The contribution of smoking and obesity to socioeconomic inequalities in health in England

Laura Vallejo-Torres, HERG, Brunel University, Laura.Vallejo-Torres@brunel.ac.uk
Steve Morris, HERG, Brunel University, Stephen.Morris@brunel.ac.uk

Abstract

Aims: (1) To measure socioeconomic inequalities in health over a period of 9 years and across regions of England. (2) To quantify the contribution of smoking and obesity to this inequality.
Methods: Data for the analysis were taken from nine rounds of the Health Survey for England (HSE; 1998–2006; *n*=100,317). We construct a health variable using the predicted values from an interval regression model regressing self-assessed health against a comprehensive set of health indicators. EQ-5D scores on the self-assessed health categories are used to set the cut-points. We use a concentration index approach to measure inequality and decomposition methods to quantify the contribution of smoking and obesity for each region of England. We use instrumental variables (IV) regression to account for the endogeneity of smoking and obesity. Standard errors of the estimated contributions are calculated using bootstrapping techniques.
Results: Income related health inequalities have increased over time and vary by region. Inequality is lowest in London and the South and highest in the North. After allowing for endogeneity, smoking explains 9% of the income-related inequality in health nationally, while the contribution of obesity is less than 1%. These are statistically significant contributions to health inequality that vary across areas.
Conclusions: Income-related health inequality is increasing over time and the extent of it varies between areas. Smoking and obesity make a significant contribution to the extent of health inequality, with smoking playing the larger role.

Introduction

Tackling health inequalities is a top priority for the British Government. Evidence shows that although Britain is now healthier than ever as a nation, the gap in health status between those at the top and bottom ends of the socioeconomic scale remains large and in some areas it continues to widen [1]. Socioeconomic inequalities in health are partly explained by differences in lifestyle and living conditions, and lifestyle can vary between socioeconomic groups. In this paper we consider the impact of two important lifestyle factors on income related inequality in health in England.

Evidence suggests that smoking is the primary reason for the gap in healthy life expectancy between the rich and poor [2]. Among men, smoking is responsible for over half the excess risk of premature death between social classes [3].This is due to the negative impact of smoking on health and the socioeconomic gradient in smoking prevalence. Smoking is responsible for one-sixth of all deaths in the UK, accounting directly for 87,000 deaths a year in England alone [4]. Between 1998 and 2006 adult smoking rates in England fell from 28 percent to 22 percent but over 9 millions of adults in England still smoke [5]. The social gradient of smoking has been most frequently documented by occupational status, showing that in 2006, the prevalence of cigarette smoking was 15% in professional and managerial groups, and 29% in routine and manual groups [6]. Decreases in smoking rates are higher among those with higher socioeconomic status, which may exacerbate health inequalities in the future.

Obesity is an increasing lifestyle problem in many countries. In England in 1980 6% of males and 8% of females were obese; by 2005 these prevalences had increased to 24% and 28%, respectively [7]. Obesity is a debilitating condition in its own right and an important risk factor for a number of diseases including coronary heart disease, type II diabetes, hypertension and stroke [8]. In England 7% of all deaths are attributable to obesity [9]. Obesity and its risks are also not spread equally across society. For adults, the gradient between socioeconomic group and obesity is more pronounced for women than it is for men, with a prevalence of obesity of 19% for women in the highest income quintile compared 32% in the lowest quintile [10].

Given the importance of these two problems in terms of their impact on public health, plus their uneven distribution across socioeconomic groups, and that they are thought to be avoidable, it is unsurprising that reducing smoking and obesity in deprived communities are believed to be effective interventions for reducing inequalities in health [11].

