

WP 12/29

Decomposing Differences in Cotinine Distribution between Children and Adolescents from Different Socioeconomic Backgrounds

Ijeoma P. Edoaka

December 2012

Decomposing Differences in Cotinine Distribution between Children and Adolescents from Different Socioeconomic Backgrounds

Ijeoma P. Edoka*

Centre for Health Economics, University of York
Department of Economics and Related Studies, University of York

6th December, 2012

Abstract

This study decomposes differences in saliva log cotinine between children/adolescents from low and high socioeconomic backgrounds using the 1997/98 cross-section of the Health Survey for England (HSE). Three decomposition methods are applied including a mean-based (Oaxaca-Blinder) decomposition method and two further methods that allow the decomposition of differences in quantiles (the quantile regression and the recentered influence function regression decomposition methods). By extending the analysis beyond differences in means, this study is able to identify the contributions of different characteristics to differences in quantiles of the log cotinine distribution. Differences in log cotinine between the two study groups are decomposed into a part explained by group differences in the distribution of characteristics (composition effect) and a part explained by group differences in the impact of these characteristics (structural effect). The composition effect accounts for a larger proportion of the total difference in log cotinine compared to the structural effect. The composition effect attributable to smoking within the home explains more of socioeconomic differences at lower quantiles indicative of passive smoking compared to higher quantiles that are indicative of active smoking while the composition effect of household income and parental smoking explains more of socioeconomic differences in active smoking compared to passive smoking. The structural effect of parental smoking and smoking within homes is indicative of underlying group differences in parents' compensatory behaviours that limit the impact of parents' risky lifestyle choices on child health.

JEL Classification: C1, I14

Keywords: cotinine; socioeconomic inequality; decomposition analysis; passive smoking; active smoking

* Correspondence to: Centre for Health Economics, University of York, Alcuin Block A, Heslington, York YO10 5DD. Email: ijeoma.edoka@york.ac.uk

1. Introduction

The adverse health consequences of passive and active smoking in children and adolescents are well established. Exposure to second-hand smoke or passive smoking has been associated with several adverse health outcomes in children including respiratory illnesses (bronchitis, pneumonia, asthma, coughing and wheezing), recurrent middle ear infection, brain tumours, leukaemia and meningitis (Tobacco Advisory Group of the Royal College of Physicians 2010; US Department of Health and Human Services 2006). In children, passive smoking has also been linked to impairments in mental development, affecting both reading and reasoning skills (Yolton et al. 2004) and repeated absence from school due to respiratory illnesses (Charlton 1996; Gilliland et al. 2003). Impairments in both the physical and mental development of the child could in turn have important consequences for future health outcomes and labour market participation in adulthood (Eriksen 2004 ; Graham and Power 2004). Similarly, active smoking in adolescents has been linked to several adverse health outcomes both in adolescence and later in adulthood (Center for Disease Control and Prevention 2004). Active smoking in adolescents represents an important public health concern since adult smoking behaviours are usually established earlier in life. For example, in the United States between 1992 and 1995, 42% of ex- and current adult smokers reported initiating smoking before the age of 16, while 75% reported initiating smoking before the age of 19 (Gruber and Zinman 2000).

This study aims at contributing to a further understanding of socioeconomic variations in passive and active smoking amongst children and adolescents. The relationship between socioeconomic status, health risk behaviours such as smoking and health is well established. In adults, higher socioeconomic status (education, income or occupation) is associated with better health outcomes and explaining the association between health and socioeconomic status has been the focus of much research (some examples include Balia and Jones (2008), Contoyannis and Jones (2004), Vallejo-Torres and Morris (2010)). These studies have shown the existence of a strong and robust correlation between socioeconomic status and health risk behaviours, suggesting that the socioeconomic gradient in health can be explained by socioeconomic-related inequalities in health risk behaviours and lifestyle choices such as smoking, excessive alcohol consumption and lack of physical exercise (Balia and Jones 2008; Contoyannis and Jones 2004; Vallejo-Torres and Morris 2010). Furthermore, limited knowledge of the adverse health consequences of health risk behaviours

and lifestyle choices may provide further explanations of the link between socioeconomic status and health. For example Kenkel (1991) showed that higher years of schooling is associated with better knowledge of the relationship between lifestyle choices and health outcomes, thereby resulting in higher household allocative efficiency in the production of health (Kenkel 1991). Other recent studies have shown links between education, health knowledge and lifestyle choices (Cutler and Lleras-Muney 2010; Peretti-Watel et al. 2007). Cutler and Lleras-Muney (2010) showed that education increases cognitive ability which in turn improves health behaviours. Peretti-Watel et al. (2007) demonstrated that persons with less education are more likely to underestimate the health risk of smoking.

The relationship between child health and parental socioeconomic status has also been widely reported (recent examples include Cameron and Williams (2009), Condliffe and Link (2008) and Currie et al. (2007)). Grossman (1972, 2000), using a health capital model, describes how parental socioeconomic status can affect child health. Child health can be ‘produced’ using a set of health inputs, the choice of which is determined by the child’s parent, subject to a budget constraint (parental income) and parental preferences. Parental socioeconomic status can therefore affect child health directly or indirectly through its effect on the choice of the health input that go into child health production function. For example the effect of parental socioeconomic status on child health may arise directly because parents with lower income are unable to afford better quality healthcare or high nutritional food for the child, or indirectly, due to parental preferences for health risk behaviours, which in turn impact adversely on child health. On the other hand, parents with lower socioeconomic status may simply have different health beliefs that make them treat health inputs differently from parents with higher socioeconomic status (Currie 2009). Smoking is a major factor contributing to socioeconomic variations in adult health and children from lower socioeconomic backgrounds are likely to be exposed to environmental factors that increase both the probability of them initiating smoking or increase the probability of exposure to second-hand smoke. Therefore parental smoking behaviour may, at least in part, explain the socioeconomic gradient in child health. However recent evidence appears to suggest that the correlation between parental socioeconomic status and child health is not mediated through parental smoking (Frijters et al. 2011; Reinhold and Jürges 2011).

While the association between adult smoking behaviour and socioeconomic status has been widely studied, very few studies exist on the relationship between parental socioeconomic status and

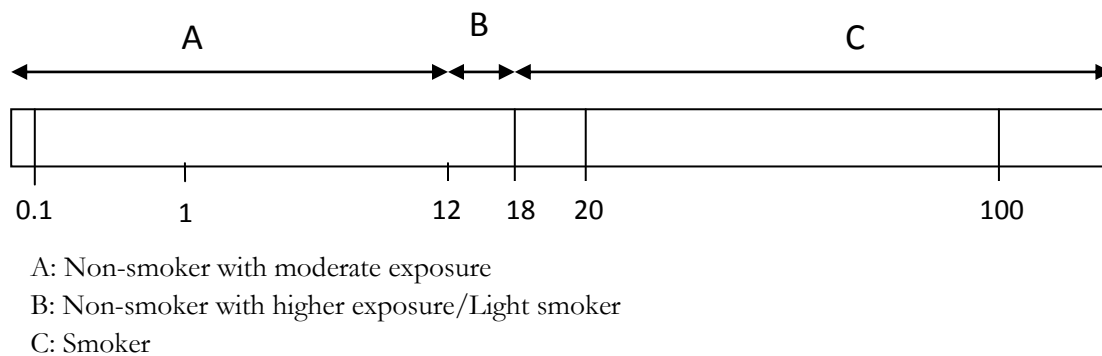
passive/active smoking in children and adolescents. In a recent study, Frijters et al. (2011) showed that household income is negatively associated with passive smoking (measured using saliva cotinine). Existing studies on the relationship between active smoking in adolescents and parental socioeconomic status have often reported conflicting findings (Blow et al. 2005; Edoka 2011; Glendinning et al. 1994; Soteriades and DiFranza 2003; Tuinstra et al. 1998). While some studies report a robust negative correlation between parental socioeconomic status and active smoking amongst adolescents (Edoka 2011; Soteriades and DiFranza 2003), other studies fail to find an association (Glendinning et al. 1994; Tuinstra et al. 1998) or the association disappears after controlling for parental smoking (Blow et al. 2005). Differences in the indicators of parental socioeconomic status (Currie et al. 2008; Gruber and Zinman 2001; Tyas and Pederson 1998), contextual differences of samples as well as the extent of measurement errors in adolescent self-reported smoking behaviour (Edoka 2011), may explain these divergent findings.

The unequal distribution of the determinants of passive and active smoking amongst individuals from different socioeconomic background may explain the social gradient of smoking in children and adolescents. For example, the attenuation or disappearance of the negative association between parental socioeconomic status and adolescent smoking, after controlling for parental smoking, suggest that parental smoking is an important mediator of the socioeconomic gradient in adolescent smoking (some examples include Blow et al. (2005) and Soteriades and DiFranza (2003)). Children from lower socioeconomic backgrounds are more likely to have parents or other family members or friends who smoke, are more likely to live in non-smoke free homes and are more likely to live in deprived neighbourhoods (Action on Smoking and Health (ASH) Research Report 2011). In addition to increasing the risks of exposure to second-hand smoke, these factors have been shown to increase the probability of active smoking amongst adolescents (Loureiro et al. 2010; Powell et al. 2005; Powell and Chaloupka 2005; Sims et al. 2010).

This study aims at furthering the understanding of the contributions of various determinants of smoking to differences in passive and active smoking between two groups of children and adolescents defined by parental socioeconomic status (high and low socioeconomic status). Saliva cotinine, a major metabolite of nicotine is used as a proxy for active and passive smoking. Cotinine is a biomarker of the extent of exposure to second-hand smoke and a quantitative indicator of active smoking (Environmental Protection Agency 1997; Jarvis et al. 2008). Cotinine levels greater than or equal to 12ng/ml identifies active smoking with high sensitivity (96.7%; Jarvis et al. 2008). However,

higher cut-points (18ng/ml) may be required due to an overlap between light smokers and non-smokers with high exposure to second-hand smoke. The advantages of using cotinine in this study are two-fold. First, because cotinine measurements are objective, they are less prone to measurement errors seen with self-reported smoking behaviours. Second, the entire distribution of cotinine can be decomposed, allowing the simultaneous identification of contributions made by each determinant to socioeconomic differences in both active and passive smoking. Figure 1 shows a diagrammatic representation of saliva cotinine cut-points for identifying active and passive smoking in the cotinine distribution (based on the findings of Jarvis et al. (2008)). In this study, we assume that the lower end of the log cotinine distribution is likely to comprise of non-smoking children/adolescents with moderate exposure to second-hand smoke while the top end of the distribution comprises of active smokers.

