HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 11/29

Parental Income and Smoking Participation in Adolescents: Implications of misclassification error in empirical studies of adolescent smoking participation

Ijeoma Peace Edoka

September 2011

york.ac.uk/res/herc/hedgwp

Parental Income and Smoking Participation in Adolescents:

Implications of misclassification error in empirical studies of adolescent smoking participation

Ijeoma Peace Edoka[*]
Centre for Health Economics, University of York, UK

September 6, 2011

Abstract

In adults, the negative relationship between smoking and income is well established. However divergent results have been reported on the impact of parental socioeconomic status on adolescent smoking. In this study we investigate the extent to which misclassification errors in self-reported smoking affects estimates of the impact of parental income on smoking in adolescents aged 11-15 years old. We use the Household Survey for England (HSE) which contains both a self-reported smoking component and an objective measure of smoking obtained through cotinine assays. Smoking participation is modelled using self-reported smoking and cotinine-validated smoking as binary dependent variables in two separate probit models. We compare marginal effects of parental income (and other independent variables) in both models. Our results suggest that self-reported smoking is misreported leading to biased estimates of the effect of parental income on adolescent smoking. Income-related inequality in smoking (the concentration index) is also underestimated when misclassification errors vary across income quintiles.

Key words: **Adolescent smoking; Self-reported smoking; Cotinine-validated smoking; Parental income; Misreporting; Concentration index**

***JEL classification codes*: I10; I14; C42**

[*]Correspondence to: Centre for Health Economics, University of York, Alcuin Block A, Heslington, York YO10 5DD, UK. E-mail: ijeoma.edoka@york.ac.uk; Tel: +441904321411; Fax: +441904321402

1. Introduction

Tobacco smoking is recognized as the single largest cause of preventable diseases and deaths and is responsible for approximately 26 million years of life lost globally (Lopez, 2005). In England, an estimated 18% of deaths of adults aged 35 years and above were attributable to smoking in 2009 (The NHS Information Centre, 2010). Adolescent smoking is increasingly becoming an important target for health policies in developed countries because a high proportion of long-term adult smokers initiate smoking as children or adolescents. Over the past decade, the UK government has put great efforts into tackling smoking amongst children and adolescents by implementing various anti-smoking policies including the 2003/2004 ban of tobacco advertising and sponsorship in print, on billboards and on the internet , and an increase in the minimum tobacco purchase age from 16 to 18 years as well as a ban on smoking in public places in 2007 (Department of Health, 2010). One of the main aims of the February 2010 White paper, *A Smoke-free Future*, was to 'stop the inflow of young people recruited as smokers'. This White paper set out targets to reduce smoking initiation rates to 1% amongst 11-15-year-olds and to 8% among 16-17-year-olds by 2020.

The progress of anti-smoking policies as well as the impact of other determinants of smoking is monitored using self-reported smoking data collected from national surveys. However, high inconsistencies have been found in self-reported smoking by adolescents in studies comparing self-reported smoking to objective biochemical indicators of smoking, raising important questions about the validity of data (Craig and Mindell, 2008; Kandel et al., 2006; Wagenknecht et al., 1992). In a discrete response model, misreporting smoking behaviour has important consequences in empirical economic analyses when self-reported smoking is used as a binary 0/1 dependent variable. A binary variable is misclassified when a one is miscoded as zero or vice versa. Thus, when an adolescent fails to report their current smoking status correctly, a positive response (one) is miscoded as a negative response (zero) or vice versa. In linear models, under classical assumptions, measurement error in a continuous dependent variable may result in less statistical precision in the estimation of the coefficients but does not lead to biased estimates of coefficients (Hausman, 2001). On the other hand, misclassification error in a binary 0/1 dependent variables will lead to biased and inconsistent estimates of the coefficients in a maximum likelihood estimation approach (Bound et al., 2001; Hausman et al., 1998). Therefore estimates of the impact of the determinants of smoking initiation or participation as well as estimates of the impact of anti-smoking policies are likely to be biased when based on self-reported data.

Inconsistencies in reporting smoking participation may also undermine studies estimating inequalities in smoking between sub-groups of the population, if there are systemic differences in the pattern of misreporting between the groups being compared. For example, Bauman and Ennett (1994) showed that under-reporting of smoking by African American adolescents accounted, in part, for the large racial difference typically seen in self-reported smoking rates between whites and blacks. In terms of income-related inequalities in smoking, if underestimation of smoking rates vary with parental income, estimates of the effect of parental income on adolescent smoking as well as measures of income-related inequality in smoking are likely to be biased.

Reducing inequalities in smoking has been an important policy target in most developed countries because smoking is increasingly concentrated in adults with lower education, lower income and lower socioeconomic status. In addition, long term smokers who initiate smoking at an earlier age are more concentrated in individuals from lower socioeconomic backgrounds. Reports on the impact of parental socioeconomic status on adolescent smoking behaviours have often been mixed[1]. These studies use parental education, social class or income as proxies for parental socioeconomic status. While some studies conclude that higher parental socioeconomic status and income are negatively associated with the probability of smoking initiation and participation amongst adolescents (Gruber and Zinman, 2001; Soteriades and DiFranza, 2003; Tyas and Pederson, 1998), others have reported a higher probability to smoke in high school seniors with more educated parents (Gruber and Zinman, 2001). Other studies fail to observe any association between parental income and smoking prevalence amongst adolescents (Blow et al., 2005). Blow et al. (2005) used the British Household Panel Survey (BHPS) and the British Youth Survey to evaluate the association between adolescent smoking and parental income. After controlling for parental education and adult smoking status, no significant association was observed between parental income and smoking participation in adolescents aged between 11 and 18 years (Blow et al., 2005).

Although the inconsistent results reported may, in part, be attributed to contextual differences in samples used or to differences in the measure of parental socioeconomic status, variations in the extent of misclassification of smoking status in different datasets may also account for the failure to produce significant and consistent estimates of the impact of

[1] Tyas and Pederson, 1998 provides an extensive review of the literature

parental socioeconomic status on adolescent smoking. These studies typically use a binary 0/1 measure of smoking participation as the dependent variable. Hausman et al. (1998) showed, using Monte Carlo simulations[2], that misclassification of a binary dependent variable by as little as 2% can result in significant levels of bias in coefficients estimated from a probit model.

The consequences of misclassification errors have been widely reported in applied economics, labour economics (Falaris, 2011; Hausman et al., 1998), epidemiology (Höfler, 2005; Magder and Hughes, 1997) and insurance (Artís et al., 2002). However, very little attention has been paid to the consequences of misclassification error in smoking participation models in health economics research. To the best of our knowledge, only one study has investigated how misclassification errors in self-reported smoking affect estimates of the impact of cigarette price changes on smoking participation (Kenkel et al., 2004). Kenkel et al. (2004) showed that, in a probit model, estimates of the effect of cigarette prices on smoking participation are biased when smoking participation is misclassified.

In this study, we make an important contribution to the literature by investigating the extent to which misclassification errors in smoking participation models bias estimates of the impact of parental socioeconomic status (parental income) on the probability of smoking participation in adolescents aged 11-15 years. In addition we investigate if there are income-related differences in self-reported smoking behaviour amongst adolescents and how this affects measurements of income-related inequality in smoking. We pool data from the 1997-2008 Health Survey for England, using annual household income as a proxy for parental socioeconomic status. The HSE contains both self-reported smoking and an objective measure of smoking obtained from cotinine assays, allowing for a direct comparison of estimates of the effects of parental income in a self-reported smoking and a cotinine-validated smoking participation model.

Epidemiological studies using objective measures of smoking to validate self-reported smoking behaviour have shown that age, ethnicity, frequency of exposure to other smokers and smoking status of parents and friends are correlated with misreporting smoking behaviour amongst adolescents (Griesler et al., 2008; Kandel et al., 2006). Younger adolescents are more likely to under-report smoking while those who perceive their friends to be smokers are less likely to under-report smoking (Griesler et al., 2008; Kandel et al., 2006).

[2] A sample size of 5000 was used for each simulation (Hausman et al. 1998)

African Americans are more likely to under-report smoking compared to whites (Bauman and Ennett, 1994; Griesler et al., 2008; Kandel et al., 2006). Therefore we examine the extent to which parental income as well as other observed characteristics predict under-reporting of smoking participation amongst adolescents. Although adolescents are also prone to over-report their smoking behaviour, we focus on under-reporting for two main reasons. First, smoking is generally considered to be socially undesirable. Therefore, adolescent non-smokers are more likely to report truthfully whereas adolescent smokers are likely to be untruthful. Second, we use data collected from a household survey and household surveys have been shown to produce lower estimates of adolescent smoking rates in comparison to school-based surveys (Craig and Mindell, 2008; Griesler et al., 2008). This suggest that under-reporting of smoking behaviour is likely to be more prevalent in household surveys.

The rest of the paper is organised as follows: In section 2 we describe the model of smoking participation with misclassification to quantify the extent to which coefficients (and marginal effects) in a probit model are biased as a result of misclassification error. In addition, a description of the HSE and the variables used in the study are detailed in section 2. The results are presented and discussed in section 3 and concluding comments are provided in the section 4.

2. Method

2.1 Model of Smoking Participation with Misclassification

Traditional economic models of smoking behaviour are based on the economic theory of demand such that the tobacco demand equation is derived from a utility maximization process in which an individual maximizes his/her utility subject to a budget constraint which may comprise both economic and social factors. Behavioural models have evolved to incorporate important dimensions that reflect decision-making unique to children and adolescents. For example, in addition to economic factors such as tobacco price and other anti-tobacco policies, social factors may influence the decision-making process of adolescents. These social factors interact to create an environment which may either reduce the perceived costs of smoking or increase the perceived benefits[3].

