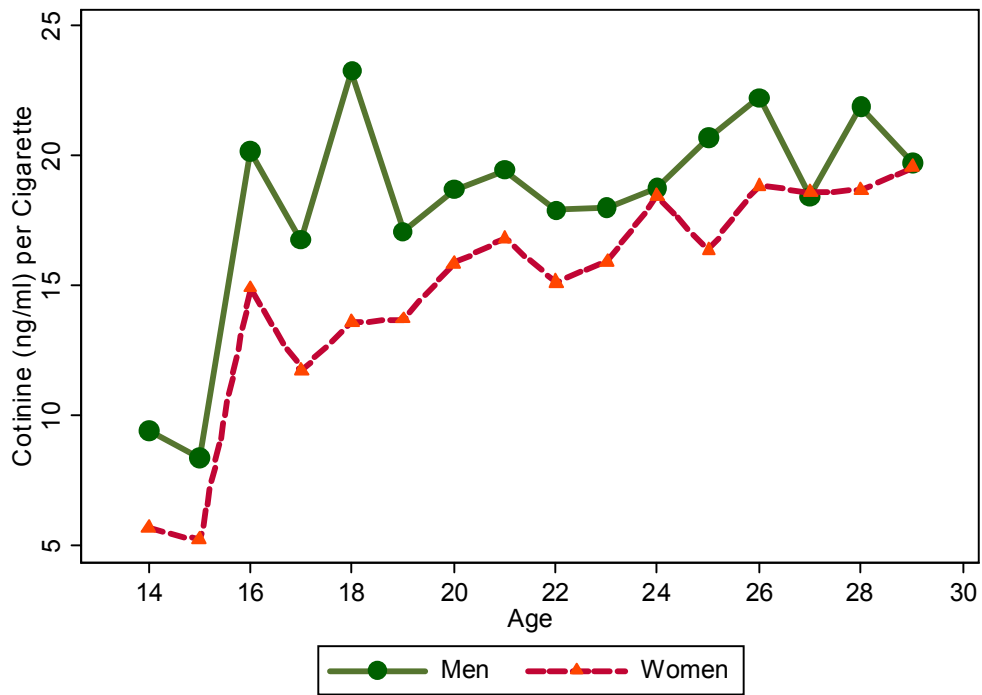


Figure 3a: Empirical Cumulative Distribution of Smoking Intensity, in 1993 and Worst Case Bounds in 1994, UK.

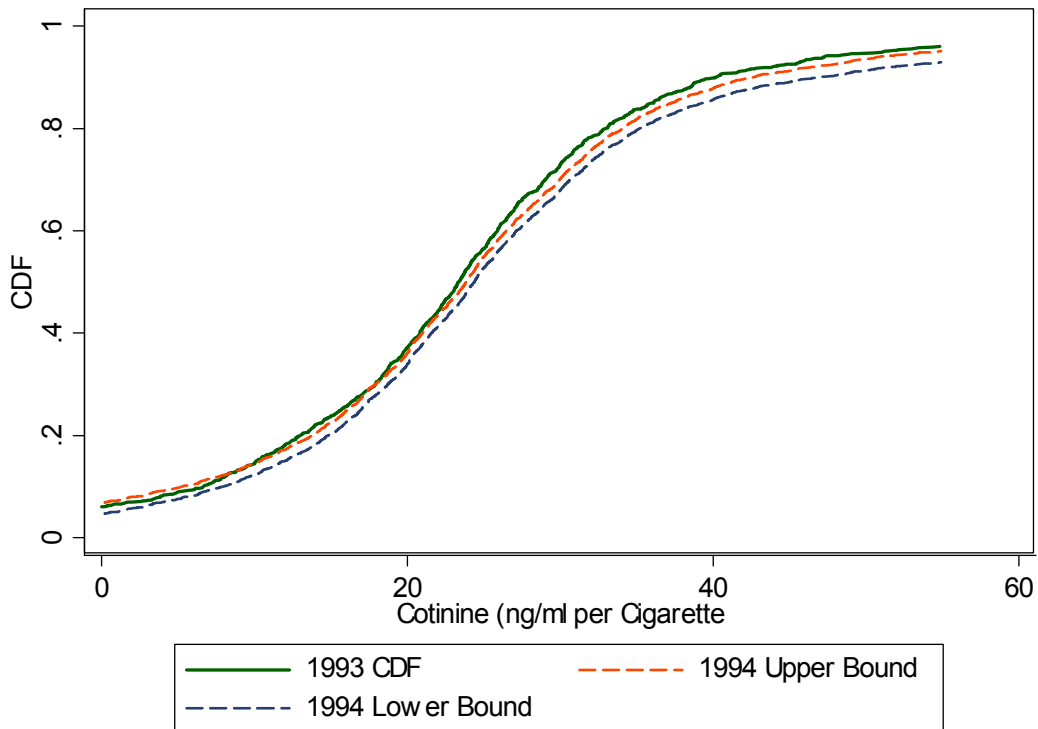


Figure 3b: Empirical Cumulative Distribution of Smoking Intensity, in 1989 and Worst Case Bounds for States in Which Prices Increased, US.