Figure 1- Saliva cotinine cut-points (ng/ml)



We use the 1997/98 Health Survey for England (HSE) which contains saliva cotinine measurements and apply decomposition methods to explain differences in the distribution of cotinine between children and adolescents from high and low socioeconomic backgrounds. Although more recent years of the HSE collected saliva specimens for cotinine assays, specimens were only collected for a small proportion of children. In addition to the nationally representative sample of children, a boost sample of children were surveyed in 1997 and saliva specimens collected, thus providing a larger sample of children with valid cotinine measurements in 1997 compared to more recent years. These decomposition methods, which were originally developed and applied in the labour economics literature (for example, in explaining gender, regional and inter-country

differences in wages, as well as in explaining changes in wage inequalities across time)¹, have now found wider application in other fields including health economics. These decomposition methods allow socioeconomic differences in the distribution of log cotinine to be decomposed into a part explained by differences in the distribution of characteristics (composition effect), and a part explained by differences in the impact of these characteristics (structural effect). Therefore, in addition to quantifying the extent to which the distribution of characteristics explain socioeconomic differences in passive and active smoking amongst children and adolescents, we are able to identify, conditional on having similar characteristics, the extent to which variations in the impact of these characteristics contribute to socioeconomic differences in smoking.

In the first instance, a mean-based decomposition approach (Blinder 1973; Oaxaca 1973) is used to decompose differences in mean log cotinine. Then, the empirical analysis is extended to decompose differences between quantiles of log cotinine (Firpo et al. 2009; Melly 2005). The decomposition of the entire distribution of log cotinine allows a deeper investigation into the extent to which various factors make greater or lesser contributions at different quantiles of the log cotinine distribution, indicative of passive or active smoking. For example, our results suggest that smoking within the home explains more of the socioeconomic difference at the lower end of the log cotinine distribution and less of the difference at the upper end of the distribution. Conversely, parental smoking explains more of the difference at the upper end of the log cotinine distribution compared to its contribution at the lower end of the distribution.

The rest of the paper is organized as follows: section 2 gives a description of the data and variables. An empirical framework motivating the choice of variables is also outlined in section 2. In section 3, we broadly define the parameters of interest in the decomposition analysis, outlining the conditions necessary for identification of these parameters and describe the estimation procedures. The results are presented and discussed in section 4 and section 5 concludes the paper.

¹ Fortin et al. (2011) provide an extensive review of the decomposition methods and the applications in labour economics.

2. Data, Variables and Empirical Framework

2.1 Data and Variables

This study uses the 1997/98 cross-section of the Health Survey for England (HSE). The HSE is a series of annual cross-sectional surveys which includes a nationally representative sample of households in England. Households are drawn from the Postcode Address file and all adults over the age of 16 years and a random selection of two children aged between 0-15 years living within selected households are interviewed.² In addition to individually self-completed questionnaires, each consenting household received a nurse visit during which objective measures of health were taken and saliva specimens collected for cotinine assay. Cotinine assay was performed using gas chromatography which detects cotinine levels as low as 0.1ng/ml. Cotinine is a metabolite of nicotine and with a half-life of about 16-20 hours, it can be detected in saliva specimens of regular or occasional smokers or in individuals exposed to second-hand smoke. Cotinine is generally accepted as a quantitative indicator of tobacco intake such that the extent of active or passive exposure to tobacco is reflected in the level of saliva cotinine detected (Environmental Protection Agency 1997; Jarvis et al. 2008).

In this study, two groups of children and adolescents are defined based on the social class of the household head which was assigned using the Registrars General's Social Class (RGSC) classification system. The RGSC classification system is based on six categories of occupation: professional (I), managerial/technical (II), non-manual skilled (IIIa), manual skilled (IIIb), partly skilled (IV), unskilled (V) and other (VI). Children and adolescents living in households where the head of the household had a professional or managerial/technical occupation were classified as the 'high social class' (HSC) group. While children and adolescents living in households where the head of the household belonged to categories IV, V or VI (partly skilled, unskilled, or any other occupation) were classified as the 'low social class' (LSC) group. Figure 2 shows distribution of log cotinine by social class. The log cotinine distribution in the LSC group lies to the right of the log cotinine distribution of the HSC group indicating higher levels of cotinine in children and adolescents in the LSC group compared to those in the HSC group at all quantiles.

² A full description of the survey design can be found in Prescott-Clarke (1998).

We make use of a wide range of demographic and socioeconomic characteristics available in the HSE. These characteristics were collected using household questionnaires (for household characteristics) and self-completed questionnaires. Characteristics of the child or adolescent include: age group (8-10/11-12/13/14-15years), gender (male/female) and ethnicity (white/non-white); household characteristics include: household location (rural/suburban/urban), non-smoke free homes (defined as whether smoking by members or non-members of the household was permitted within the home) and household income. Finally, parental characteristics include: parental smoking behaviour, highest academic qualification, marital status and age group³. These were obtained by linking parents' responses in the individual questionnaires to each child.

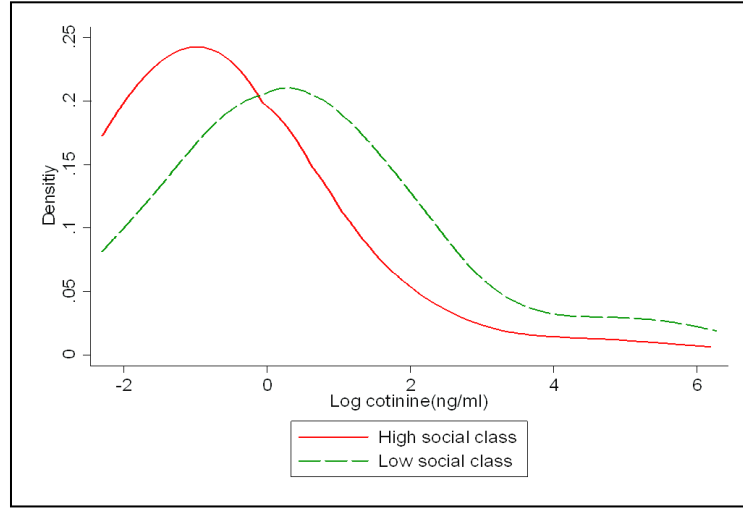
Table 1 shows a comparison of average characteristics of both groups. On average, children and adolescents in the HSC group, have more educated and older parents compared to those in the LSC group (Table 1). In addition, household income is significantly higher in the HSC group compared to the LSC group (Table 1). A significantly higher proportion of mothers (25% vs. 11%) and fathers (14% vs. 9%) smoke within the LSC group compared to the HSC group (Table 1). Similarly, a higher proportion of families in the LSC group permit smoking within homes compared to those in the HSC group (59% vs. 24%; Table 1).

In 1997, the number of children surveyed was boosted by surveying more households⁴. Although household questionnaires were completed by the head of the household, adults (including parents) from the boost sample were not surveyed. Parent information is therefore missing for all those in the boost sample as well as for those living in a single parent household or in a two-parent household but one parent was absent during the interview. In this study, we include dummy variables for missing information on parental smoking behaviour, highest academic qualification, marital status and age group. The final sample consists of 2355 children and adolescents with 1397 individuals from the HSC group and 958 individuals from the LSC group.

³ A full description of all variables including parents' characteristics is provided in Table A1 of the Appendix.

⁴ Approximately 45% of our final sample comprise of children from the boost sample.

Figure 2- Log cotinine distribution by social class



2.2 Empirical Framework

The decomposition methods rely on modelling log cotinine as a function of a set of covariates. The conceptual framework for the model adopted in this study is based on Rawls' principle of Justice (1971): *"each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others"*. This principle has been adapted to measuring socioeconomic inequality in health (Bommier and Stecklov 2002). The Bommier & Stecklov (2002) approach is based on the notion that all individuals should have equal opportunities to achieve their health potential and inequalities in health arise due to inequalities in the distribution of unobserved natural factors (or "luck"). These natural factors generally reflect circumstances that are largely beyond the individual's control. The empirical analysis described in this paper focuses on variations in the log cotinine distribution of a young cohort (8-15 year olds) that can be explained by variations in circumstances such as family/parental socioeconomic background that are beyond the individuals' control. These circumstances form the social environment which either reduces the perceived cost of smoking (for active smokers) or increases exposure to second-hand smoke (for non-smokers). Therefore in this study, log cotinine (*COT*) is modelled as a function of a set of covariates that reflect these circumstances.

$$COT_g = \theta_g(X_g, \varepsilon_g)$$

where g denotes HSC or LSC group membership; X is a vector of characteristics including demographic characteristics of the child/adolescent (age, gender and ethnicity) and other characteristics that define the social environment of the child including household characteristics (smoking within homes, home location and household income) and parental characteristics (parents' age, academic qualification and current smoking status); ε is a vector of unobservable characteristics.

These characteristics have been shown to influence active smoking participation as well as passive smoking amongst children and adolescents. For example parental smoking has been shown to increase the probability of active and passive smoking in children and adolescents (Frijters et al. 2011; Loureiro et al. 2010) while parental income has been shown to be negatively correlated with both passive and active smoking (Edoka 2011; Frijters et al. 2011; Soteriades and DiFranza 2003). Since the choice of parents or parents' lifestyle choices are beyond the control of the child, the unequal distribution of these characteristics reflects inequality of opportunity in active and passive smoking amongst children and adolescents.

3. The Decomposition Methods

To decompose differences in log cotinine between children and adolescents from the HSC and LSC groups, we use three approaches: a mean-based decomposition approach, the Oaxaca-Blinder (OB) decomposition approach (Blinder 1973; Oaxaca 1973), and two approaches which allow the decomposition of differences in distributional statistics other than the mean, the quantile regression (QR) decomposition method (Machado and Mata 2005; Melly 2005) and the recentered influence function regression (RIFR) decomposition method (Firpo et al. 2009; Fortin et al. 2011). These methods allow socioeconomic differences in log cotinine to be decomposed into a part attributable to group differences in the distribution of characteristics (composition effect) and a part attributable to group differences in coefficients (structural effect). In the following sub-sections the restriction assumptions required for identification of the composition and structural effects are outlined formally⁵.

⁵ Fortin et al. (2011) provides an extensive discussion of these assumptions. Only assumptions that apply to the decomposition methods applied in this study are highlighted here.

3.1 Identification

The decomposition methods rely on estimating unconditional counterfactual distributions of the outcome variable. For two mutually exclusive groups, HSC (H) and LSC (L) groups, we observe log cotinine distributions for each group (COT_H and COT_L respectively). The unconditional counterfactual distribution is constructed to simulate what the log cotinine distribution of individuals in the HSC group would be if they belonged to the LSC group, or conversely, what the log cotinine distribution of individuals in LSC group would have been if they belonged to HSC group⁶. To construct these counterfactual distributions, the decomposition methods explore the relationship between log cotinine and a set of observed and unobserved characteristics.

$$COT_g = \theta_g(X_g, \varepsilon_g), \quad g \in \{H, L\}$$

$$\Delta_{COT} = COT_H - COT_L = [\theta_H(X_H, \varepsilon_H)] - [\theta_L(X_L, \varepsilon_L)] \quad (1)$$

where X_H and X_L are vectors of observable characteristics, θ_H and θ_L are the functional forms of the log cotinine equation and ε_H and ε_L are vectors of unobservable characteristics for the HSC and LSC groups respectively.