[3] A detailed discussion is provided by O'donoghue & Rabin, 2001

Empirical studies typically model the determinants of smoking participation using a binary 0/1 dependent variable of self-reported smoking. The extent of misclassification error in a non-linear model for smoking participation can be shown using the framework of Hausman et al. (1998). The tobacco demand equation can be expressed as a latent variable model:

$$Y^* = \mathrm{X}\beta + \varepsilon \qquad (1)$$

where Y^* is a latent variable and is a linear function of a vector of covariates X and an error term, ε, which is independent and identically distributed (i.i.d).

An adolescent decides to smoke when the net benefits of smoking (for example, social acceptability amongst peers) outweighs the costs. In this study we observe cotinine-validated smoking behaviour as the true smoking behaviour. We define a binary variable, Y^{CV} which equals one when saliva cotinine level are 12ng/ml or over and zero otherwise. Saliva cotinine levels greater than or equal to 12ng/ml have been shown to be indicative of active smoking in adolescents (Jarvis et al., 2008).

$$\mathrm{Y^{CV}} = \begin{cases} 1 \; if \quad \mathrm{Y^*} \geq 0 \\ 0 \; otherwise \end{cases} \qquad (2)$$

In addition, we observe self-reported smoking behaviour, Y^{SR}, which is equal to one when an adolescent reports smoking and zero otherwise. When self-reported smoking is misreported, Y^{SR} is misclassified. Two misclassification probabilities, p01 and p10 can be identified: the probability that a zero is misclassified as a one (equation 3) and the probability that a one is misclassified as a zero (equation 4), respectively:

$$p01 = \Pr(Y^{SR} = 1 \big| Y^{\mathrm{CV}} = 0) \qquad (3)$$

$$p10 = \Pr(Y^{SR} = 0 \big| Y^{\mathrm{CV}} = 1) \qquad (4)$$

The model of misclassification assumes that misclassification probabilities p01 and p10 are only dependent on the value of Y^{CV} but are independent of X (Hausman et al., 1998).

The expected values of cotinine-validated smoking participation (Y^{CV}) and the self-reported smoking participation (Y^{SR}) can be written as:

$$E(Y^{CV}|X) = \Pr(Y^{CV} = 1|X) = F(X\beta) \quad (5)$$

$$E(Y^{SR}|X) = \Pr(Y^{SR} = 1|X) = p01 + (1 - p01 - p10)F(X\beta) \quad (6)^4$$

where *F* is the cumulative distribution for a probit model.

When Y^{SR} is measured without error, p01 and p10 are equal to zero and Pr (Y^{SR} =1| *X*) = Pr (Y^{CV} =1| *X*) =*F*(Xβ). However when misclassification errors are present, equation (6) suggests that coefficient estimates (β) are biased towards zero. Similarly, the marginal effects of each independent variable will be biased towards zero. The downwards bias in marginal effects can be seen by estimating the partial derivative of equation 6 with respect to any independent variable:

$$\frac{\partial \Pr\,(\,Y^{SR} = 1|X)}{\partial x_j} = \;\; (1 - p01 - p10)\, f(X\beta)\beta_j \quad (7)$$

Where *j* is any independent variable

For example if *j* is parental income and self-reported smoking (Y^{SR}) is misreported, then the marginal effect of parental income will be understated by a value '*ω*', where *ω* = 1-p01- p10. Since the misclassification probabilities p01 and p10 are assumed to be constant across observed characteristics, it is expected that the marginal effect of each covariate will be biased by a constant value, *ω*. In this study, the probability that a zero is misclassified as a one (p01) is approximately 0.02 and the probability that a one is misclassified as a zero is approximately 0.51. This implies that the marginal effects (and coefficients) of observable characteristics in the self-reported smoking participation model will be underestimated by approximately 53% (p01+p10=0.53). We investigate the assumption that p01 and p10 are independent of the observed characteristics using two separate bivariate probit models (equations 8 and 9).

$$p01 = \Pr(Y^{SR} = 1 \big| Y^{\mathrm{CV}} = 0, X) \quad (8)$$

$$p10 = \;\Pr(Y^{SR} = 0 \big| Y^{\mathrm{CV}} = 1, X) \quad (9)$$

[4]See Appendix for proof

Self-reported and cotinine-validated smoking participation are modelled separately as a function of a set of observed characteristics in two separate probit models to estimate the impact of parental income on smoking participation. Following the approach of Blow et al. (2005), only adolescent characteristics (age, gender and ethnicity) are controlled for in a basic model. The basic model is then extended by first including household characteristics and then, other observed parental characteristics which are likely to be correlated with parental income.

2.2. Data Description and Variables

This study uses the HSE and pools data from 1997 to 2008. The HSE is a series of annual cross-sectional surveys designed to include adults and children living in a representative sample of households in England. Every year a nationally representative sample of households is 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.[5] Each year, the survey includes a set of core health and lifestyle topics including smoking, drinking, and general health. Objective measures of health including saliva specimens for cotinine assay were collected in a nurse visit approximately one week after the first interview.

Self-reported smoking behaviour

Data on smoking participation was collected for all individuals aged 8 and above, using self-completed questionnaires. In this study, current smoking status of children aged 11-15 years was obtained from replies to two questions. First, 'Now read all the following sentences carefully and tick the box next to the one which best describes you: (1) I have never smoked; (2) I have only smoked once or twice; (3) I used to smoke sometimes, but I never smoke a cigarette now; (4) I sometimes smoke, but I don't smoke every week; (5) I smoke between one and six cigarettes a week and (6) I smoke more than six cigarettes a week. Second, 'Did you smoke any cigarettes last week?' Those who chose options 4, 5 or 6 in the first questions are classified as current smokers and those who chose options 1, 2, or 3 as non-smokers provided they did not answer 'yes' to the second question. Adolescents who chose options 1, 2 or 3 in the first question and answered 'yes' to the second questions were

[5] A full description of the survey design can be found in Prescott-Clarke, 1998.

classified as current smokers[6]. We define a binary variable of self-reported smoking as one for current smokers and zero for non-smokers.

Objective indicator of smoking

In all years except in 2000, saliva specimens were collected for cotinine assay in a nurse visit one week after the self-reported questionnaires were completed. Cotinine assays were performed using gas chromatography which can detect cotinine levels as low as 0.1ng/ml. Cotinine is a metabolite of nicotine and has a half-life of about 16-20 hours. It can be detected in saliva specimens of regular smokers and occasional smokers if smoking occurred a few days prior to the collection of the specimen. Cotinine is generally accepted as a quantitative indicator of tobacco intake with very high specificity (percentage of non-smokers classified as non-smokers) and sensitivity (percentage of smokers classified as smokers) (Benowitz, 1996; Jarvis et al., 2008). Despite its objectivity, in adolescents, saliva cotinine levels may be subject to the extent of exposure to second-hand smoke. However recent studies have now shown that a cut-point of 12ng/ml will detect active smoking with sensitivity of 95.8% in children aged 8-15 years (Jarvis et al., 2008). In this study we define cotinine-validated smokers as those with cotinine levels greater than or equal to 12ng/ml. However, because the cut-point may vary widely between 8-18ng/ml in non-smokers depending on the extent of exposure to second-hand smoke (Jarvis et al., 2008), we use a cut-point of 18ng/ml in a sensitivity analysis.

Socioeconomic measure and other variables

Equivalised annual household income is used as a proxy for parental socioeconomic status in this study. From 1997 onwards, the HSE collected data on annual household income using cards displaying 31 income bands ranging from less than £520 to greater than £150,000 per annum. These cards were completed by the household reference person or their partner. They were asked to estimate the total annual household income including their own income, the income of their partners and any other persons living within the household. Equivalised income was calculated using the McClements scoring system (McClements, 1977) to account for the number of persons in the household including children.

This study makes use of a wide range of demographic and socioeconomic information available in the HSE. These include demographic characteristics of adolescents: age at the

[6] Less than 1% of those who chose options 1,2 or 3 in the first question answered 'yes' to the second question

time of the interview (11-15 years), gender (male/female) and ethnicity (white/non-white); household characteristics: home ownership, household location (rural/suburban/urban) and social class of the household responsible person. Social class was assigned using the Registrars General's Social Class (RGSC) classification system. Characteristics of parents controlled for include: fathers and mothers highest academic qualification, occupation, marital status and current smoking status. These were obtained by linking parents' responses in the individual questionnaires to each adolescent. Current smoking status of parents was obtained from a positive response to the question: 'Do you smoke cigarettes at all nowadays?'

Finally, 'non-smoke-free' households were identified using the response to the question completed by the household reference person: 'Does anyone smoke inside this house/flat on most days?' Interviewers were explicitly asked to code as 'no' if smoking by any household member was reported, but occurred outside the home and to code as 'yes' if non-household members smoked within the home. Non-smoke-free homes were defined by a binary variable which equals one if respondent answered 'yes' and zero otherwise. Table 1 gives a description of all the variables.

The final sample consists of 7421 observations. This includes all adolescents who had valid cotinine results and complete information on household income and demographic characteristics. Data from 1999, 2000 and 2004 were excluded because in 2000 saliva specimens were not collected and in 1999 and 2004, saliva specimens were collected from adolescents within ethnic minority groups only. For adolescents who had missing information on parents, we include dummy variables to account for the missing parent. A distinction was made between those for whom parental information was missing because they were from a single parent household and those from a two-parent household but where one parent's information was missing because that parent was absent during the interview.