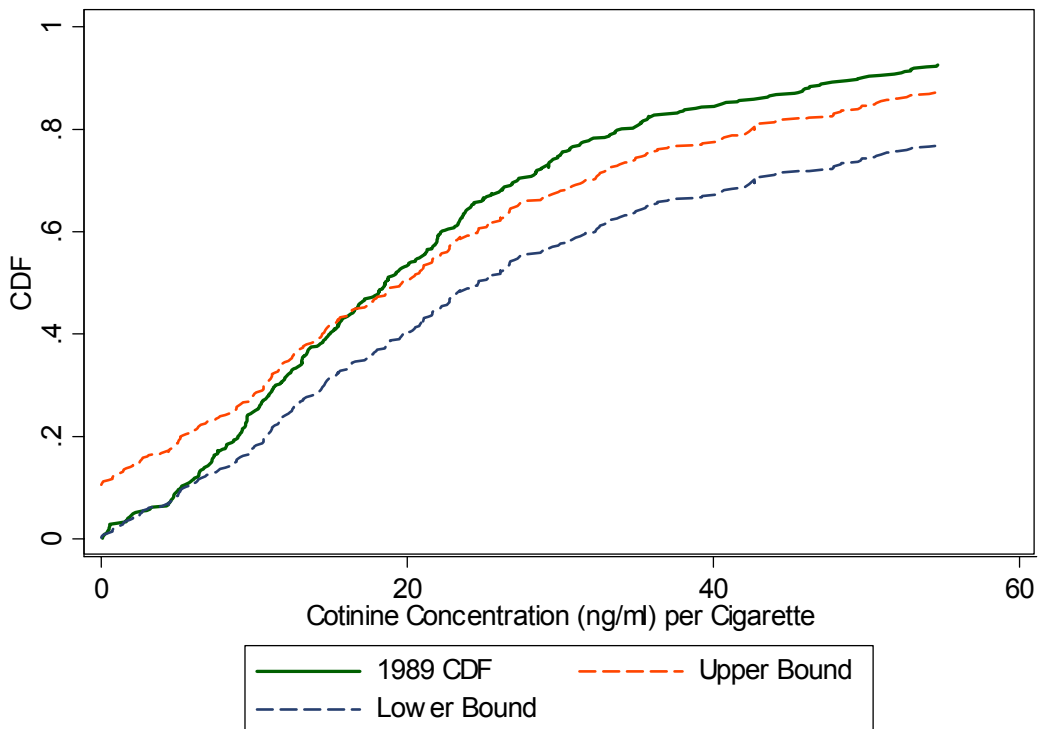


Figure 4a: Quantities of Cigarette per Day, by age, birth cohort and sex, UK.

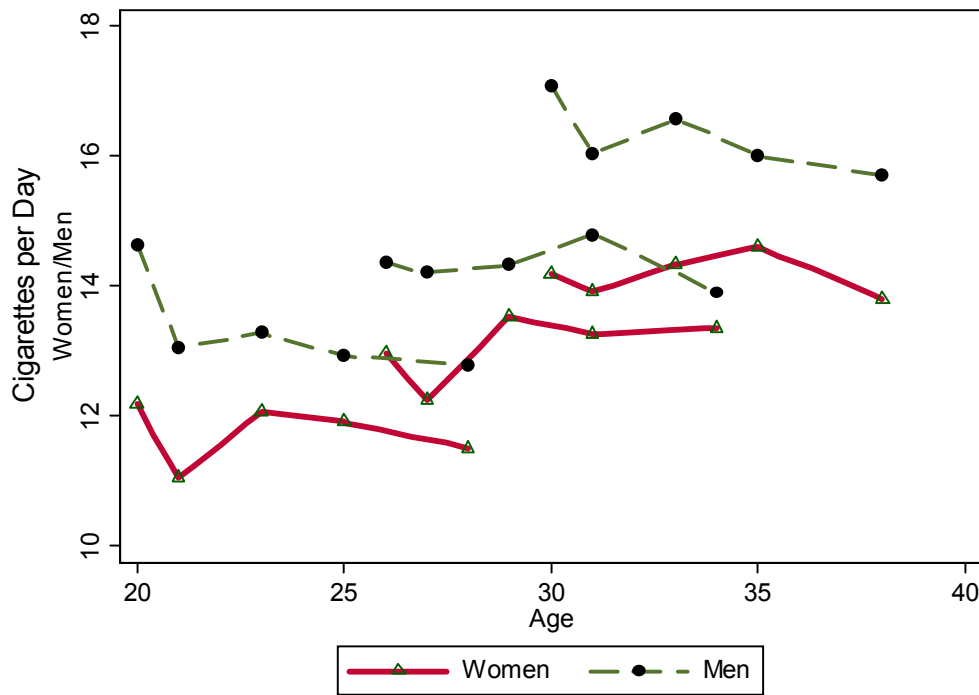


Figure 4b: Cotinine Concentration, by age, birth cohort and sex, UK.