The unconditional counterfactual distribution of log cotinine is generated by integrating the conditional distribution of log cotinine given a set of covariates in one group over the marginal distribution of covariates in the other group. If the unconditional distribution of log cotinine of each group is given by:

$$F_{COT_g}(cot) = \int F_{COT_g|X}(cot|X=x) \cdot dF_{X_g}(x), \quad g \in \{H, L\} \quad (2)$$

(where $F_{COT_g|X}(cot|X=x)$ is the conditional distribution of log cotinine and $F_{X_g}(x)$ is the marginal distribution of X), the unconditional counterfactual distribution can be generated by either replacing the conditional distribution of log cotinine in one group with the corresponding conditional distribution of the other group or by substituting marginal distribution of covariates. In this study we use the LSC as the reference group and construct a counterfactual distribution of log cotinine, $F_{COT_H}^C$, by replacing $F_{COT_H|X}(cot|X=x)$ with $F_{COT_L|X}(cot|X=x)$ in equation (2) when $g = H$:

⁶ In this study, we construct the former unconditional counterfactual distribution.

$$F_{COT_H}^C(cot) = \int F_{COT_L|X}(cot|X=x) \cdot dF_{X_H}(x) \quad (3)$$

The unconditional counterfactual distribution $F_{COT_H}^C(cot)$ represents the distribution of log cotinine that would have prevailed in the HSC group if the distribution of characteristics were similar to the LSC group.

From equation (1), it follows that the total difference in log cotinine between the two groups can be written as:

$$\Delta_{COT} = \Delta_{\theta} + \Delta_X + \Delta_{\varepsilon}$$

where Δ_{θ} captures group differences in the θ functions (**A**), Δ_X captures group differences in the distribution of observable characteristics (**B**), and Δ_{ε} captures group differences in the distribution of unobservable characteristics (**C**). In constructing the unconditional counterfactual distribution $F_{COT_H}^C$, replacing the conditional distribution of log cotinine of the HSC group with that of the LSC group replaces both θ and the conditional distribution of ε . Therefore group difference in θ will be confounded by group differences in the distribution of ε ⁷. To separate the group differences ε from the group differences in θ (and X), an identification restriction is imposed on the distribution of ε . Under the conditional independence/ignorability assumption, the conditional distribution of ε given X is the same for both groups and is independent of group membership ($M_g \perp \varepsilon | X, g = H, L$).

In addition to the conditional independence/ignorability assumption, the overlapping support assumption is imposed to rule out cases where observable and unobservable characteristics in the cotinine structural model are different for both groups. This assumption also ensures that no single characteristic can identify membership into any one group (Fortin et al., 2011).

Under these two assumptions, the total difference in log cotinine between the two groups, Δ_{COT}^v (where v represents a distributional statistics of log cotinine such as the mean or quantiles), can be separated and identified in an aggregate decomposition as:

$$\Delta_{COT}^v = \Delta_{\theta}^v + \Delta_X^v$$

⁷ The conditional distribution $F_{COT_g|X}(cot|X=x)$ depends on the distribution of ε as follows (Fortin et al. 2011): $F_{COT_g|X}(cot|X=x) = \Pr(\varepsilon \leq \theta_g^{-1}(X, cot) | X=x)$

where $\Delta_{\theta}^v = v(F_{COT_H} - F_{COT_H}^C)$, a part explained by group differences in the log cotinine structure (structural effect) and $\Delta_X^v = v(F_{COT_H}^C - F_{COT_L})$, a part explained by group differences in the distribution of the observed characteristics (composition effect).

The structural and composition effects can further be decomposed into contributions attributable to each characteristic (detailed decomposition). For the detailed decomposition, additional assumptions are required for the identification of the contribution of each characteristic. These assumptions are specific to the decomposition method and are discussed further in the estimation procedure described for each method in the following sub-section.

3.2 Estimation procedures

3.2.1 Oaxaca-Blinder (OB) decomposition method

The mean-based OB decomposition method is based on the assumption that the relationship between log cotinine and a set of characteristics is linear and additive:

$$COT_g = X_g \beta_g + \vartheta_g, \quad E(\vartheta_g) = 0, \quad g \in \{H, L\}$$

where X is a vector of observable characteristics, β is a vector of the slope parameters including the intercept and ϑ is the error term. Given that $E(\vartheta_g) = 0$, the total difference in mean log cotinine, Δ_{COT}^{μ} or $\mu(F_{COT_H} - F_{COT_L})$, can be decomposed as follows:

$$\Delta_{COT}^{\mu} = E(COT_H) - E(COT_L) = \underbrace{E(X_H)\beta_H - E(X_H)\beta_L}_A + \underbrace{E(X_H)\beta_L - E(X_L)\beta_L}_B \quad (4)^8$$

where $E(X_H)\beta_L$ is the unconditional counterfactual distribution of log cotinine at the mean⁹. Rearranging equation (5), we obtain:

$$\Delta_{COT}^{\mu} = (E(X_H)[\beta_H - \beta_L]) + ([E(X_H) - E(X_L)]\beta_L) \quad (5)$$

Equation (5) is a special case of a more general decomposition. Following Jones and Kelley (1984), equation (4) can be rearranged to obtain:

⁸ These two terms, A and B, are analogous to components A and B described in section 3.1

⁹ The counterfactual distribution is generated as described in equation (3) at the sample means $\mu(F_{COT_H}^C) \rightarrow E(COT_H^C) = E(X_H) \beta_L$

$$\Delta_{COT}^{\mu} = E(X_L)[\beta_H - \beta_L] + [E(X_H) - E(X_L)]\beta_L + [E(X_H) - E(X_L)][\beta_H - \beta_L] \quad (6)$$

Replacing $E(X_H)$ and $E(X_L)$ by their sample means \bar{X}_H and \bar{X}_L , as well as β_H and β_L by their ordinary least square (OLS) estimates, $\hat{\beta}_H$ and $\hat{\beta}_L$, equation (6) can be written as:

$$\hat{\Delta}_{COT}^{\mu} = \underbrace{\bar{X}_L (\hat{\beta}_H - \hat{\beta}_L)}_{\hat{\Delta}_{\theta}^{\mu}} + \underbrace{(\bar{X}_H - \bar{X}_L) \hat{\beta}_L}_{\hat{\Delta}_X^{\mu}} + \underbrace{(\bar{X}_H - \bar{X}_L)(\hat{\beta}_H - \hat{\beta}_L)}_{\hat{\Delta}_I^{\mu}} \quad (7)^{10}$$

The first term, $\hat{\Delta}_{\theta}^{\mu}$, represents contributions to the total difference in log cotinine between the HSC and LSC groups attributable to group differences in the coefficients including the intercept. The second term, $\hat{\Delta}_X^{\mu}$, represents contributions attributable to group differences in the distribution of mean characteristics. There is no clear interpretation for the third term, $\hat{\Delta}_I^{\mu}$, because it represents an interaction between group differences in characteristics and coefficients as well as differences in unobserved characteristics.

An attractive feature of the OB decomposition method is that it provides a way of performing a detailed decomposition of the composition and structural effects into contributions attributable to each covariate. This is possible because of the additive linearity assumption. The total composition and structural effect is simply the sum of the contribution of individual covariates:

$$\hat{\Delta}_X^{\mu} = \sum_{k=1}^P (\bar{X}_{Hk} - \bar{X}_{Lk}) \hat{\beta}_{Lk} \quad (8)$$

and

$$\hat{\Delta}_{\theta}^{\mu} = (\hat{\beta}_{H0} - \hat{\beta}_{L0}) + \sum_{k=1}^P (\hat{\beta}_{Hk} - \hat{\beta}_{Lk}) \bar{X}_{Lk} \quad (9)$$

where k represents the k th covariate and $\hat{\beta}_{H0}$ and $\hat{\beta}_{L0}$ are the estimated intercept coefficients of the HSC and LSC group respectively.

For categorical variables, the result of the detailed decomposition is not invariant to the choice of the base or omitted category. Changing the base category alters the contributions of the

¹⁰ In this study, these components are estimated using the ‘Oaxaca’ STATA command and the three-fold option (Jann 2008).

other categories as well as the contribution of the categorical variable as a whole¹¹. This is accounted for by applying a normalization approach (Yun 2005). This approach imposes a normalization on the coefficients of the categories by restricting the coefficients of the first category to be equal to the unweighted average of the coefficients on the other categories. In addition, the sum of the coefficients are restricted to sum up to zero (Yun 2005).

3.2.2 Quantile Regression (QR) decomposition method

The QR decomposition method (Melly 2005)¹² goes beyond the mean and decomposes differences between the two groups across the entire distribution of log cotinine. It allows for the identification of the total structural and composition effect at different quantiles. The unconditional counterfactual distribution ($F_{COT_H}^C$) is generated as defined in equation (3) by integrating the conditional distribution of cotinine in the LSC group ($F_{COT_L|X}$) over the marginal distribution of covariates in the HSC group (F_{X_H}). But for quantiles, the conditional distribution of cotinine in the LSC group is given as:

$$F_{COT_L|X}(cot|X = x) = \int_0^1 1[F_{COT_L|X}^{-1}(\tau|X = x) \leq q_\tau] \cdot d\tau$$

where q_τ is the τ th quantile of the unconditional distribution of log cotinine. The counterfactual distribution of log cotinine can be expressed as:

$$F_{COT_H}^C(q_\tau) = \int \left[\int_0^1 1[F_{COT_L|X}^{-1}(\tau|X = x) \leq q_\tau] \cdot d\tau \right] \cdot dF_{X_H}(x) \quad (10)$$

Replacing $F_{COT_L|X}^{-1}(\tau|X = x)$ in equation (10) with its consistent conditional quantile regression estimator, $X_L \hat{\beta}_L(\tau)$,¹³ and inverting the distribution function $F_{COT_H}^C$, the unconditional quantiles of the counterfactual distribution of log cotinine can be recovered.

The decomposition of the total difference at the τ th quantile, Δ_{COT}^τ or $\tau(F_{COT_H} - F_{COT_L})$, can then be performed as follows:

¹¹ This mainly affects results of the detailed decomposition of the structural effect. The contribution to the composition effect is unaffected by the choice of the omitted category.

¹² The QR decomposition method was first proposed by Mata and Machado (2005) and is similar to Melly (2005). We use the Melly's (2005) STATA command (rqdeco3) because it is less computationally demanding.

¹³ The conditional quantile function is estimated using quantile regression $F_{COT_L|X}^{-1}(\tau|X = x) = X_L \beta_L(\tau), \forall \tau \in (0,1)$.

$$\begin{aligned} \hat{q}_\tau(\hat{\beta}_H, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_L) \\ = \underbrace{[\hat{q}_\tau(\hat{\beta}_H, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_H)]}_{\hat{\Delta}_\theta^\tau} + \underbrace{[\hat{q}_\tau(\hat{\beta}_L, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_L)]}_{\hat{\Delta}_X^\tau} \end{aligned} \quad (11)$$

where $\hat{q}_\tau(\hat{\beta}_L, X_H)$ is the unconditional counterfactual distribution at the τ th quantile, $\hat{q}_\tau(\hat{\beta}_H, X_H)$ and $\hat{q}_\tau(\hat{\beta}_L, X_L)$ are the unconditional distribution of log cotinine in the HSC and LSC groups, respectively.