Table 2 shows a descriptive summary of the characteristics of adolescents in the full sample and by income quintiles. In the full sample (Table 2, column 1), the proportion of self-reported smokers (6.4%) is approximately 3 percentage points less than the proportion of cotinine-validated smokers (9.3%). Interestingly, the underestimation of the proportion of self-reported smokers in the full sample varies greatly across income quintiles (Fig. 1). The largest difference between the proportions of cotinine-validated and self-reported smokers is observed at the bottom end (first quintile) of the income distribution with a difference of approximately 7 percentage points in comparison to a difference of 1 percentage point in the

highest (fifth) income quintile. Overall, both self-reported and cotinine-validated smoking follows an income gradient with smoking rates highest amongst the poorest.

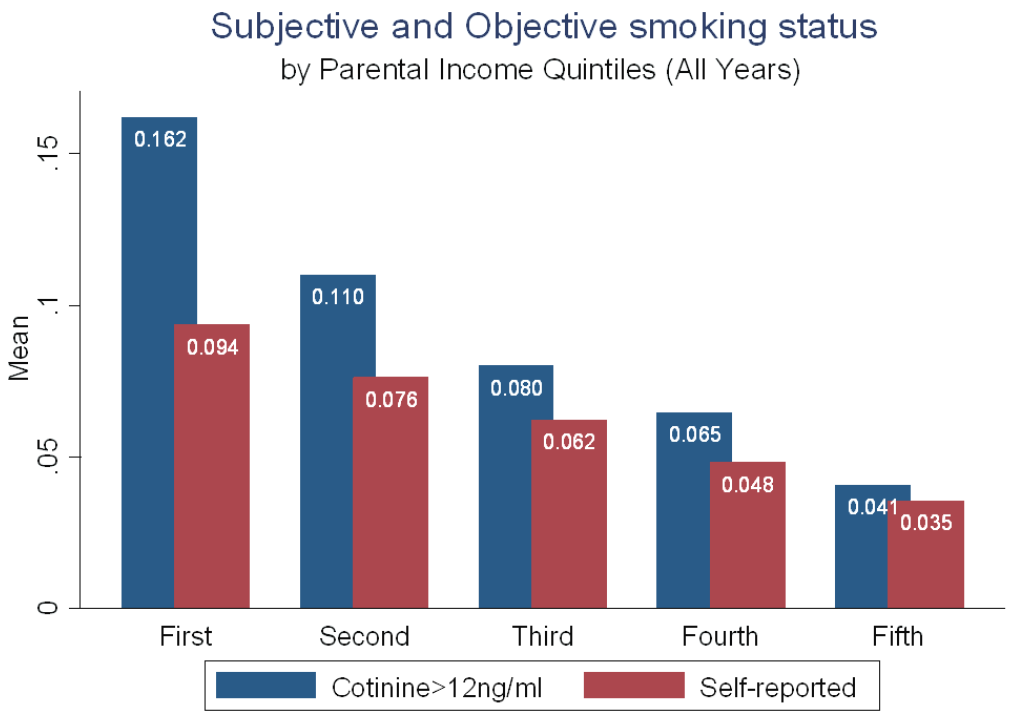

Fig. 1 Self-reported and cotinine-validated smoking by parental income quintiles

The distribution of cotinine by self-reported smoking behaviour (Figs. 2a and 2b) shows further evidence of discrepancies in self-reported smoking. For example, very high levels of cotinine were detected in the saliva sample of a proportion of those who reported to have 'not smoked last week' (Fig. 2a) or in those who reported 'to have smoked sometimes before but never smoke now' (Fig. 2b). On the other hand, in a proportion of those who reported to 'smoke sometimes, but not every week' (Fig. 2b) or to have 'smoked last week' (Fig. 2a), we observe low levels of cotinine not indicative of active smoking (cotinine: 0 to < 12ng/ml). This discrepancy could occur if a non-smoker reports smoking or if an occasional smoker smoked many days before providing a saliva sample. This may result in a failure of the assay to detect cotinine levels indicative of active smoking.

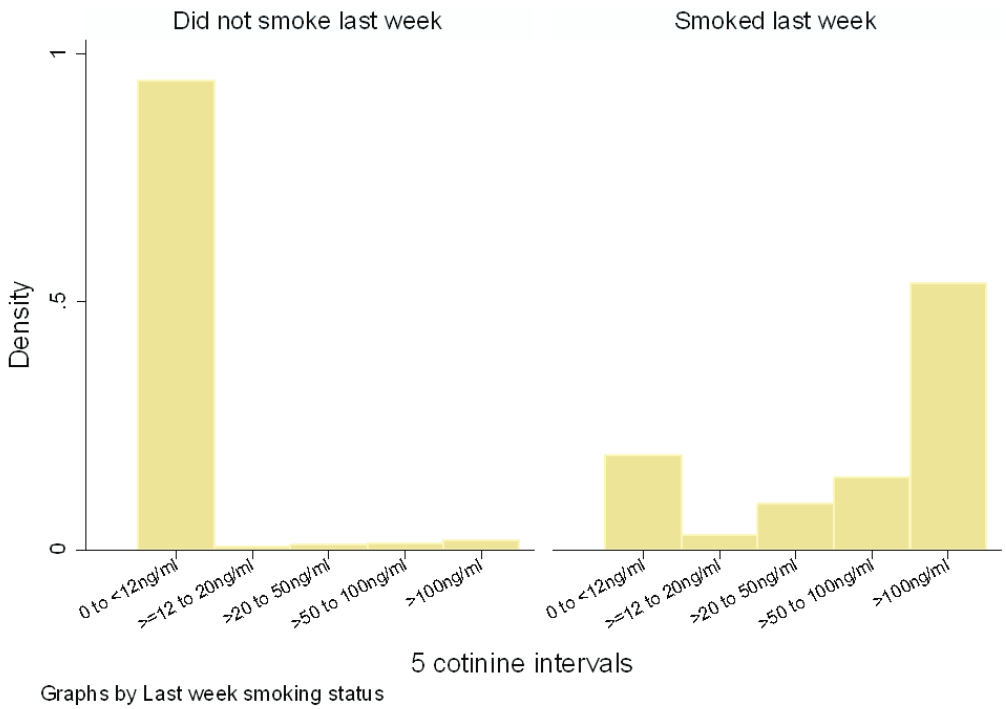

Fig. 2a cotinine distribution by 'last week' self-reported smoking

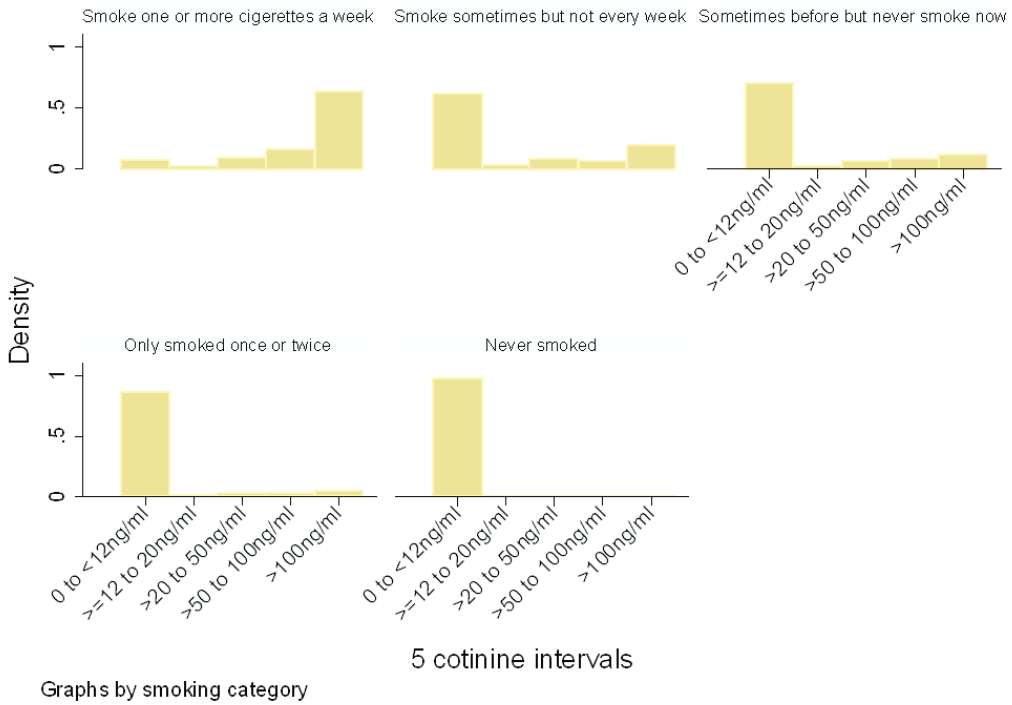

Fig. 2b cotinine distribution by five categories of self-reported smoking behaviour

3. Results and Discussion

The results (marginal effects) of the bivariate probit models estimating the effect of observable characteristics on the misclassification probabilities are presented in Table 3. The probability of over-reporting smoking participation ($p01 = \Pr(Y^{SR} = 1 \big| Y^{\mathrm{CV}} = 0, X)$) is significantly lower in adolescents whose parents own a house, in adolescents whose parents are married and in adolescents whose fathers are not in paid employment but look after the home. In addition, adolescents living in homes within which smoking is permitted are significantly more likely to over-reporting smoking participation. On the other hand, the probability of under-reporting smoking participation ($p10 = \Pr(Y^{SR} = 0 \big| Y^{\mathrm{CV}} = 1, X)$) is higher in adolescents whose parents own a house and in those with fathers who are not in employment but look after the home/family. The probability of under-reporting smoking participation (p10) is lower in adolescents with fathers possessing an NVQ2 or GCE O level qualification (in comparison to those whose fathers possess a degree or equivalent) and in those whose mothers are in full time education or on a government training scheme (in comparison to those with mothers in paid employment). Put differently, adolescents whose parents own a home and whose fathers look after the home/family are more likely to under-report smoking participation and less likely to over-report smoking participation. This suggests that adolescents from higher socioeconomic backgrounds may be less willing to disclose their true smoking behaviour, perhaps due to the perceived undesirability of their behaviour within the household.