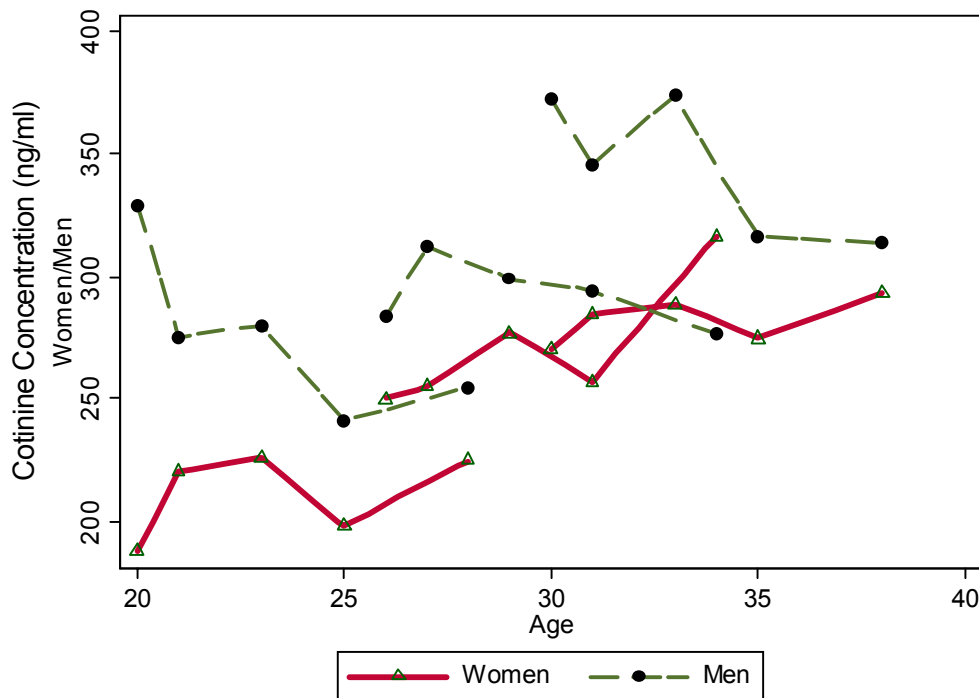


Figure 4c: Average Smoking Intensity for Smokers, by age, birth cohort and sex, UK

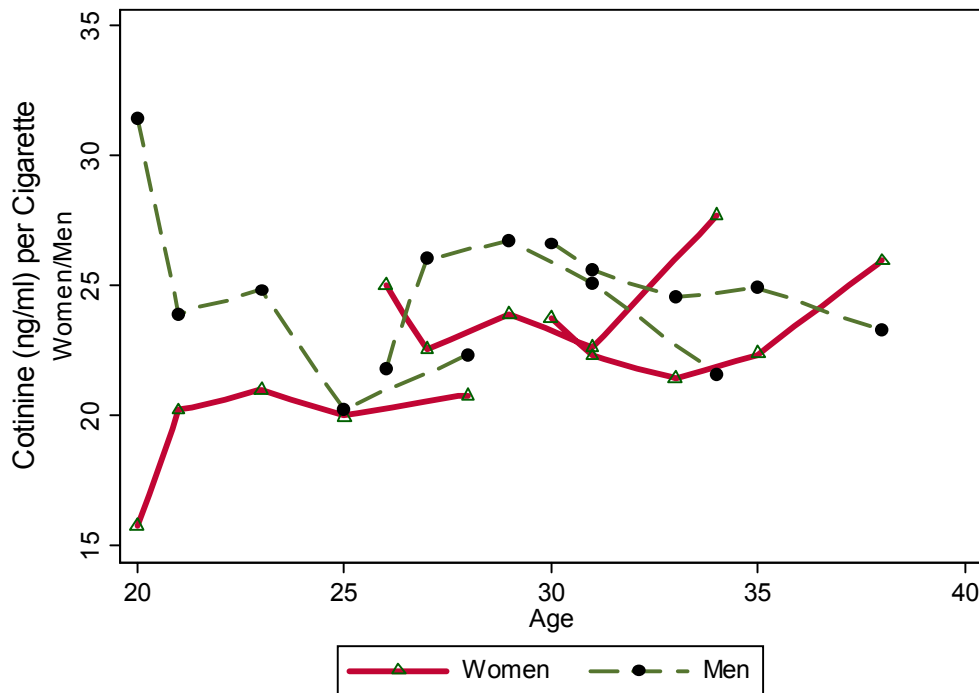


Figure 5: Average Smoking Intensity, by Age, Race and Birth Cohort, US