Melly (2005) proposes a way of separating the effects of the coefficients from the effects of the residuals by defining an $N \times 1$ vector, $\hat{\beta}_{HL}$, with its n th component defined as: $\hat{\beta}_{HL}(\tau_n) = \hat{\beta}_H(0.5) - \hat{\beta}_L(0.5) + \hat{\beta}_L(\tau_n)$, where $\hat{\beta}_H(0.5)$ and $\hat{\beta}_L(0.5)$ are the coefficient vectors of the median regressions for the HSC and LSC groups, respectively.

The overall decomposition in equation (11) can then be expressed as:

$$\begin{aligned} \hat{\Delta}_{COT}^\tau &= \hat{q}_\tau(\hat{\beta}_H, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_L) \\ &= \underbrace{[\hat{q}_\tau(\hat{\beta}_{HL}, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_H)]}_{\hat{\Delta}_\theta^\tau} + \underbrace{[\hat{q}_\tau(\hat{\beta}_L, X_H) - \hat{q}_\tau(\hat{\beta}_L, X_L)]}_{\hat{\Delta}_X^\tau} \\ &\quad + \underbrace{[\hat{q}_\tau(\hat{\beta}_H, X_H) - \hat{q}_\tau(\hat{\beta}_{HL}, X_H)]}_{\hat{\Delta}_R^\tau} \end{aligned}$$

where $\hat{q}_\tau(\hat{\beta}_{HL}, X_H)$ is the distribution of log cotinine at the τ th quantile that would have prevailed if median coefficients had been similar to those of the HSC group but the residuals had been similar to those of the LSC group. $\hat{\Delta}_\theta^\tau$ represents contributions attributable to group differences in (median) coefficients at the τ th quantile, $\hat{\Delta}_R^\tau$ represents contributions attributable to group differences in residuals and $\hat{\Delta}_X^\tau$ represents contributions attributable to group differences in the distribution of characteristics.

3.2.3 Recentered Influence Function regression (RIFR) decomposition method

One major limitation of the QR approach is that it cannot be extended to a detailed decomposition. To assess the contributions of individual covariates at different quantiles, we apply the RIFR decomposition method (Firpo et al. 2009). The RIFR decomposition approach, which is based on an unconditional quantile estimator, is analogous to the mean-based OB decomposition

method. The RIFR¹⁴ provides a way of estimating the marginal effect of a vector of covariates (X) on an unconditional distributional statistic of an outcome variable. The marginal effect of X is estimated by regressing a function of the outcome variable, known as the recentered influence function (RIF), on X .

In this study the RIF of log cotinine at each quantile is estimated directly from the data by first computing the sample quantile q and then estimating the density at that quantile using kernel density methods. An estimate of the RIF of each observation is then obtained using the following equation:

$$RIF(cot; q_\tau) = q_\tau + \frac{\tau - 1[cot \leq q_\tau]}{f_{cot}(q_\tau)} \quad (12)$$

where q_τ is the τ th quantile of log cotinine and $f_{cot}(q_\tau)$ is the unconditional density of log cotinine at the τ th quantile and $1[cot \leq q_\tau]$ is an indicator function for whether the outcome variable is smaller or equal to the τ th quantile. At each quantile, the coefficients on X for groups H and L are then estimated by regressing the RIF on X ¹⁵:

$$q_{g,\tau} = E_X \left[E \left[\widehat{RIF}(cot_g; q_{g,\tau}) | X_g \right] \right] = E[X_g] \hat{\delta}_{g,\tau}, \quad g \in \{H, L\} \quad (13)$$

where $q_{g,\tau}$ is the unconditional τ th quantile of log cotinine for group $g \in \{H, L\}$ and $\hat{\delta}_{g,\tau}$ is the coefficient of the unconditional quantile regression which captures the marginal effect of a change in distribution of X on the unconditional quantile of log cotinine. Equation (13) is analogous to the basis of the OB decomposition at the mean. Thus, by applying the same logic, the difference in log cotinine between the two groups at the τ th quantile of log cotinine can be decomposed as follows:

$$\begin{aligned} \hat{\Delta}_{COT}^\tau &= [\widehat{RIF}(cot_H, q_{H,\tau})] - [\widehat{RIF}(cot_L, q_{L,\tau})] \\ \hat{\Delta}_{COT}^\tau &= \underbrace{\bar{X}_L (\hat{\delta}_{H,\tau} - \hat{\delta}_{L,\tau})}_{\hat{\Delta}_\theta^\tau} + \underbrace{(\bar{X}_H - \bar{X}_L) \hat{\delta}_{L,\tau}}_{\hat{\Delta}_X^\tau} + \underbrace{(\bar{X}_H - \bar{X}_L)(\hat{\delta}_{H,\tau} - \hat{\delta}_{L,\tau})}_{\hat{\Delta}_f^\tau} \end{aligned}$$

¹⁴ Firpo et al. (2009) describe the RIFR as an unconditional quantile regression, distinct from the conditional quantile regression, because it estimates the marginal effect of X on the *unconditional* quantile of log cotinine.

¹⁵ This can be performed using the STATA 'rifreg' command which is available for download as an RIF-regression STATA ado file from Firpo et al. (2009): <http://faculty.arts.ubc.ca/nfortin/datahead.html>.

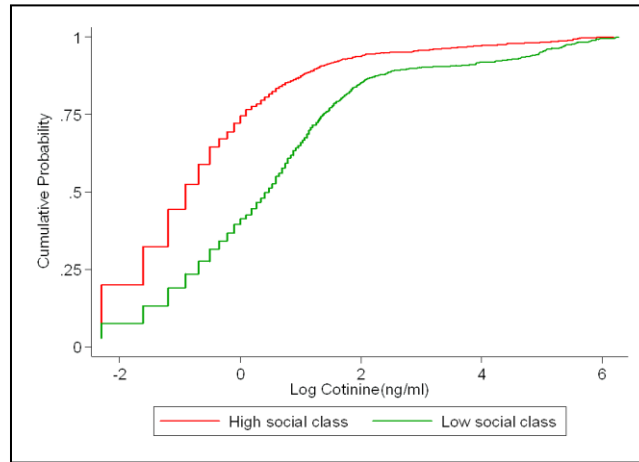
Similarly, the composition and structural effects can be further decomposed into contributions of each covariate at the τ th quantile in a detailed decomposition similar to equations (8) and (9).

4. Results

4.1 Aggregate Decomposition

Children and adolescents in the LSC group are more likely to be exposed to social and environmental factors that either increase the probability of them becoming active smokers or increase the risk of exposure to second-hand smoke. On average, log cotinine in the LSC group is significantly higher than the HSC group (0.612 vs. -0.639; Table 2), by approximately 1.251 points. Similarly, at all quantiles, log cotinine is higher in the LSC group compared to the HSC group, with the gap between the two groups increasing at higher quantiles (Table 2). This is depicted graphically in Figure 3 which shows a widening of the gap between the two groups moving up the log cotinine distribution.

Figure 3-Cumulative distribution of log cotinine by social class

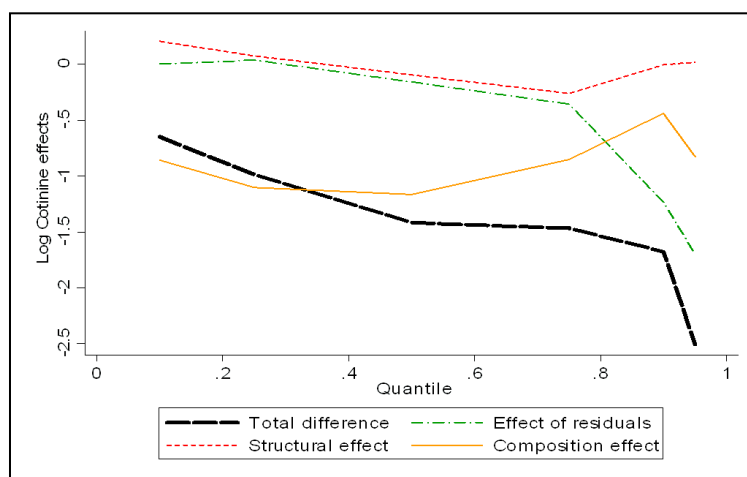


The results of the aggregate decomposition analysis are shown in Table 3. The OB decomposition shows that differences in mean characteristics account for a large proportion of the total difference between the two groups. If mean characteristics of the HSC group had been distributed similar to those of the LSC group, the total difference in average log cotinine between

both groups would decrease by approximately 1.115 points (upper panel, Table 3). Therefore approximately 89% of the total difference in average log cotinine is explained by differences in the distribution of characteristics (composition effect). However, the composition effect does not entirely account for the total difference and approximately 34% of the total difference is attributable to group differences in coefficients (structural effect). On the other hand, the interaction effect is positive. However, the interpretation of this effect is not unambiguous since it captures not only group differences in unobserved characteristics but also the interaction between group differences in characteristics and coefficients.

Table 3 (middle and lower panels) shows the results of the aggregate decomposition at different quartiles. Similar to the mean, the difference in log cotinine attributable to differences in characteristics explains a larger proportion of the total difference across all three quartiles, compared to the structural effect. In addition, the composition effect is greatest at lower quartiles compared to higher quartiles. The difference in log cotinine attributable to differences in characteristics, coefficients and residuals is depicted graphically in Figure 4. In Figure 4, from the lowest quantile up to the median, the ‘composition effect line’ lies close to, and follows the same direction as the ‘total difference line’. This implies that up to the median, differences in the distribution of characteristics explain a large proportion of the difference between the two groups. From the median, the two lines diverge implying that the composition effect explains less of the total difference between the two groups. Since lower quantiles of the log cotinine distribution are likely to comprise of passive smokers and higher quantiles, of active smokers, this result suggest that the distribution of observed characteristics explains more of the socioeconomic differences in passive smoking compared to active smoking. Interestingly this trend corresponds to an increasing contribution of the residuals to the total difference at higher quartiles. Unlike the interaction effect in the OB and RIFR decomposition methods, the residual effect in the QR decomposition can be interpreted as the extent to which differences in residuals contribute to the total difference in log cotinine (Melly 2005). At the third quartile, the residuals account for approximately 24% of the total difference. Therefore, the contribution of the residuals may reflect group differences in unobserved characteristics such as attitude towards risk or rate of time preference which may be more important in explaining socioeconomic differences in active smoking (compared to passive smoking) amongst children and adolescents.