Finally, the misclassification probability p01 is significantly lower in 2006-2008, whereas p10 is higher within the same period in comparison to 1997-2001. Other observed characteristics modelled do not predict misclassification errors and the assumption that misclassification probabilities are independent on observed characteristics appears to be true for these covariates.

Results of the cotinine-validated and self-reported smoking participation models are presented in Table 4a and 4b. The results (marginal effects) are interpreted as estimates of the association between observable characteristics and the probability of smoking and do not reflect causality between observable characteristics and adolescent smoking. After controlling for adolescents' characteristics alone, a negative association is observed between parental income and the probability of smoking in both the cotinine-validated and self-reported smoking participation models (columns 1-4, Table 4a). An increase in parental income is

associated with a decrease in the probability of being a smoker. This income effect is statistically significant in both models, but the size of the income effect in the self-reported smoking model is approximately 50% less than that of the cotinine-validated smoking model. This is close to the estimation of the downward bias (53%) introduced by misclassification error in self-reported smoking. A 1% increase in parental income is associated with a decrease in the probability of being a cotinine-validated smoker by approximately 4 percentage points, but by approximately 2 percentage points in the self-reported smoking model.

With regards to other covariates, age is positively associated with the probability of being a smoker in both models. However, while an increase in age results in an increase in the probability of being a cotinine-validated smoker by approximately 15 percentage points, the probability of being a self-reported smoker increases by approximately 7 percentage points. Being white significantly increases the probability of being a smoker by approximately 3 percentage points in the cotinine-validated smoking model but by approximately 2 percentage points in the self-reported smoking model.

When the model is extended to include household characteristics, we observe a reduction in the estimated size of the association between cotinine-validated or self-reported smoking and parental income (columns 5-8, Table 4a). After controlling for home ownership, social class of the head of the household and smoking within the home, the association between income and the probability of being a cotinine-validated smoker remains negative and statistically significant while the effect of parental income on the probability of being a self-reported smoker becomes insignificant. We observe a similar effect when we consider another important socio-economic variable, the social class of the head of the household. The probability of being a cotinine-validated smoker increases by approximately 3 percentage points when an adolescent lives in lower social class household (versus living in a higher social class household). This effect is not significant in the self-reported smoking model. In addition, the effect of ethnicity and age remain significant in the cotinine-validated smoking model but become insignificant in the self-reported smoking model after controlling for household characteristics. In both models, smoking within the home significantly increases the probability of being a cotinine-validated and a self-reported smoker, but, again this effect is underestimated in the self-reported smoking model.

After controlling for the full set of observed characteristics of parents including variables which are likely to be correlated with income such as education and employment status, the association between parental income and cotinine-validated smoking remains negative and significant[7] (Table 4b). On the other hand, the parental income effect in the self-reported smoking model is insignificant (Table 4b). This result is similar to those reported by Blow et al., 2005 using the BHPS. After controlling for parental education and smoking, the observed association between parental income and self-reported smoking disappears (Blow et al., 2005). With respect to other parental characteristics, we observe no association between mother's education and cotinine-validated or self-reported smoking. Adolescents whose fathers have a higher education but below a degree (compared to fathers with a degree or equivalent) are significantly more likely to be cotinine-validated smokers but not self-reported smokers.

For some covariates measured with statistical precision in the self-reported smoking participation model, estimates of the marginal effects are not always 50% less than estimates of marginal effects in the cotinine-validated smoking participation model. This may be because the misclassification probabilities are not independent of these covariates. For example, the misclassification probabilities p01 and p10 are dependent on parental home ownership (Table 3) and the size of the marginal effect of parental home ownership in the self-reported smoking model is approximately equal to the marginal effect in the cotinine-validated smoking model (Table 4b).

Sensitivity tests

The robustness of the results is tested using a higher cut-point (18ng/ml) for defining cotinine-validated smokers. In adolescents, it has been suggested that very high exposure to second-hand smoke could result in saliva cotinine levels as high as 18ng/ml in non-smokers

[7] This result remained robust to the functional form of parental income used. After controlling for the full set of covariates, the probability of being a cotinine-validated smoker increases by approximately 4 percentage points (t-stats=2.7) for the first (lowest) income quintile and approximately 2 percentage points (t-stats=1.79) for the second income quintile in comparison to adolescents from the fifth (highest) income quintile. On the other hand, in the self-reported smoking model, the size of the marginal effects of the income quintiles are less than those observed in the cotinine-validated smoking model and statistically insignificant.

(Jarvis et al., 2008). Therefore a cut-point of 12ng/ml may misclassify non-smokers as smokers if they are exposed to high levels of second-hand smoke.

The results are robust to the cut-point for defining cotinine-validated smokers. Using a cut-point of 18ng/ml, we observe similar results to those obtained with a cut-point of 12ng/ml (last two columns, Table 4b). The marginal effects of all covariates as well as the *t*-statistics obtained using a cut-point of 12ng/ml are similar to those obtained using a cut-point of 18ng/ml. For example, a 1% increase in parental income significantly reduces the probability of being a cotinine-validated smoker by approximately 1 percentage point using a cut-point of either 12ng/ml or 18ng/ml; being white increases the probability of being a cotinine-validated smoker by approximately 2 percentage points in both cases; an increase in age is positively associated with an increase in the probability of being a cotinine-validated smoker by approximately 13 and 12 percentage points using a cut point of 12ng/ml and 18ng/ml respectively. These results are unsurprising given that the proportion of adolescents classified as cotinine-validated smokers using a cut-point of 12ng/ml and 18ng/ml do not differ markedly (9.26 and 8.63 percent respectively; table 2 first column). This implies that less than 1% of adolescents classified as smokers by a cotinine cut-point of 12ng/ml are classified as non-smokers by a cut-point of 18ng/ml.

Determinants of under-reporting of smoking participation

Several factors may explain why adolescents misreport their smoking behaviour (Dolcini et al., 1996). Fear of disclosure and the perceived social undesirability of smoking may result in adolescents under-reporting their true smoking behaviour. Being interviewed by the same interviewer who interviewed a parent is significantly correlated with under-reporting of smoking participation in adolescents (Griesler et al., 2008; Kandel et al., 2006). Higher smoking rates have been observed when anonymous questionnaires are administered in comparison to named questionnaires (Adams et al., 2008). Exit questionnaires completed by adolescents participating in a six months intervention study aimed at reducing the prevalence of smoking showed that approximately 4% of adolescents admitted to have consistently misreported their smoking status during the intervention study (Stein et al., 2002). Of these, approximately 25% were worried the information they give in the study will be relayed back to their parents (Stein et al., 2002). In other studies on substance-abuse, adolescents admit to denying the use of substances to their physicians when their parents are present (Friedman et al., 1990).

Parental smoking or living with other smokers may be associated with misreporting smoking behaviour. For example, in the 2008 Smoking, Drinking and Drug Use Survey, adolescents who smoked openly (i.e. those whose families were aware of their smoking behaviour) were less likely to perceive that their families would disapproved of their smoking in comparison to secret smokers (Fuller, 2009). In households where no other household member smoked, approximately 40% of adolescent smokers were open smokers while 60% were secret smokers. On the other hand, in households where three or more household members were smokers, as high as 71% were open smokers (Fuller, 2009). This suggests that adolescents living in households where other members smoke may be less likely to under-report their smoking. In adults, smoking has been shown to follow a socioeconomic gradient, with smoking concentrated more in adults from lower socioeconomic backgrounds. Taken together, it can be expected that adolescents from higher socioeconomic backgrounds may have a greater motivation not to disclose their true smoking behaviour due to the perceived undesirability of their behaviour.

The setting in which data on smoking is collected is also important. The HSE consistently produces lower estimates of smoking rates in comparison to the Smoking Drinking and Drugs Use Survey, a school-based survey (Craig and Mindell, 2008). In the United States, a study comparing self-reported smoking status in a sample of adolescents within a school and household setting showed that approximately 9-20% of the same adolescents who had reported smoking in the last 12 months in the school-based survey reported never tried smoking in the household counterpart (Griesler et al., 2008).

Due to the sporadic and experimental nature of adolescent smoking, poor recall may result in an unintentional misreporting of smoking behaviour. For example, in the 2008 Smoking Drinking and Drugs Use Survey, 65% of those who reported usually smoking up to six cigarettes a week in the interviews had diary records of smoking more than seven cigarettes in the previous week. Similarly, 46% of those who reported smoking sometimes but less than once a week, had diary entries suggesting they had smoked at least one cigarette in the previous week (Fuller, 2009).