Figure 4-QR decomposition (Melly, 2005)



Although it has been recently disputed¹⁶, some authors argue that the socioeconomic gradient in health risk behaviours can be explained by differences in the degree of risk aversion and rates of time preferences (Becker and Mulligan 1997; Leigh 1986). Adults with lower socioeconomic status are more likely to underestimate the potential health hazards of smoking (Peretti-Watel et al. 2007), are more likely to be present-oriented or have less incentive to invest in future health benefits and thus have higher discount rates of time preference in comparison to those with higher socioeconomic status (Becker and Mulligan 1997; Leigh 1986). The existence of a strong intergenerational transmission of the willingness to take risk (including health risk) and rates of time preferences (Breuer et al. 2011; Dohmen et al. 2008), implies that children in the LSC group may be more likely to adopt their parents' rate of time preference or attitude towards risk. In addition to reducing the perceived future cost of engaging in health risk behaviours, the social environment of a child may directly influence consumption preferences, with children emulating the consumption preference of their parents.

¹⁶ Some examples include Cutler and Lleras-Muney (2010) and Khwaja et al. (2007).

4.2 Detailed Decomposition

4.2.1 Composition effects

The results of the detailed decomposition provide more insight into the contributions of individual covariates to the composition and structural effects. Table 4 shows the results of the detailed decomposition. At the mean, the composition effect is driven mainly by difference in the distribution of homes within which smoking is permitted, household income, fathers' education and mothers' smoking status.

Smoking within the home makes the largest contribution to the overall composition effect, accounting for approximately 44% of the total composition effect at the mean. This is unsurprising, given that on average, a larger proportion of children and adolescents in the LSC group live in non-smoke free homes in comparison to those in the HSC group (59% vs. 24%; Table 1). The resulting higher levels of saliva cotinine in the LSC group may either be as a direct consequence of higher exposure to second-hand smoke or indirectly through less discouragement of experimentation and/or active smoking participation. The detailed decomposition of the entire log cotinine distribution sheds more light on the contributions of smoking within the home to the composition effect at different quartiles. Similar to the difference in mean, differences in the distribution of non-smoke free homes make the highest contribution to the composition effect at all three quartiles, compared to the contributions of other covariates. The composition effect of non-smoke free homes varies at different quartiles and a distinct pattern is observed. At the first quartile, differences in the distribution of non-smoke free homes account for approximately 57% of the total composition effect. This decreases to 45% at the median and approximately 29% at the last quartile. This suggests that smoking within homes explains more of the difference in passive smoking compared to its contribution to differences in active smoking.

At the mean, differences in the distribution of household income account for approximately 21% of the total composition effect. In addition, at the first and last quartile, although statistically insignificant, differences in the distribution of household income account for approximately 12% and 19% of the total composition, respectively. At the median, the contribution of income is statistically significant and accounts for approximately 18% of the total composition effect. Interestingly, our results suggest that household income explains more of the socioeconomic difference in active smoking compared to passive smoking. A possible explanation for the

contribution of household income to the composition effect at the third quartile may be due to income-related differences in parental smoking. Parental smoking has been shown to be an important mediator of the (parental) socioeconomic gradient in active smoking amongst adolescents (Blow et al. 2005; Soteriades and DiFranza 2003). Thus, children and adolescents in the LSC group living in low-income households are more likely to have parents who smoke, which in turn increases the probability of active smoking participation in children/adolescents. In addition, household income is negatively correlated with passive smoking within the home (Frijters et al. 2011). Therefore, the income effect at the first quartile can be explained by income-related differences in the exposure to second-hand smoke.

Results on the contribution of mothers' smoking status to the composition effect support these arguments. Differences in the distribution of mothers' smoking status make statistically significant contributions to the composition effect accounting for approximately 16% of the total composition effect at the mean. Differences in the distribution of mothers smoking equally makes statistically significant contributions at all three quartiles of the log cotinine distribution with the highest contribution observed at the last quartile (23% compared to 7% at the first quartile). This pattern (as well as that observed with household income) is in direct contrast to contributions of non-smoke free homes, where smoking within homes makes the highest contribution to the composition effect at the lowest quartile, suggesting that parental smoking (and household income) explains more of the socioeconomic differences in active smoking compared to passive smoking. Since children/adolescents whose parents smoke are more likely to become smokers themselves (Loureiro et al. 2010; Powell and Chaloupka 2005), and that we observe a higher proportion of mothers who smoke in the LSC group compared to the HSC group, the contribution of mothers' smoking at the highest quartile suggests a higher proportion of child/adolescent active smokers within the LSC group compared to the HSC group. The contribution of mothers' smoking behaviour to the composition effect at the first quartile suggests a higher exposure of those in the LSC group to second-hand smoke.

Father's education makes statistically significant contributions to the composition effect at the mean. Similarly, the composition effect of mothers' education is negative at the mean and at all quartiles but is statistically significant only at the first quartile, accounting for 15% of the total difference in log cotinine at the first quartile. The composition effect of parental education is likely to reflect group differences in the distribution of parental smoking which in turn could result in higher levels of exposure to second-hand smoke and active smoking amongst those in the LSC

group. In adults, the link between years of schooling/education, health knowledge and lifestyle choices including smoking behaviours are well established (Cutler and Lleras-Muney 2010; Kenkel 1991; Peretti-Watel et al. 2007). Therefore, given the potential intergenerational transmission of consumption preferences, adolescents with less educated parents who smoke are likely to be active smokers themselves.

4.2.2 Structural effects

Results of the detailed decomposition provide further interesting insights into the contributions of individual covariates to the total structural effects. At the mean and median, father's education makes statistically significant contributions to the total structural effect. This suggests a differential impact of father's education across socioeconomic groups. More insight into group differences in the impact of father's education is shown in Table A2-A5 of the Appendix. For example, in the HSC group, for children/adolescents whose fathers have no qualification, average log cotinine is approximately 11% higher compared to those whose fathers have a university degree (Table A2). The corresponding estimate in the LSC group is approximately 43% (Table A2). Similar results are observed across all categories of father's education except for the NVQ3 qualification/equivalent category (Table A2). The result of the decomposition analysis suggests that group differences in the coefficients of father's education contribute significantly to the difference in log cotinine observed between both groups. The definition of social class in this study is based on parent's occupation (the RGSC classification system). Thus, for fathers with similar levels of education, those in the HSC group are likely have a higher occupational status, earn more and perhaps live in more affluent neighbourhoods compared to fathers in the LSC group. For example, for fathers in the HSC group with lower levels of education, the impact of father's education on child's log cotinine levels may be mitigated by other favourable social circumstances that limit exposure to second-hand smoke or discourage active smoking amongst children and adolescents in the HSC group.

At the lowest quartile and at the median, smoking within homes contributes significantly to the total structural effect. This suggests that smoking within homes also exerts a differential impact on log cotinine in children/adolescents from both groups. Although smoking within homes

increases log cotinine in both groups, this effect is less in the HSC group¹⁷. This suggests possible group differences in the behaviours of parents. For example parents in the HSC group may adopt avoidance behaviours which limit the child's exposure to tobacco smoke, such as restricting smoking within the home to specific rooms or to specific times when the child is unlikely to be present. In our sample, in households where either parent smokes, 94% of parents in the LSC group permit smoking within the home compared to 72% in the HSC group. This proportion increases to 100% and 87% respectively, when both parents smoke. The contribution of smoking within homes at the third quartile also suggests differential impact of smoking within homes on smoking behaviours of children/adolescents. However unlike at lower quartiles, the impact of smoking within homes is greater in the HSC group compared to the LSC group at the third quartile.

Further evidence of possible differences in parental behaviour is demonstrated in the differential impact of parental smoking particularly at the third quartile. The structural effect of father's smoking is negative at the mean, median and last quartile but is statistically significant only at the last quartile. Similarly, the structural effect of mothers' smoking is negative at the mean and across all quartiles (although statistically insignificant). Given the strong correlation between education, health knowledge and health risk perception, (Cutler and Lleras-Muney 2010; Kenkel 1991; Peretti-Watel et al. 2007), parents in the HSC group with higher levels of education are likely to be more knowledgeable of the adverse health consequences of their lifestyle choices on child health, thus explaining group differences in parental attitudes towards protecting their child from the harmful effects of smoking.

Taken together, these results suggests that not only do a smaller proportion of wealthier and more educated parents engage in health risk behaviours such as smoking and smoking within the home, parents in the HSC group who engage in these health risk behaviours adopt other compensatory or avoidance behaviours that either discourage their child from active smoking participation or limit their child's exposure to second-hand smoke.

5. CONCLUSION

Socioeconomic inequality in adolescent smoking, has received very little attention in the economics literature. This study sheds more light on the factors that contribute to socioeconomic

¹⁷ See Tables A2-A4 in the Appendix

differences in both passive and active smoking amongst children and adolescents aged 8 to 15 years old. In the first instance, a mean-based decomposition method is applied to assess contributions of various characteristics or determinants of smoking to differences in average log cotinine between two groups of children and adolescents defined by parental socioeconomic status. The analysis is then extended to decompose the entire distribution of log cotinine. The lower end of the log cotinine distribution is likely to comprise of non-smokers exposed to second-hand smoke (passive smokers) while the upper end of the distribution is likely to comprise of active smokers. Therefore, the extension of the decomposition analysis to different quartiles provides useful insight into the distinct roles of these determinants in explaining socioeconomic differences in active and passive smoking amongst children and adolescents. Log cotinine is modelled as a function of a set of characteristics, which are considered to explain within- and between-group variations in log cotinine. Differences in log cotinine between the two groups are decomposed into a part explained by differences in the distribution of characteristics (composition effect) and a part explained by differences in the impact of these characteristics (structural effect).

The results show that the composition effect accounts for a large proportion of the total difference between the two groups both at the mean and at different quartiles of the log cotinine distribution. At the lower end of the log cotinine distribution (indicative of passive smoking), the composition effect explains more of the difference in log cotinine and less of the difference at the upper end of the distribution (indicative of active smoking). Conversely, the distribution of unobserved characteristics are more important in explaining socioeconomic differences in active smoking and less of the difference in passive smoking.

Characteristics making the largest contribution to the total composition effect are smoking within the home, father's education, household income and mother's smoking. These characteristics make different contributions at different quartiles of the log cotinine distribution, suggesting distinct roles in explaining socioeconomic differences in passive versus active smoking. Smoking within the home explains more of the socioeconomic difference in passive smoking compared to the socioeconomic difference in active smoking while household income and parental smoking behaviour explain more of the socioeconomic difference in active smoking compared to passive smoking. Given that smoking in adults follows a socioeconomic gradient, children and adolescents in the LSC group are more likely to have parents who smoke or live in homes where smoking is permitted within the home. These factors are likely to increase the probability of passive or active smoking in children/adolescents, thus explaining, at least in part, the higher levels of log cotinine

observed in the LSC group.

Although the composition effect explains a large part of the difference between the two groups, it does not tell the whole story. Conditional on having similar characteristics, we observe differential impacts of these characteristics (structural effect). The structural effect attributable to smoking within homes and parental smoking suggest group differences in parental behaviours or attitudes which limit or increase the impact of parental smoking in the HSC and LSC groups, respectively.

References

- Action on Smoking and Health (ASH) Research Report (2011), 'Passive smoking: the impact on children' (Available at: <http://www.ash.org.uk/information/secondhand-smoke>)¹.
- Balia, Silvia and Jones, Andrew M. (2008), 'Mortality, lifestyle and socio-economic status', *Journal of Health Economics*, 27 (1), 1-26.
- Becker, Gary S. and Mulligan, Casey B. (1997), 'The Endogenous Determination of Time Preference', *The Quarterly Journal of Economics*, 112 (3), 729-58.
- Blinder, AS (1973), 'Wage discrimination: reduced form and structural estimates', *Journal of Human Resources*, 8 (4), 436 - 55.
- Blow, Laura, Leicester, Andrew, and Windmeijer, Frank (2005), 'Parental Income and Children's Smoking Behaviour: Evidence from the British Household Panel Survey', *Institute of Fiscal Studies Working Paper Series* (London: Institute of Fiscal Studies).
- Bommier, Antoine and Stecklov, Guy (2002), 'Defining health inequality: why Rawls succeeds where social welfare theory fails', *Journal of Health Economics*, 21 (3), 497-513.
- Breuer, Wolfgang , et al. (2011), 'Time Preferences, Culture, and Household Debt Maturity Choice', *NCCR Working Paper 674* (Available at: http://www.nccr-finrisk.uzh.ch/media/pdf/wp/WP674_A1.pdf).
- Cameron, Lisa and Williams, Jenny (2009), 'Is the relationship between socioeconomic status and health stronger for older children in developing countries?', *Demography*, 46 (2), 303-24.
- Center for Disease Control and Prevention (2004), 'The Health Consequences of Smoking: A report of the Surgeon General's Executive Summary' (Available at: http://www.cdc.gov/tobacco/data_statistics/sgr/2004/pdfs/executivesummary.pdf).
- Charlton, Anne (1996), 'Children and smoking: the family circle', *British Medical Bulletin*, 52 (1), 90-107.
- Condliffe, Simon and Link, Charles R. (2008), 'The Relationship between Economic Status and Child Health: Evidence from the United States', *American Economic Review*, 98 (4), 1605-18.
- Contoyannis, Paul and Jones, Andrew M. (2004), 'Socio-economic status, health and lifestyle', *Journal of Health Economics*, 23 (5), 965-95.
- Currie, Alison, Shields, Michael A., and Price, Stephen Wheatley (2007), 'The child health/family income gradient: Evidence from England', *Journal of Health Economics*, 26 (2), 213-32.
- Currie, Candace, et al. (2008), 'Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale', *Social Science & Medicine*, 66 (6), 1429-36.

- Currie, Janet (2009), 'Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development', *Journal of Economic Literature*, 47 (1), 87-122.
- Cutler, David M. and Lleras-Muney, Adriana (2010), 'Understanding differences in health behaviors by education', *Journal of Health Economics*, 29 (1), 1-28.
- Dohmen, Thomas J, et al. (2008), 'The Intergenerational Transmission of Risk and Trust Attitudes', *CEPR Discussion Paper no. 6844* (London: Centre for Economic Policy Research (Available at: <http://www.cepr.org/pubs/dps/DP6844.asp>)).
- Edoka, I.P (2011), 'Parental Income and Smoking Participation in Adolescents: Implications of misclassification error in empirical studies of adolescent smoking participation', *Health Econometrics and Data Group working paper WP 11/29* (York: University of York).
- Environmental Protection Agency (1997), 'Health Effects of Exposure to Environmental Tobacco Smoke', (California Environmental Protection Agency: Sacramento Available at: <http://oehha.ca.gov/pdf/chapter2.pdf>).
- Eriksen, W (2004), 'Do people who were passive smokers during childhood have increased risk of long-term work disability?', *European Journal of Public Health*, 14, 296–300.
- Firpo, Sergio, Fortin, Nicole M., and Lemieux, Thomas (2009), 'Unconditional Quantile Regressions', *Econometrica*, 77 (3), 953-73.
- Fortin, Nicole, Lemieux, Thomas, and Firpo, Sergio (2011), 'Chapter 1 - Decomposition Methods in Economics', in Ashenfelter Orley and Card David (eds.), *Handbook of Labor Economics* (Volume 4, Part A: Elsevier), 1-102.
- Frijters, Paul, et al. (2011), 'Quantifying the cost of passive smoking on child health: evidence from children's cotinine samples', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174 (1), 195-212.
- Gilliland, F. D., et al. (2003), 'Environmental Tobacco Smoke and Absenteeism Related to Respiratory Illness in Schoolchildren', *American Journal of Epidemiology*, 157 (10), 861-69.
- Glendinning, Anthony, Shucksmith, Janet, and Hendry, Leo (1994), 'Social class and adolescent smoking behaviour', *Social Science & Medicine*, 38 (10), 1449-60.
- Graham, H. and Power, C (2004), 'Childhood Disadvantage and Adult Health: a Lifecourse Framework', (National Health Service Health Development Agency, London).
- Grossman, Michael (1972), 'On the Concept of Health Capital and the Demand for Health', *Journal of Political Economy*, 80 (2), 223-55.
- (2000), 'Chapter 7 The human capital model', in J. Culyer Anthony and P. Newhouse Joseph (eds.), *Handbook of Health Economics* (Volume 1, Part A: Elsevier), 347-408.
- Gruber, Jonathan and Zinman, Jonathan (2000), 'Youth Smoking in the U.S.: Evidence and Implications', *National Bureau of Economic Research Working Paper Series*, No. 7780.

- (2001), 'Youth Smoking in the United States: Evidence and Implications', in Jonathan Gruber (ed.), *Risky Behavior Among Youths: An Economic Analysis* (The University of Chicago Press), 69-120.
- Jann, B. (2008), 'The Blinder-Oaxaca decomposition for linear regression models', *Stata Journal*, 8 (4), 453-79.
- Jarvis, Martin J., et al. (2008), 'Assessing smoking status in children, adolescents and adults: cotinine cut-points revisited', *Addiction*, 103 (9), 1553-61.
- Jones, F. L. and Kelley, Jonathan (1984), 'Decomposing Differences between Groups', *Sociological Methods & Research*, 12 (3), 323-43.
- Kenkel, Donald S. (1991), 'Health Behavior, Health Knowledge, and Schooling', *Journal of Political Economy*, 99 (2), 287-305.
- Khwaja, Ahmed, Silverman, Dan, and Sloan, Frank (2007), 'Time preference, time discounting, and smoking decisions', *Journal of Health Economics*, 26 (5), 927-49.
- Leigh, J. Paul (1986), 'Accounting for Tastes: Correlates of Risk and Time Preferences', *Journal of Post Keynesian Economics*, 9 (1), 17-31.
- Loureiro, Maria L., Sanz-de-Galdeano, Anna, and Vuri, Daniela (2010), 'Smoking Habits: Like Father, Like Son, Like Mother, Like Daughter?', *Oxford Bulletin of Economics and Statistics*, 72 (6), 717-43.
- Machado, José A. F. and Mata, José (2005), 'Counterfactual decomposition of changes in wage distributions using quantile regression', *Journal of Applied Econometrics*, 20 (4), 445-65.
- Melly, Blaise (2005), 'Decomposition of differences in distribution using quantile regression', *Labour Economics*, 12 (4), 577-90.
- Murasko, Jason E (2008), 'An evaluation of the age-profile in the relationship between household income and the health of children in the United States', *Journal of Health Economics*, 27 (6), 1489-502.
- Oaxaca, Ronald (1973), 'Male-female wage differentials in urban labour markets', *International Economic Review*, 9, 693 - 709.
- Peretti-Watel, Patrick, et al. (2007), 'Smoking too few cigarettes to be at risk? Smokers' perceptions of risk and risk denial, a French survey', *Tobacco Control*, 16 (5), 351-56.
- Powell, Lisa M. and Chaloupka, Frank J. (2005), 'Parents, public policy, and youth smoking', *Journal of Policy Analysis and Management*, 24 (1), 93-112.
- Powell, Lisa M., Tauras, John A., and Ross, Hana (2005), 'The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior', *Journal of Health Economics*, 24 (5), 950-68.

- Prescott-Clarke, Patricia and Primatesta, Paola (eds.) (1998), *Health Survey for England 1996* (The Stationery Office).
- Propper, Carol, Rigg, John, and Burgess, Simon (2007), 'Child health: evidence on the roles of family income and maternal mental health from a UK birth cohort', *Health Economics*, 16 (11), 1245-69.
- Rawls, J. A (1971), 'Theory of Justice', (Harvard University Press Cambridge, MA).
- Reinhold, Steffen and Jürges, Hendrik (2011), 'Parental income and child health in Germany', *Health Economics* (Available from: DOI: 10.1002/hec.1732).
- Sims, Michelle, et al. (2010), 'Trends in and predictors of second-hand smoke exposure indexed by cotinine in children in England from 1996 to 2006', *Addiction*, 105 (3), 543-53.
- Soteriades, Elpidoforos S. and DiFranza, Joseph R. (2003), 'Parent's Socioeconomic Status, Adolescents' Disposable Income, and Adolescents' Smoking Status in Massachusetts', *American Journal of Public Health*, 93 (7), 1155-60.
- Tobacco Advisory Group of the Royal College of Physicians (2010), ' Report on passive smoking and children'.
- Tuinstra, J., et al. (1998), 'Socio-economic differences in health risk behavior in adolescence: Do they exist?', *Social Science & Medicine*, 47 (1), 67-74.
- Tyas, Suzanne L. and Pederson, Linda L. (1998), 'Psychosocial factors related to adolescent smoking: a critical review of the literature', *Tobacco Control*, 7 (4), 409-20.
- US Department of Health and Human Services (2006), 'The health consequences of involuntary exposure to tobacco smoke: A report of the surgeon general (Available at: www.surgeongeneral.gov/library/secondhandsmoke/)'.
- Vallejo-Torres, Laura and Morris, Stephen (2010), 'The contribution of smoking and obesity to income-related inequalities in health in England', *Social Science & Medicine*, 71 (6), 1189-98.
- Yolton, Kimberly, et al. (2004), 'Exposure to Environmental Tobacco Smoke and Cognitive Abilities among U.S. Children and Adolescents', *Environmental Health Perspective*, 113 (1).
- Yun, Myeong-Su (2005), 'A simple solution to the identification problem in detailed wage decompositions ', *Economic Inquiry*, 43 (4), 766-72.