We investigate the association between observable characteristics and the probability to under-report smoking behaviour using a subsample consisting of only cotinine-validated smokers (N=694). A binary variable Y^{UR}, was generated which equals one if a cotinine-validated smoker reports being a non-smoker and zero otherwise. We estimate the probability

that Y^{UR} equals one conditional on a set of covariates, Pr (Y^{UR} =1|X), using maximum likelihood estimation approach. We do not consider those who, despite reporting being smokers, had cotinine levels less than 12ng/ml. This form of discrepancy (over-reporting) is less common and in our sample the probability of over-reporting smoking is approximately 0.02 while the probability of under-reporting smoking in approximately 0.51.

Table 2 (last column) shows a summary of the characteristics of the sub-sample of cotinine-validated smokers. On average, in comparison to the full sample, the sub-sample consists of adolescents from a lower socioeconomic family background. For example, approximately 36% of those in the sub-sample (versus 22% in the full sample) live in household where the household head is from the lowest social class. Average household income is lower in the sub- sample compared to the full sample (£15,000 vs. £20,000 per annum). In addition a higher proportion of cotinine-validated smokers (61% versus 34% in the full sample) live in non-smoke-free homes. Table 2 (last column) also shows that approximately 50% of the cotinine-validated smokers correctly reported their smoking behaviour. When the proportion of adolescents, under-reporting and correctly reporting smoking status are plotted by parental income quintiles (Fig. 3), we observe no clear pattern. A higher proportion of adolescents in the highest income quintile under-report smoking, equal proportions of adolescents in the second and fourth income quintile under-report and report smoking status correctly, while a higher proportion of adolescents in the middle of the income distribution report smoking status correctly.

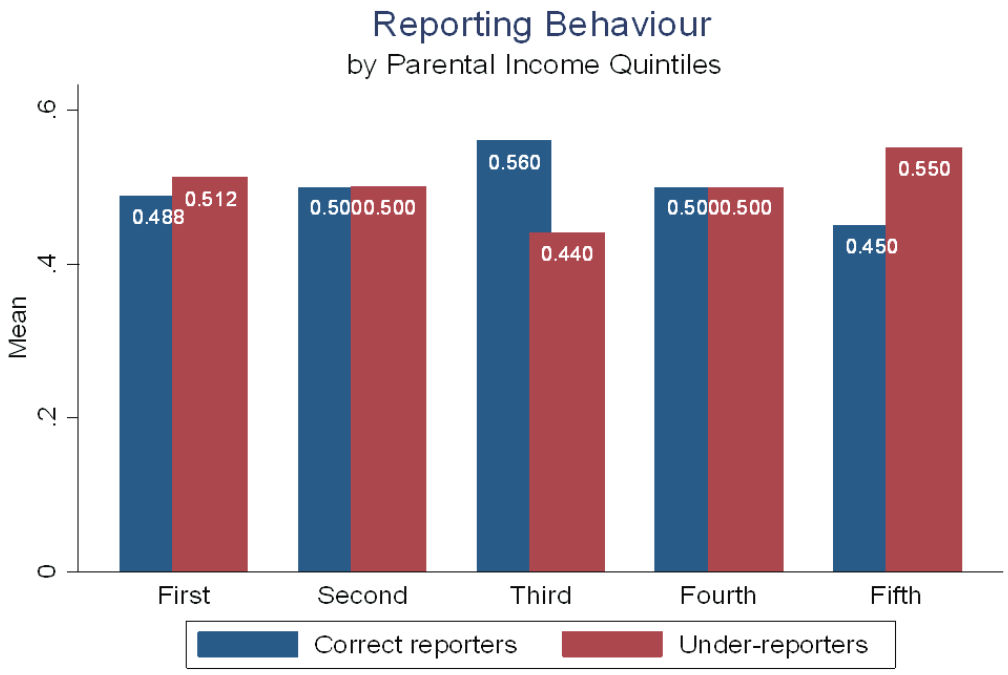

Fig. 3 Self-reported and cotinine –validated smoking in the cotinine-validated sub-sample

Table 5 shows the results (marginal effects) of the probit model of under-reporting smoking, conditional on observable characteristics. In the basic model controlling for only adolescent characteristics, the association between parental income and under-reporting is positive, meaning that as household income increases, the probability of under-reporting smoking participation increase. Although this effect is statistically insignificant, it is consistent with the notion that adolescents living in wealthier households are more likely to be untruthful about their smoking behaviour. When the model is extended to include household characteristics, the association between home ownership (another measure of socioeconomic status) is positive and statistically significant. The probability of under-reporting increases by approximately 12 percentage points when adolescents are living in homes owned by their parents. Smoking within the household reduces the probability of under-reporting by approximately 7 percentage points. Although this effect is statistically insignificant, it is plausible given that adolescents living in homes within which smoking is permitted are likely to perceive their smoking behaviour as acceptable.

With respect to adolescents' characteristics, we find no statistically significant effects of age, ethnicity and gender on the probability to under-report smoking behaviour. Nevertheless, the direction of the association between these variable and reporting behaviour

is worth noting. After controlling for the full set of covariates, older adolescents are less likely to under-report smoking behaviour. This is consistent with findings from recent epidemiological studies which have reported negative associations between age and the probability to under-reporting smoking in adolescents (Griesler et al., 2008; Kandel et al., 2006).

Misreporting smoking participation and income-related inequality in smoking

In addition to producing biased estimates of the effects of observable characteristics on the probability of smoking participation, misclassification errors are likely to affect measures of income-related inequality in smoking if there are systemic differences in misreporting smoking amongst adolescents from different socioeconomic backgrounds. Fig 1 suggests an income (parental) gradient in misreporting smoking participation. The difference between cotinine-validated smoking rates and self-reported smoking rates is highest (approximately 7 percentage points) in adolescents at the lowest parental income quintile. This difference decreases going up the parental income quintile and at the highest parental income quintile the difference falls to less than 1 percentage point.

The concentration index has been widely used in measuring income-related health inequality and depends on the relationship between the variable of interest and the rank position in the income distribution (Kakwani et al., 1997; Wagstaff et al., 1991). It is most suitable when the variable of interest is unbounded and measured on a ratio scale (Erreygers, 2009; Wagstaff, 2005). In this case the value of the concentration index ranges from -1 (in which case the variable of interest is disproportionately concentrated amongst the poor) and +1 (in which case the variable of interest is disproportionately concentrated amongst the rich). The concentration index, *C*, can be defined as:

$$C = \frac{2}{\bar{y}} \mathrm{cov}(Y_i , R_i) \qquad (10)$$

where Y_i is the variable of interest (smoking status for the *i*th individual), R_i is the fractional rank of the *i*th individual in the income distribution and $\bar{y}$ is the mean of Y across all individuals.

However, when the variable of interest is binary, such as in smoking participation, the possible range of values the concentration index can take will depend on $\bar{y}$ such that as $\bar{y}$ increases, the range the concentration index takes becomes narrower. This poses a problem

when comparing the degree of inequality between populations with different means. Therefore, to ensure comparability of the concentration indices estimated using cotinine-validated smoking and self-reported smoking, we use the corrected concentration index (Erreygers, 2009). The corrected concentration index, *CCI* is defined as:

$$CCI = \frac{4\bar{y}}{z - w}C \qquad (11)$$

where *C* is the concentration index derived from equation (10), *z* and *w* are the upper (one) and lower (zero) bounds of *Y*, respectively.

Estimates of the corrected concentration indices based on cotinine-validated and self-reported smoking are presented in Table 6. The corrected concentration indices for both cotinine-validated and self-reported smoking are negative, meaning that smoking is concentrated disproportionately in the poor. However, parental income-related inequality in self-reported smoking is significantly less than inequality in cotinine-validated smoking by approximately 50%. This is unsurprising given that on average, self-reported smoking rates are underestimated to a larger extent in adolescents from poorer households in comparison to adolescents from richer households (Fig 1).

Fig 4 shows a graphical representation (concentration curves) of income-related inequality in self-reported and cotinine-validated smoking. The concentration curve plots the cumulative proportion of smoking by the cumulative proportion of adolescents ranked from the poorest to the richest. The concentration curves of both cotinine-validated and self-reported smoking lie above the diagonal line of equality, meaning that smoking is disproportionately concentrated in the poor. However, the concentration curve of cotinine-validated smoking lies everywhere above the concentration curve of self-reported smoking, suggesting that income-related inequality in cotinine-validated smoking is greater. This may have significant implications when comparing inequality in smoking between populations or within a population across different time periods. If the extent or pattern of misreporting varies significantly across different populations or time, incorrect conclusions may be reached on the extent or trend of income-related inequalities in adolescent smoking.