Table 1 Mean characteristic by parents' socioeconomic class

Variables	High social class	Low social class	Mean difference	t-statistics of difference
Log household income	9.872	8.925	0.947	33.96
Household income	23976.72	9247.87	14728.85	25.70
8-10 years old	0.387	0.424	-0.037	-1.81
11-12 years old	0.257	0.254	0.003	0.18
13 years old	0.113	0.104	0.009	0.66
14-15 years old	0.243	0.218	0.025	1.42
White	0.928	0.888	0.039	3.32
Male	0.497	0.517	-0.019	-0.92
Urban	0.102	0.177	-0.075	-5.30
Suburb	0.590	0.647	-0.057	-2.81
Rural	0.308	0.175	0.132	7.33
Non-smoke free homes	0.238	0.587	-0.348	-18.27
Mother(M) smokes	0.105	0.251	-0.145	-9.52
Father (F) smokes	0.089	0.138	-0.048	-3.70
M Degree	0.143	0.020	0.123	10.35
M Below degree	0.077	0.023	0.054	5.71
M NVQ3	0.079	0.025	0.054	5.56
M NVQ2/NVQ1	0.193	0.246	-0.053	-3.09
M No qualification	0.067	0.194	-0.127	-9.52
F Degree	0.204	0.011	0.193	14.44
F Below degree	0.102	0.027	0.075	7.02
F NVQ3	0.064	0.022	0.043	4.80
F NVQ2/NVQ1	0.083	0.098	-0.015	-1.26
F No qualification	0.029	0.136	-0.106	-9.97
M age ≤ 35 years	0.091	0.235	-0.144	-9.82
M age 36-45 years	0.364	0.230	0.135	7.02
M age ≥ 46 years	0.103	0.045	0.058	5.16
F age ≤ 35 years	0.044	0.076	-0.033	-3.35
F age 36-45 years	0.268	0.163	0.106	6.07
F age ≥ 46 years	0.171	0.056	0.115	8.41
Parent single	0.087	0.239	-0.152	-10.38
M missing (single father/mother not home)	0.017	0.016	0.002	0.28
F missing (single mother/father not home)	0.094	0.230	-0.136	-9.26
Boost sample	0.423	0.476	-0.053	-2.54
Observations	1397	958		

Table 2: Descriptive statistics: Differences in log cotinine distribution

	High social class	Low social class	Difference	t-statistics of difference
Mean	-0.639	0.612	-1.251	-15.67
25th percentile	-1.609	-0.693	-0.916	-9.50
50th percentile	-0.916	0.47	-1.386	-18.14
75th percentile	0	1.482	-1.482	-14.01
Sample weights applied				

Table 3: Aggregate Decomposition of socioeconomic differences in log cotinine distribution

Oaxaca-Blinder Decomposition						
Difference attributable to:	Mean	% Change				
Characteristics	-1.115*** (0.132)	89%				
Coefficients	-0.424*** (0.0977)	34%				
Interaction	0.288* (0.140)	-23%				
Total difference	-1.251*** (0.0818)	100%				
CQR Decomposition (Melly 2005)						
Difference attributable to:	Q25	%Change	Q50	%Change	Q75	%Change
Characteristics	-1.101*** (0.193)	112%	-1.166*** (0.147)	82%	-0.851*** (0.127)	58%
Coefficients	0.076 (0.234)	-8%	-0.094 (0.199)	7%	-0.261 (0.188)	18%
Residuals	0.038 (0.125)	-4%	-0.158 (0.112)	11%	-0.356** (0.142)	24%
Total difference	-0.987*** (0.103)	100%	-1.418*** (0.071)	100%	-1.467*** (0.095)	100%
RIFR Decomposition (Firpo et al. 2009)						
Difference attributable to:	Q25	%Change	Q50	%Change	Q75	%Change
Characteristics	-1.389*** (0.206)	120%	-1.379*** (0.161)	106%	-0.839*** (0.164)	65%
Coefficients	-0.644* (0.314)	56%	-0.626* (0.248)	48%	-0.0268 (0.302)	2%
Interaction	0.874*** (0.219)	-75%	0.703*** (0.173)	-54%	-0.424** (0.198)	33%
Total difference	-1.158*** (0.297)	100%	-1.302*** (0.229)	100%	-1.290*** (0.263)	100%

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied

Table 4: Detailed decomposition of socioeconomic differences in log cotinine distribution

Difference attributable to:	OB Decomposition		RIFR Decomposition					
	Mean	%Change	Q25	%Change	Q50	%Change	Q75	%Change
Characteristics								
Log household income	-0.231* (0.0926)	21%	-0.168 (0.106)	12%	-0.255* (0.112)	18%	-0.161 (0.114)	19%
Gender	0.000683 (0.00189)	0%	-0.000439 (0.00306)	0%	0.000221 (0.00300)	0%	0.00442 (0.00734)	-1%
Ethnicity	0.0120 (0.00747)	-1%	0.0113 (0.00900)	-1%	0.0152 (0.00853)	-1%	0.00160 (0.00857)	0%
Age last birthday	0.0399 (0.0250)	-4%	0.00650 (0.00865)	0%	0.00143 (0.00917)	0%	0.0335 (0.0215)	-4%
Home location	-0.00801 (0.0218)	1%	0.0507 (0.0270)	-4%	-0.00967 (0.0241)	1%	-0.00900 (0.0280)	1%
Non-smoke-free homes	-0.496*** (0.0593)	44%	-0.797*** (0.0774)	57%	-0.615*** (0.0666)	45%	-0.242*** (0.0649)	29%
Mother's smoking status	-0.183*** (0.0468)	16%	-0.0917* (0.0375)	7%	-0.0799* (0.0378)	6%	-0.190*** (0.0465)	23%
Father's smoking status	-0.0405 (0.0668)	4%	0.00506 (0.0170)	0%	-0.0360 (0.0186)	3%	-0.0418 (0.0220)	5%
Mother's education	-0.0527 (0.0483)	5%	-0.203* (0.0954)	15%	-0.136 (0.0750)	10%	-0.0870 (0.0717)	10%
Father's education	-0.388** (0.147)	35%	-0.0623 (0.160)	4%	-0.193 (0.118)	14%	-0.0485 (0.122)	6%
Mother's age	0.0365 (0.0428)	-3%	-0.0303 (0.0388)	2%	0.00864 (0.0372)	-1%	0.0551 (0.0466)	-7%
Father's age	0.225 (0.124)	-20%	0.0494 (0.0325)	-4%	-0.0203 (0.0299)	1%	-0.0456 (0.0384)	5%
Marital status	-0.0315 (0.0649)	3%	-0.159** (0.0602)	11%	-0.0592 (0.0507)	4%	-0.109 (0.0625)	13%
Total	-1.115*** (0.132)	100%	-1.389*** (0.206)	100%	-1.379*** (0.161)	100%	-0.839*** (0.164)	100%

Table 4 Continued

Difference attributable to:	OB Decomposition		RIFR Decomposition					
	Mean	%Change	Q25	%Change	Q50	%Change	Q75	%Change
Coefficients								
Log household income	0.851 (1.023)	-201%	0.336 (1.173)	-52%	1.132 (1.169)	-181%	-0.676 (1.311)	2522%
Gender	0.00265 (0.0681)	-1%	0.0410 (0.0756)	-6%	-0.00441 (0.0727)	1%	0.137 (0.0923)	-511%
Ethnicity	-0.103 (0.194)	24%	-0.323 (0.246)	50%	-0.357 (0.213)	57%	0.350 (0.283)	-1306%
Age last birthday	0.0750 (0.0383)	-18%	0.0323 (0.0400)	-5%	-0.0322 (0.0380)	5%	0.0831 (0.0506)	-310%
Home location	-0.000782 (0.0432)	0%	0.00731 (0.0512)	-1%	-0.0882 (0.0467)	14%	0.0140 (0.0575)	-52%
Non-smoke-free homes	-0.0716 (0.113)	17%	-0.954*** (0.128)	148%	-0.312** (0.115)	50%	0.854*** (0.153)	-3187%
Mother's smoking status	-0.00410 (0.0525)	1%	-0.0474 (0.0744)	7%	-0.0252 (0.0741)	4%	-0.0120 (0.0996)	45%
Father's smoking status	-0.00454 (0.0208)	1%	0.0571 (0.0487)	-9%	-0.0527 (0.0480)	8%	-0.141* (0.0646)	526%
Mother's education	-0.0620 (0.0911)	15%	-0.333 (0.207)	52%	-0.150 (0.160)	24%	0.121 (0.190)	-451%
Father's education	-0.495** (0.186)	117%	-0.0650 (0.109)	10%	-0.177* (0.0890)	28%	0.00579 (0.102)	-22%
Mother's age	0.378* (0.153)	-89%	-0.0530 (0.0639)	8%	-0.105* (0.0505)	17%	0.0275 (0.0598)	-103%
Father's age	0.0845 (0.0717)	-20%	-0.00692 (0.0254)	1%	-0.0120 (0.0219)	2%	0.00473 (0.0280)	-18%
Marital status	0.0240 (0.0284)	-6%	0.0151 (0.0127)	-2%	0.00387 (0.00862)	-1%	0.00203 (0.0107)	-8%
Constant	-1.098 (1.106)	259%	0.650 (1.252)	-101%	-0.446 (1.239)	71%	-0.796 (1.400)	2970%
Total	-0.424*** (0.0977)	100%	-0.644* (0.314)	100%	-0.626* (0.248)	100%	-0.0268 (0.302)	100%

Table 4 Continued

Difference attributable to:	OB Decomposition		RIFR Decomposition					
	Mean	%Change	Q25	%Change	Q50	%Change	Q75	%Change
Interaction								
Log household income	0.0914 (0.110)	32%	0.0361 (0.126)	4%	0.122 (0.126)	17%	-0.0726 (0.141)	17%
Gender	-0.000075 (0.00193)	0%	-0.00116 (0.00411)	0%	0.000125 (0.00361)	0%	-0.00387 (0.00722)	1%
Ethnicity	-0.00396 (0.00758)	-1%	-0.0124 (0.0110)	-1%	-0.0137 (0.00999)	-2%	0.0134 (0.0125)	-3%
Age last birthday	-0.0154 (0.0114)	-5%	-0.00123 (0.00982)	0%	0.00477 (0.00901)	1%	-0.0167 (0.0141)	4%
Home location	0.0000751 (0.0257)	0%	-0.0179 (0.0319)	-2%	0.00934 (0.0290)	1%	-0.0104 (0.0347)	2%
Non-smoke-free homes	0.0431 (0.0679)	15%	0.574*** (0.0816)	66%	0.188** (0.0700)	27%	-0.514*** (0.0951)	121%
Mother's smoking status	0.0460 (0.0532)	16%	0.0278 (0.0438)	3%	0.0147 (0.0436)	2%	0.00702 (0.0585)	-2%
Father's smoking status	0.00596 (0.0837)	2%	-0.0230 (0.0205)	-3%	0.0212 (0.0201)	3%	0.0569* (0.0288)	-13%
Mother's education	0.0476 (0.0559)	17%	0.141 (0.102)	16%	0.0963 (0.0808)	14%	0.124 (0.0864)	-29%
Father's education	0.382* (0.150)	133%	0.0798 (0.173)	9%	0.207 (0.131)	29%	-0.0319 (0.143)	8%
Mother's age	-0.0342 (0.0547)	-12%	0.0188 (0.0536)	2%	0.0397 (0.0504)	6%	-0.0287 (0.0655)	7%
Father's age	-0.251 (0.131)	-87%	-0.0745 (0.0422)	-9%	-0.0187 (0.0382)	-3%	0.0348 (0.0502)	-8%
Marital status	-0.0239 (0.0763)	-8%	0.127 (0.0719)	15%	0.0325 (0.0619)	5%	0.0170 (0.0786)	-4%
Total	0.288* (0.140)	100%	0.874*** (0.219)	100%	0.703*** (0.173)	100%	-0.424* (0.198)	100%

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied

Appendix

Table A1 Description of Variables

Variable name	Variable Label
Male	1 if male
White	1 if white
8-10 years old (base group)	1 if aged between 8 and 10 years
11-12 years old	1 if aged between 11 and 12 years
13 years old	1 if aged 13 years
14-15 years old	1 if aged between 14 and 15 years
<u>Household characteristics</u>	
Log household income	Log household income
Urban	1 if lives in inner city/other dense urban or city centre
Suburb	1 if lives in a suburb residential (city/large town outskirts)
Rural	1 if lives in rural residential/village centre/rural agric. with isolated dwelling
Non-smoke free homes	1 if no one smokes inside the house/flat on most days
Mother(M) smokes	1 if mother is an smoker
Father (F) smokes	1 if father is a smoker
<u>Father's highest qualification</u>	
F Degree	1 if NVQ4/NVQ5/Degree or equivalent
F Below degree	1 if higher education below degree
F NVQ3	1 if NVQ3/GCE A level equivalent
F NVQ2/NVQ1	1 if NVQ2/GCE O level/NVQ1/CSE or equivalent/foreign/other qualification
F No qualification	1 if no qualification
<u>Mother's highest qualification</u>	
M Degree	1 if NVQ4/NVQ5/Degree or equivalent
M Below degree	1 if higher education but below degree
M NVQ3	1 if NVQ3/GCE A level equivalent
M NVQ2/NVQ1	1 if NVQ2/GCE O level/NVQ1/CSE or equivalent/foreign/other qualification
M No qualification	1 if no qualification
<u>Parent's age</u>	
F/M age ≤ 35 years	1 if father/mother's age is less than or equal to 35 years
F/M age 36-45 years	1 if father/mother is aged between 36 and 45 years
F/M age ≥ 46 years	1 if father/mother's age is greater or equal to 46 years
<u>Mother's employment status</u>	
M age ≤ 35 years(base group)	1 if in paid employment or self employed
M age 36-45 years	1 if unemployed/looks after home/retired/full-time education
M age ≥ 46 years	1 if retired/permanently unable to work due to long-term sickness/disability
<u>Parents marital status</u>	
Parent married (Base group)	1 if parents are married/cohabiting
Parent single	1 if parent is single/divorced/widowed/separated
<u>Parents missing variable indicator</u>	
F missing	1 if father information is missing because mother is single/father not home
M missing	1 if mother information is missing because father is single/mother not home
Boost sample	1 if information on both parents are missing (boost sample)

Table A2: OLS regression of Log cotinine

Variables	High social class		Low social class	
	Coeff.	Std. error	Coeff.	Std. error
Log household income	-0.150*	0.0613	-0.244*	0.0965
Male	-0.0385	0.0767	-0.0574	0.108
White	0.229	0.133	0.353*	0.173
8-10 years old (base group)				
11-12 years old	-0.00630	0.0697	-0.00767	0.0945
13 years old	0.213	0.111	0.627**	0.225
14-15 years old	0.876***	0.131	1.337***	0.196
Urban (base group)				
Suburb	0.0225	0.126	0.0409	0.144
Rural	-0.0456	0.132	-0.0354	0.183
Non-smoke free homes	1.253***	0.120	1.363***	0.144
Mother(M) smokes	0.635***	0.148	0.817***	0.194
Father (F) smokes	0.0737	0.144	0.166	0.236
M Degree (base group)				
M Below degree	-0.0662	0.160	0.236	0.438
M NVQ3	0.00323	0.184	0.806	0.422
M NVQ2/NVQ1	-0.0470	0.158	0.447	0.350
M No qualification	0.00673	0.201	0.622	0.368
F Degree (base group)				
F Below degree	0.145	0.166	0.261	0.552
F NVQ3	-0.0198	0.145	-0.584	0.455
F NVQ2/NVQ1	0.247	0.181	0.345	0.428
F No qualification	0.109	0.240	0.426	0.433
M age ≤35 years(base group)				
M age 36-45 years	0.0216	0.139	0.182	0.185
M age ≥46 years	0.0383	0.206	0.107	0.360
F age ≤35 years(base group)				
F age 36-45 years	0.0241	0.157	0.0332	0.228
F age ≥46 years	-0.199	0.195	-0.360	0.332
Parent single	0.429*	0.178	0.468*	0.213
M missing	0.109	0.345	1.993**	0.716
F missing	0.0638	0.216	0.0179	0.488
Boost sample	0.448*	0.185	1.362**	0.498
Constant	-0.182	0.671	-0.0879	1.032
Observations	1397		958	
R ²	0.275		0.380	

Standard errors in second column; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied

Table A3: Q25 RIF regression of Log cotinine

Variable	High social class		Low social class	
	Coeff.	Std. error	Coeff.	Std. error
Log household income	-0.140	0.0719	-0.177	0.110
Male	0.110	0.0887	0.0291	0.117
White	-0.0346	0.167	0.330	0.221
8-10 years old (base group)				
11-12 years old	0.0910	0.112	-0.0731	0.150
13 years old	0.105	0.155	0.335	0.210
14-15 years old	0.162	0.120	0.124	0.167
Urban (base group)				
Suburb	0.103	0.160	0.161	0.173
Rural	0.276	0.173	0.428	0.223
Non-smoke free homes	0.618***	0.109	2.206***	0.178
Mother(M) smokes	0.363**	0.140	0.528*	0.210
Father (F) smokes	0.268	0.148	-0.0712	0.239
M Degree (base group)				
M Below degree	0.137	0.220	0.489	0.701
M NVQ3	0.0760	0.235	1.196	0.766
M NVQ2/NVQ1	0.276	0.187	0.940	0.582
M No qualification	0.185	0.232	1.072	0.584
F Degree (base group)				
F Below degree	0.106	0.186	-0.155	0.798
F NVQ3	0.0392	0.223	-0.994	0.779
F NVQ2/NVQ1	0.0644	0.215	0.0673	0.696
F No qualification	-0.188	0.279	0.524	0.682
M age ≤ 35 years(base group)				
M age 36-45 years	-0.114	0.190	-0.112	0.202
M age ≥ 46 years	0.00388	0.271	-0.344	0.395
F age ≤ 35 years(base group)				
F age 36-45 years	-0.173	0.239	0.405	0.284
F age ≥ 46 years	-0.243	0.281	0.522	0.376
Parent single	0.190	0.208	0.824**	0.305
M missing	0.513	0.382	0.535	0.782
F missing	0.0937	0.319	-0.315	0.770
Boost sample	0.510	0.267	1.785*	0.783
Constant	-0.736	0.791	-2.552*	1.257
Observations	1397		958	
R ²	0.106		0.372	

Standard errors in second column; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied

Table A4: Q50 RIF regression of Log cotinine

	High social class		Low social class	
	Coeff.	Std. error	Coeff.	Std. error
Log household income	-0.140*	0.0596	-0.269*	0.117
Male	-0.0219	0.0794	-0.0171	0.117
White	0.0420	0.158	0.446*	0.180
8-10 years old (base group)				
11-12 years old	-0.0686	0.100	-0.231	0.137
13 years old	0.0320	0.132	-0.174	0.207
14-15 years old	0.234*	0.107	0.143	0.174
Urban (base group)				
Suburb	-0.127	0.138	0.198	0.148
Rural	-0.0531	0.149	0.00988	0.195
Non-smoke free homes	1.184***	0.106	1.702***	0.159
Mother(M) smokes	0.376**	0.132	0.464*	0.214
Father (F) smokes	0.218	0.138	0.521*	0.241
M Degree (base group)				
M Below degree	0.101	0.183	0.108	0.550
M NVQ3	-0.0231	0.184	0.954	0.520
M NVQ2/NVQ1	0.0904	0.156	0.348	0.387
M No qualification	0.120	0.193	0.712	0.389
F Degree (base group)				
F Below degree	0.000485	0.159	0.0950	0.586
F NVQ3	0.110	0.189	-0.503	0.550
F NVQ2/NVQ1	0.224	0.184	0.472	0.492
F No qualification	-0.00473	0.212	0.637	0.484
M age ≤ 35 years(base group)				
M age 36-45 years	0.272	0.177	0.205	0.193
M age ≥ 46 years	0.219	0.232	-0.541	0.316
F age ≤ 35 years(base group)				
F age 36-45 years	-0.314	0.219	-0.0689	0.258
F age ≥ 46 years	-0.419	0.248	-0.271	0.347
Parent single	0.144	0.183	0.305	0.261
M missing	0.632	0.357	1.419**	0.548
F missing	-0.102	0.279	0.539	0.591
Boost sample	0.379	0.233	1.473*	0.579
Constant	0.0909	0.675	-0.0874	1.228
Observations	1397		958	
R ²	0.198		0.349	

Standard errors in second column; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied

Table A5: Q75 RIF regression of Log cotinine

	High social class		Low social class	
	Coeff.	Std. error	Coeff.	Std. error
Log household income	-0.246**	0.0868	-0.173	0.119
Male	-0.0395	0.108	-0.307*	0.143
White	0.440*	0.213	0.0474	0.235
8-10 years old (base group)				
11-12 years old	-0.0941	0.130	-0.0602	0.166
13 years old	0.116	0.191	0.577*	0.265
14-15 years old	0.635***	0.151	1.092***	0.222
Urban (base group)				
Suburb	0.0183	0.183	0.0137	0.195
Rural	-0.132	0.197	-0.0606	0.240
Non-smoke free homes	2.090***	0.180	0.670***	0.176
Mother(M) smokes	1.051***	0.231	1.093***	0.245
Father (F) smokes	-0.213	0.236	0.610*	0.285
M Degree (base group)				
M Below degree	0.00955	0.215	0.106	0.472
M NVQ3	0.0532	0.242	0.865	0.520
M NVQ2/NVQ1	-0.138	0.196	0.122	0.342
M No qualification	-0.361	0.261	0.550	0.373
F Degree (base group)				
F Below degree	0.130	0.203	0.0116	0.573
F NVQ3	-0.0348	0.216	-0.336	0.537
F NVQ2/NVQ1	0.607*	0.251	0.474	0.499
F No qualification	0.482	0.354	0.312	0.492
M age ≤ 35 years(base group)				
M age 36-45 years	0.121	0.242	0.275	0.246
M age ≥ 46 years	0.0368	0.297	0.305	0.382
F age ≤ 35 years(base group)				
F age 36-45 years	0.102	0.302	-0.209	0.314
F age ≥ 46 years	-0.185	0.339	-0.564	0.454
Parent single	0.496*	0.247	0.581	0.319
M missing	-0.260	0.480	1.906**	0.718
F missing	0.0809	0.361	-0.0691	0.587
Boost sample	0.478	0.339	1.003	0.577
Constant	1.088	1.011	1.181	1.251
Observations	1397		958	
R ²	0.283		0.217	

Standard errors in second column; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; sample weights applied