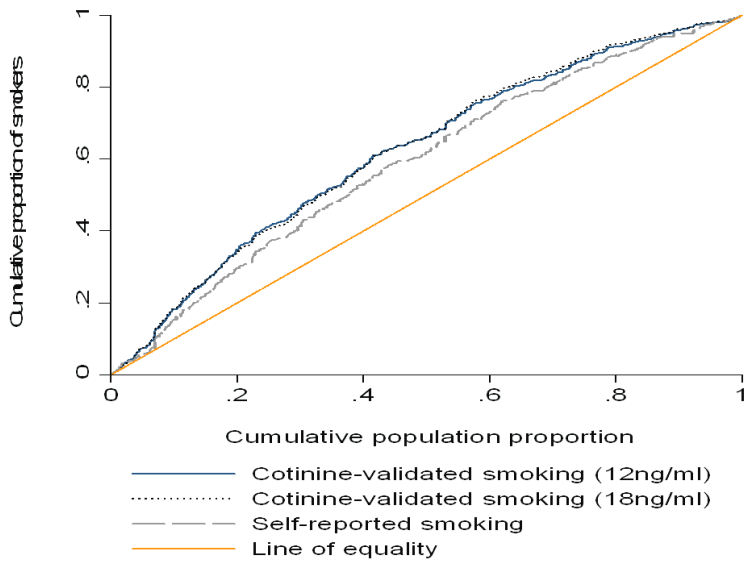

Fig. 4 ***Concentration curves of smoking participation***

4. Conclusion

In this study we show that misreporting of smoking behaviour by adolescents has important consequences in empirical studies estimating the impact of various factors on adolescent smoking participation. The failure to observe an effect of parental income on adolescent self-reported smoking in previous empirical studies may be due to variations in the extent of misreporting errors within available data. Similarly, misreporting of smoking participation is likely to bias estimates of the impact of various anti-tobacco policies and may explain the divergent results often reported in empirical studies on the impact of anti-tobacco policies such as the effect of price, on adolescent smoking (for example see Rice et al., 2009). Our results suggest that misclassification errors in smoking participation models will result in an underestimation of the effects of observable characteristics as well as understate the precision of the estimates. In addition, misclassification errors may have important consequences when measuring income-related inequality in self-reported smoking, resulting in an underestimation of the extent of inequality in smoking. From a policy maker's perspective, this represents a serious problem when comparing inequality in smoking between populations or across time periods.

In this study we use saliva cotinine levels greater than 12ng/ml as the true measure of smoking participation. However, cotinine assays do not always represent a gold standard for determining true smoking status particularly in adolescents in whom high variations of saliva cotinine may occur due to variations in the pattern of cigarette consumption or due to exposure to second-hand smoke (Patrick et al., 1994). Nevertheless, we show that varying the cut-point does not affect estimates of the effect of parental income and other observable characteristics on smoking participation.

Although cotinine assay may produce more reliable data on smoking participation, collecting such data is not without its problems. It is expensive, more difficult to obtain and may result in large proportions of missing data if participants are unable to provide sufficient saliva samples. Self-reported smoking questionnaires, on the other hand, are easier and cheaper to administer and represent an important method of collecting data on smoking behaviour in population surveys. Therefore there is the need to develop suitable methods or to adapt existing methods to account for misclassification errors in empirical studies on self-reported smoking participation in adolescents. This will ensure that the impact of observable characteristics including anti-tobacco smoking policies are consistently estimated using available data.

Appendix

$$\Pr(Y^{SR}=1|X) = \Pr(Y^{SR}=1, Y^{CV}=1|X) + \Pr(Y^{SR}=1, Y^{CV}=0|X)$$

$$= \Pr(Y^{SR}=1|\, Y^{CV}=1, X) * \Pr\,(Y^{CV}=1|X) + \Pr(Y^{SR}=1|\, Y^{CV}=0, X) * \Pr\,(Y^{CV}=0|X)$$

$$= \Pr(Y^{SR}=1|\, Y^{CV}=1) * \Pr\,(Y^{CV}=1|X) + \Pr(Y^{SR}=1|\, Y^{CV}=0) * \Pr\,(Y^{CV}=0|X)$$

$$= (1 - p10)\Pr(Y^{CV}=1|X) + p01\,(1 - \Pr(Y^{CV}=1|X))$$

$$\{where\ p01 = \Pr(Y^{SR}=1\big|Y^{\mathrm{CV}}=0\big) \quad and \quad p10 = \Pr(Y^{SR}=0\big|Y^{\mathrm{CV}}=1\big)\}$$

$$= p01 + (1 - p01 - p10)\Pr(Y^{CV}=1|X)$$

References

Adams J, Parkinson L, Sanson-Fisher RW, Walsh RA. Enhancing self-report of adolescent smoking: The effects of bogus pipeline and anonymity. Addictive Behaviors 2008;33; 1291-1296.

Artís M, Ayuso M, Guillén M. Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims. Journal of Risk and Insurance 2002;69; 325-340.

Bauman KE, Ennett SE. Tobacco use by black and white adolescents: the validity of self-reports. American Journal of Public Health 1994;84; 394-398.

Benowitz NL. Cotinine as a Biomarker of Environmental Tobacco Smoke Exposure. Epidemiologic Reviews 1996;18; 188-204.

Blow L, Leicester A, Windmeijer F. Parental Income and Children's Smoking Behaviour: Evidence from the British Household Panel Survey. Institute of Fiscal Studies Working Paper Series. Institute of Fiscal Studies London; 2005.

Bound J, Brown C, Mathiowetz N. Chapter 59 Measurement error in survey data. In: Heckman JJ, Leamer E (Eds), Handbook of Econometrics, vol. Volume 5. Elsevier; 2001.

Craig R, Mindell J. Health Survey for England 2006. Vol 2: Obesity and other risk factors in children. The Information Centre; 2008.

Department of Health. A Smokefree Future: A Comprehensive Tobacco Control Strategy for England. London; 2010.

Dolcini M, Adler N, Ginsberg D. Factors influencing agreement between self-reports and biological measures of smoking among adolescents. Journal of Research on Adolescence 1996;6; 515–542.

Erreygers G. Correcting the Concentration Index. Journal of Health Economics 2009;28; 504-515.

Falaris EM. Misclassification of the dependent variable in binary choice models: evidence from five Latin American countries. Applied Economics 2011;43; 1315 - 1327.

Friedman L, Johnson B, Brett A. Evaluation of substance-abusing adolescents by primary care physicians. Journal of Adolescent Health Care 1990;11; 227-230.

Fuller E. Smoking, drinking and drug use among young people in England in 2008. The NHS Information Centre for Health and Social Care; 2009.

Griesler PC, Kandel DB, Schaffran C, Hu M-C, Davies M. Adolescents' Inconsistency in Self-Reported Smoking. Public Opinion Quarterly 2008;72; 260-290.

Gruber J, Zinman J. Youth Smoking in the United States: Evidence and Implications. In: Gruber J (Ed), Risky Behavior Among Youths: An Economic Analysis. The University of Chicago Press; 2001.

Hausman J. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. Journal of Economic Perspectives 2001;15; 57-67.

Hausman JA, Abrevaya J, Scott-Morton FM. Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics 1998;87; 239-269.

Höfler M. The effect of misclassification on the estimation of association: a review. International Journal of Methods in Psychiatric Research 2005;14; 92-101.

Jarvis MJ, Fidler J, Mindell J, Feyerabend C, West R. Assessing smoking status in children, adolescents and adults: cotinine cut-points revisited. Addiction 2008;103; 1553-1561.

Kakwani N, Wagstaff A, van Doorslaer E. Socioeconomic inequalities in health: measurement, computation, and statistical inference. Journal of Econometrics 1997;77; 87 - 103.

Kandel DB, Schaffran C, Griesler PC, Hu M-C, Davies M, Benowitz N. Salivary Cotinine Concentration Versus Self-Reported Cigarette Smoking: Three Patterns of Inconsistency in Adolescence. Nicotine & Tobacco Research 2006;8; 525-537.

Kenkel DS, Lillard DR, Mathios AD. Accounting for misclassification error in retrospective smoking data. Health Economics 2004;13; 1031-1044.

Lopez A. The evolution of the Global Burden of Disease framework for disease, injury and risk factor quantification: developing the evidence base for national, regional and global public health action. Globalization and Health 2005;1; 5.

Magder LS, Hughes JP. Logistic Regression When the Outcome Is Measured with Uncertainty. American Journal of Epidemiology 1997;146; 195-203.

McClements D. Equivalence scales for children. Journal of Public Economics 1977;8; 191-210.

Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. American Journal of Public Health 1994;84; 1086-1093.

Rice N, Godfrey C, Slack R, Sowden A, Worthy G. A systematic review of the effects of price on the smoking behaviour of young people. Public Health Research Consortium; 2009.

Soteriades ES, DiFranza JR. Parent's Socioeconomic Status, Adolescents' Disposable Income, and Adolescents' Smoking Status in Massachusetts. American Journal of Public Health 2003;93; 1155-1160.

Stein L, Colby S, O'Leary T, Monti P, Rohsenow D, Spirito A, Riggs S, Barnett N. Response Distortion in Adolescents Who Smoke: a Pilot Study. Journal of Drug Education 2002;32; 271–286.

The NHS Information Centre. Statistics on Smoking: England http://www.ic.nhs.uk/pubs/smoking10 2010.

Tyas SL, Pederson LL. Psychosocial factors related to adolescent smoking: a critical review of the literature. Tobacco Control 1998;7; 409-420.

Wagenknecht LE, Burke GL, Perkins LL, Haley NJ, Friedman GD. Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. American Journal of Public Health 1992;82; 33-36.

Wagstaff A. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Economics 2005;14; 429 - 432.

Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. Social Science and Medicine 1991;33; 545 - 557.

Table 1 *Description of Variables*

Variable name	Variable Label
Cotinine≥12ng/ml	1 if cotinine ≥12ng/ml
Cotinine≥18ng/ml	1 if cotinine ≥18ng/ml
SR Smoking	1 if a self-reported smoker
Under-reporting	1 if cotinine12=1 and SR smoker=0
Age	Age last birthday
Age squared	Age squared
Male	1 if male
White	1 if white
Household characteristics	
Income	Annual total household income
Log Income	Log household income
HH Social class 1 &2 (Base group)	1 if household head(HH) is professional, managerial and technical occupations
HH Social class 3	1 if skilled occupation (manual & non-manual) or armed services
HH Social class 4&5	1 if partly skilled/unskilled/students/never occupied/unclassifiable
Own a house	1 if owns outright/mortgage/shared ownership(part rent part mortgage)
Urban (Base group)	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
Smoke in home	1 if anyone smokes inside the house/flat on most days
Mother's highest qualification	
M Degree (Base group)	1 if NVQ4/NVQ5/Degree or equivalent
M Below degree	1 if higher education below degree
M NVQ3/A levels	1 if NVQ3/GCE A level equivalent
M NVQ2/O levels	1 if NVQ2/GCE O level or equivalent
M NVQ1/CSE	1 if NVQ1/CSE/other grade equivalent/foreign or other qualification
M No qualification	1 if no qualification
Father's highest qualification	
F Degree (Base group)	1 if NVQ4/NVQ5/Degree or equivalent
F Below degree	1 if higher education but below degree
F NVQ3/A levels	1 if NVQ3/GCE A level equivalent
F NVQ2/O levels	1 if NVQ2/GCE O level or equivalent
F NVQ1/CSE	1 if NVQ1/CSE/other grade equivalent/foreign or other qualification
F No qualification	1 if no qualification
Mother's employment status	
M Employed (Base group)	1 if in paid employment or self employed
M Unemployed	1 if unemployed
M Keep home	1 if looking after home/family or doing something else
M Sick/Retired	1 if retired/permanently unable to work due to long-term sickness or disability
M Student	1 if in full time education/ on a government training scheme
Father's employment status	
F Employed (Base group)	1 if in paid employment or self employed
F Unemployed	1 if unemployed
F Keep home	1 if looking after home/family or doing something else
F Retired	1 if retired
F Sick	1 if permanently unable to work due to long-term sickness or disability
F Student	1 if in full time education/ on a government training scheme

Table 1 *Continued*

Variable name	Variable Label
Parents current smoking status	
M current smoker	1 if mother smokes
F current smoker	1 if father smokes
M & F current smokers	1 if both parents smoke
Parents marital status	
Parent married	1 if parents are married/cohabiting
Parent single (Base group)	1 if parent is single
Survey years	
1997-2001 (Base group)	1 if year of survey is from 1997-2001
2002-2005	1 if year of survey is from 2002-2005
2006-2008	1 if year of survey is from 2006-2008
Parents missing variable indicator	
Single mum	1 if father information are missing because mother is single
Single dad	1 if mother information are missing because father is single
F not home	1 if father information missing because father not home
M not home	1 if mother information missing because mother not home
Both parent missing	1 if information on both parents are missing

HH- Head of household

Table 2 *Descriptive statistics of full sample and cotinine-validated sub-sample*

	Full Sample						CV Sub-sample
	All	Income Quintiles					All
Variable		1	2	3	4	5	
Adolescents Characteristics							
Cotinine≥12ng/ml	0.0927	0.162	0.109	0.0805	0.0647	0.0408	-
Cotinine≥18ng/ml	0.0864	0.148	0.105	0.0794	0.0598	0.0351	-
SR Smoking	0.064	0.0939	0.077	0.0624	0.0483	0.0354	0.503
Age	12.94	12.88	12.96	12.97	12.95	12.97	14.07
Age squared	169.60	167.80	170.00	170.10	169.80	170.30	199.10
Male	0.502	0.510	0.493	0.511	0.507	0.489	0.467
White	0.902	0.814	0.890	0.931	0.947	0.933	0.933
Household Characteristics							
Income	20386.40	5603.90	10566.40	15947.30	23650.00	47754.40	14903.90
Log Income	9.634	8.568	9.256	9.671	10.060	10.710	9.324
HH Social class 1 &2	0.394	0.110	0.191	0.339	0.549	0.814	0.243
HH Social class 3	0.386	0.421	0.485	0.490	0.364	0.161	0.397
HH Social class 4&5	0.220	0.468	0.324	0.172	0.087	0.026	0.361
Own a house	0.703	0.329	0.597	0.767	0.910	0.945	0.523
Urban	0.298	0.312	0.305	0.289	0.275	0.307	0.269
Suburb	0.490	0.538	0.501	0.505	0.488	0.418	0.539
Rural	0.212	0.150	0.194	0.207	0.238	0.276	0.192
Smoke in home	0.343	0.547	0.445	0.323	0.230	0.153	0.620
Mother's highest qualification							
M Degree	0.104	0.021	0.032	0.064	0.131	0.283	0.048
M Below degree	0.074	0.033	0.050	0.073	0.106	0.114	0.036
M NVQ3/A levels	0.086	0.048	0.072	0.097	0.116	0.100	0.067
M NVQ2/O levels	0.230	0.181	0.264	0.290	0.256	0.159	0.233
M NVQ1/CSE	0.058	0.077	0.081	0.059	0.041	0.029	0.078
M No qualification	0.112	0.224	0.171	0.081	0.046	0.029	0.174
Father's highest qualification							
F Degree	0.106	0.016	0.029	0.069	0.136	0.290	0.026
F Below degree	0.079	0.027	0.061	0.093	0.116	0.103	0.065
F NVQ3/A levels	0.065	0.025	0.054	0.060	0.106	0.080	0.029
F NVQ2/O levels	0.109	0.065	0.109	0.158	0.140	0.076	0.088
F NVQ1/CSE	0.031	0.032	0.031	0.046	0.030	0.018	0.029
F No qualification	0.085	0.121	0.132	0.083	0.052	0.030	0.115
Mother's Employment Status							
M Employed	0.465	0.179	0.432	0.533	0.609	0.591	0.392
M Unemployed	0.010	0.023	0.008	0.006	0.006	0.005	0.010
M Keep home	0.158	0.325	0.191	0.102	0.062	0.099	0.194
M Sick/Retired	0.018	0.036	0.028	0.014	0.011	0.002	0.024
M Student	0.014	0.020	0.013	0.009	0.009	0.017	0.015
Father's Employment Status							
F Employed	0.417	0.133	0.336	0.491	0.565	0.583	0.275
F Unemployed	0.013	0.045	0.012	0.004	0.002	0.002	0.023
F Keep home	0.013	0.028	0.026	0.003	0.003	0.002	0.019

Table 2 *Continued*

	Full Sample						CV Sub-sample
	All	Income Quintiles					All
Variables		1	2	3	4	5	
F Retired	0.005	0.008	0.007	0.002	0.001	0.005	0.004
F Sick	0.021	0.064	0.029	0.004	0.004	0.000	0.028
F Student	0.006	0.007	0.008	0.005	0.005	0.003	0.006
Parents current smoking status							
M current smoker	0.197	0.286	0.242	0.195	0.164	0.092	0.364
F current smoker	0.132	0.143	0.167	0.138	0.123	0.089	0.165
M & F current smokers	0.064	0.067	0.080	0.064	0.066	0.040	0.099
Parents marital status							
Parent married	0.518	0.302	0.448	0.549	0.631	0.678	0.372
Parent single	0.172	0.308	0.250	0.143	0.085	0.059	0.297
Survey Years							
1997-2001	0.389	0.511	0.431	0.388	0.341	0.263	0.443
2002-2005	0.420	0.376	0.398	0.421	0.461	0.451	0.421
2006-2008	0.190	0.113	0.171	0.191	0.198	0.286	0.136
Parents missing indicator							
Single mum	0.150	0.283	0.225	0.120	0.064	0.045	0.258
Single dad	0.016	0.019	0.016	0.018	0.013	0.012	0.032
F not home	0.066	0.042	0.056	0.064	0.072	0.096	0.058
M not home	0.009	0.009	0.010	0.010	0.007	0.010	0.003
Both parent missing	0.310	0.390	0.302	0.308	0.284	0.264	0.330
N	***7421***	***1490***	***1488***	***1474***	***1484***	***1485***	***694***

Table 3 *Bivariate probit models of misreporting smoking*

	Under-reporting (*p10*)		Over-reporting (*p01*)	
	$Pr(Y^{SR}=0 \mid Y^{CV}=1)$		$Pr(Y^{SR}=1 \mid Y^{CV}=0)$	
Variables	**ME**	***t*-stats**	**ME**	***t*-stats**
Log Income	-0.0280	-1.21	0.0009	0.52
Age	0.0835	0.29	0.0089	0.46
Age squared	-0.0063	-0.59	0.0001	0.09
Male	0.0380	1.46	-0.0035	-1.94
White	-0.0185	-0.38	0.0040	1.42
2002-2005	0.0276	0.99	-0.0021	-1.09
2006-2008	**0.1191**	**2.66**	**-0.0084**	**-3.42**
HH Social class 3	0.0536	1.54	-0.0032	-1.32
HH Social class 4&5	0.0448	1.12	-0.0011	-0.39
Own a house	**0.0787**	**2.37**	**-0.0089**	**-2.86**
Suburb	0.0518	1.49	-0.0031	-1.23
Rural	-0.0043	-0.1	0.0006	0.2
Smoke in home	-0.0183	-0.53	**0.0069**	**2.32**
M Below degree	-0.0211	-0.27	0.0001	0.02
M NVQ3/A levels	0.0452	0.66	-0.0024	-0.55
M NVQ2/O levels	0.0729	1.22	-0.0044	-1.17
M NVQ1/CSE	0.0322	0.49	-0.0013	-0.29
M No qualification	0.0555	0.81	-0.0030	-0.71
F Below degree	-0.0663	-0.81	0.0117	1.19
F NVQ3/A levels	-0.0451	-0.49	0.0046	0.56
F NVQ2/O levels	-0.1584	-1.95	0.0222	1.77
F NVQ1/CSE	0.0030	0.03	0.0028	0.3
F No qualification	-0.0664	-0.76	0.0112	1.1
M Unemployed	0.1216	1	-0.0072	-1.43
M Keep home	-0.0197	-0.48	0.0007	0.24
M Sick/Retired	0.0814	0.64	-0.0051	-0.8
M Student	**-0.2436**	**-2.48**	0.0354	1.72
F Unemployed	0.0220	0.21	-0.0002	-0.02
F Keep home	**0.2200**	**2.39**	**-0.0100**	**-3.29**
F Retired	0.0197	0.09	-0.0032	-0.27
F Sick	-0.1805	-1.95	0.0169	1.32
F Student	0.1819	1.36	-0.0087	-1.97
M current smoker	0.0314	0.66	-0.0001	-0.03
F current smoker	-0.0180	-0.26	-0.0001	-0.03
M & F current smokers	0.0082	0.1	-0.0003	-0.04
Parent married	0.2473	1.77	**-0.0267**	**-2.1**
N	*7421*		*7421*	

Marginal effects (ME) calculated at sample mean. Standard errors adjusted for clustering at primary sample unit level. Omitted categories: HH social class 1&2 and urban for household characteristics; degree or equivalent for parents' qualification; employed for parents' employment status; single for parents' marital status; and 1997-2001 for survey year.

Table 4a *Probit model of cotinine-validated and self-reported (SR) smoking*

	SR smoker		Cotinine≥12ng/ml		SR smoker		Cotinine≥12ng/ml	
Variables	**ME**	***t*-stats**	**ME**	***t*-stats**	**ME**	***t*-stats**	**ME**	***t*-stats**
Log Income	**-0.0198**	-7.09	**-0.0424**	-10.52	-0.0054	-1.83	**-0.0134**	-3.34
Age	0.0701	1.74	**0.1474**	3.05	0.0663	1.74	**0.1374**	3.18
Age squared	-0.0014	-0.93	**-0.0039**	-2.11	-0.0014	-0.94	**-0.0036**	-2.2
Male	**-0.0081**	-2.11	-0.0101	-1.89	**-0.0076**	-2.2	**-0.0095**	-1.99
White	**0.0162**	2.78	**0.0345**	4.76	**0.0118**	2.04	**0.0235**	3.32
2002-2005	-0.0016	-0.38	0.0028	0.43	-0.0034	-0.89	0.0006	0.11
2006-2008	**-0.0164**	-3.41	**-0.0182**	-2.59	**-0.0168**	-3.43	-0.0132	-1.8
HH Social class 3					-0.0037	-0.78	0.0084	1.19
HH Social class 4&5					0.0045	0.74	**0.0283**	2.67
Own a house					**-0.0251**	-4.18	**-0.0290**	-3.97
Suburb					-0.0031	-0.6	0.0055	0.82
Rural					0.0002	0.03	0.0004	0.05
Smoke in home					**0.0310**	5.96	**0.0648**	8.6
R-squared	*0.160*		*0.157*		*0.190*		*0.204*	
N	*7421*		*7421*		*7421*		*7421*	

Marginal effects (ME) calculated at sample mean. Standard errors adjusted for clustering at primary sample unit level. Omitted categories: HH social class 1&2 and urban for household characteristics; degree or equivalent for parents' qualification; employed for parents' employment status; single for parents' marital status; and 1997-2001 for survey year.

Table 4b Extended Probit model of cotinine-validated and self-reported (SR) smoking

	SR smoker		Cotinine≥12ng/ml		Cotinine≥18ng/ml	
Variables	ME	*t*-stats	ME	*t*-stats	ME	*t*-stats
Log Income	-0.002	-0.74	**-0.008**	-2.05	**-0.009**	-2.29
Age	0.062	1.8	**0.133**	3.23	**0.122**	3.17
Age squared	-0.001	-0.97	**-0.003**	-2.24	**-0.003**	-2.19
Male	**-0.007**	-2.05	-0.008	-1.68	-0.006	-1.49
White	**0.010**	1.98	**0.021**	3.05	**0.019**	2.98
2002-2005	-0.004	-1.19	-0.001	-0.15	-0.002	-0.46
2006-2008	**-0.016**	-3.6	-0.012	-1.72	-0.010	-1.51
HH Social class 3	-0.004	-0.95	0.005	0.71	0.006	0.92
HH Social class 4&5	0.002	0.4	**0.021**	2.14	**0.019**	2.08
Own a house	**-0.019**	-3.56	**-0.021**	-3.08	**-0.017**	-2.78
Suburb	-0.003	-0.69	0.004	0.66	0.005	0.75
Rural	0.002	0.35	0.003	0.32	0.005	0.67
Smoke in home	**0.027**	4.35	**0.053**	5.8	**0.047**	5.4
M Below degree	-0.003	-0.29	-0.009	-0.86	-0.008	-0.81
M NVQ3/A levels	-0.001	-0.14	0.008	0.57	0.005	0.37
M NVQ2/O levels	-0.006	-0.81	0.005	0.41	0.006	0.6
M NVQ1/CSE	0.004	0.38	0.012	0.73	0.013	0.85
M No qualification	-0.003	-0.35	0.007	0.55	0.007	0.6
F Below degree	0.033	1.75	**0.050**	2.16	**0.054**	2.28
F NVQ3/A levels	0.005	0.41	0.006	0.36	0.012	0.67
F NVQ2/O levels	**0.042**	2.14	0.031	1.62	0.038	1.94
F NVQ1/CSE	0.014	0.69	0.029	1.18	0.025	1.1
F No qualification	0.026	1.48	0.044	1.95	**0.048**	2.12
M Unemployed	-0.011	-0.85	-0.007	-0.33	-0.004	-0.19
M Keep home	-0.001	-0.21	-0.004	-0.61	-0.005	-0.72
M Sick/Retired	-0.013	-1.19	-0.008	-0.55	-0.009	-0.65
M Student	**0.055**	2.01	0.030	1.15	0.021	0.96
F Unemployed	0.006	0.39	0.013	0.63	0.017	0.82
F Keep home	**-0.019**	-2.16	0.016	0.55	-0.009	-0.55
F Retired	-0.007	-0.37	-0.018	-1.05	-0.015	-0.95
F Sick	0.024	1.26	-0.006	-0.42	-0.002	-0.17
F Student	-0.012	-0.89	0.010	0.25	-0.019	-0.99
M current smoker	0.006	0.8	0.020	1.92	0.016	1.67
F current smoker	-0.003	-0.39	-0.011	-1.01	-0.013	-1.44
M & F current smokers	0.002	0.14	0.005	0.31	0.009	0.56
Parent married	**-0.045**	-2.08	-0.048	-1.38	-0.020	-0.68
R-squared	*0.209*		*0.219*		*0.219*	
N	*7421*		*7421*		*7421*	

Table 5 Probit model for under-reporting of smoking behaviour

Variables	ME	*t*-stats	ME	*t*-stats	ME	*t*-stats
Log Income	0.0385	1.34	-0.0103	-0.29	-0.0228	-0.6
Age	-0.6011	-1.22	-0.6428	-1.27	-0.6382	-1.23
Age squared	0.0165	0.91	0.0177	0.96	0.0172	0.91
Male	0.0475	1.18	0.0442	1.1	0.0433	1.03
White	0.0558	0.72	0.0371	0.46	0.0375	0.44
2002-2005	-0.0379	-0.82	-0.0247	-0.52	-0.0280	-0.56
2006-2008	0.0661	1.03	0.1187	1.56	0.1297	1.55
HH Social class 3			0.0233	0.43	-0.0076	-0.13
HH Social class 4&5			0.0373	0.62	0.0035	0.06
Own a house			**0.1247**	2.39	0.0890	1.61
Suburb			0.0695	1.19	0.0718	1.2
Rural			0.0447	0.6	0.0244	0.31
Smoke in home			-0.0701	-1.55	-0.0874	-1.57
M Below degree					-0.0751	-0.53
M NVQ3/A levels					-0.0567	-0.43
M NVQ2/O levels					0.0301	0.28
M NVQ1/CSE					-0.2187	-1.66
M No qualification					0.0058	0.05
F Below degree					0.0970	0.62
F NVQ3/A levels					0.1867	1.02
F NVQ2/O levels					0.1244	0.84
F NVQ1/CSE					0.2117	1.15
F No qualification					0.2146	1.45
M Unemployed/keep home					-0.0291	-0.43
M Sick/Retired					**0.3608**	2.02
M Student					**-0.6045**	-2.9
F Unemployed					-0.0721	-0.47
F Keep home					0.1253	0.61
F Retired					0.2427	0.64
F Sick/Student					-0.1729	-1.31
M current smoker					-0.0182	-0.24
F current smoker					0.0633	0.58
M & F current smokers					-0.1584	-1.15
Parent married					0.3002	1.35
R-squared	*0.06577*		*0.07895*		*0.1241*	
N	*694*		*694*		*694*	

Marginal effects (ME) calculated at sample mean. Standard errors adjusted for clustering at primary sample unit level. Omitted categories: HH social class 1&2 and urban for household characteristics; degree or equivalent for parents' qualification; employed for parents' employment status; single for parents' marital status; and 1997-2001 for survey year

Table 6 *Corrected concentration index of smoking participation*

Year	SR smoking	Cotinine>12ng/ml	Cotinine>18ng/ml	N
1997-2001	-0.0316**	-0.0837***	-0.0785***	3149
2002-2005	-0.0505***	-0.0890***	-0.0865***	2783
2006-2008	-0.0487***	-0.0867***	-0.0791***	1489
All years	-0.0468***	-0.0904***	-0.0855***	7421

** p<0.01, *** p<0.001