Previous literature has attempted to measure the contribution of lifestyle factors, including smoking and obesity, to socioeconomic inequalities in health by looking at the change in the influence of socioeconomic factors on health before and after controlling for the effect of lifestyles. Borg and Kirstensen (2000) [12] using Danish data estimated odd ratios for reporting a reduction in self-assessed health (SAH) in 1995 versus 1990 across social class groups. They found that 17% of the odds ratio for the highest social class group relative to the lowest was explained by tobacco consumption and obesity, after controlling for age, gender and illness. Tobacco and obesity separately accounted for 4% and 12% of the odds ratio, respectively. Lantz et al. (2001) [13] looked at changes in various health measures in a US panel of individuals in 1994 compared with those in 1986 and used multinomial logistic regression to investigate the change in the odd ratios of the observed socioeconomic gradient after including risky health behaviours. They found that lifestyles explained a small part of the association between socioeconomic status and health status. Contoyannis and Jones (2004) [14] used similar indicators but accounted for unobserved heterogeneity using the 1984 and 1991 Health and Lifestyle Surveys (HALS) and found that lifestyle factors (smoking, drinking, breakfast, sleep patterns, obesity, sporting activities) explained a large proportion of the relation between social class and health, accounting for a 53% reduction of the social class gradient after accounting for the endogeneity of lifestyle variables. In a more recent study, instead of investigating extremes of the socioeconomic distribution Balia and Jones (2008) [15] look at the whole distribution and compute and decompose the Gini coefficient of total health inequality. They looked at the reduction of the contribution of socioeconomic variables when lifestyle factors (similar to those used by Contoyannis and Jones, 2004) were included. After accounting for endogeneity they found that the contributions of social class, education and age diminish considerably after controlling for lifestyle variables, which explain 25% of overall inequality in health.

Robust methods for the measurement and the decomposition by contribution factors of socioeconomic inequalities in health have been developed in the literature [16, 17, 18, 19]. They allow for the quantification of inequalities in health with respect to socioeconomic status, measured by the concentration index of income-related inequality. They also allow the concentration index to be decomposed by factor to find the variables that play a significant role in explaining inequality. In this paper we use these methods to explore in details the contribution of smoking and obesity to income-related inequalities in health. We undertake a decomposition analysis, which allows us to investigate whether the contribution of lifestyle variables is mainly due to the distribution of risky behaviours between income groups, or to the impact of these behaviours on health. Following Contoyannis and Jones (2004) and Balia and Jones (2008) an important issue when investigating the role of lifestyle factors in health inequality is the endogeneity of lifestyle. Hence, the aim of this paper is to quantify the contribution of smoking and obesity to income related inequality in health in England accounting for the potential endogeneity of smoking and obesity. Our analysis has a number of components. First, we construct a continuous health variable for use in our analysis using a recently proposed method [20]. Second, we calculate income related inequality in health by area (Government Office Region) of England and by year using a concentration index approach. Third, we decompose these measures into area components to determine whether the overall inequality is due to gaps between richer and poorer areas in England or due to differences between rich and poor individuals within these areas or both. Fourth, we quantify the contribution of smoking and obesity to income-related inequality in health using a decomposition approach. In this last component we use instrumental variables (IV) regression to account for the endogeneity of smoking and obesity. We also repeat the analysis for each area in our dataset.

Methods

Data

The analysis is based on data from nine rounds (1998-2006) of the *Health Survey for England* (HSE). The HSE is a cross-sectional representative national survey which draws a different sample every year of individuals living in England. Respondents are interviewed on a range of topics including their age, their socioeconomic status, their health status, and individual lifestyle. We included data for individuals over the age of 16 year-old. The total sample size for the pooled sample is 100,317 individuals.

Measurement of health

We construct a continuous health variable using the approach proposed and validated by van Doorslaer and Jones [20]. Following this approach, respondents' self-assessed health (SAH) is regressed against age, gender and a detailed set of health indicators using an interval regression model. This model provides a more efficient alternative to the ordered probit model when the values of the cut-points of the ordered categories are known. The cut-points were computed using information on the empirical distribution of EQ-5D scores (www.euroqol.org) and mapping these to the cumulative frequency of the SAH categories. The assumption underpinning this method is that there is a stable mapping from EQ-5D scores to the (latent) variable that determines SAH. Our health measure is predicted health status based on the predictions from the interval regression model. The predictions provide a continuous health variable with an upper bound of unity (representing full health) and a value of zero being equivalent to death. Negative values represent health states worse than death. The SAH variable is based on responses to the HSE question: 'How is your health in general? Would you say it was… very good, good, fair, bad or very bad?'. The predictors used in the interval regression are: gender; a cubic function of individual age; interactions between age and gender; whether or not the individual has one of 14 longstanding illnesses by broad disease category; whether or not these longstanding illnesses limit the activity of the respondent in any way; the number of longstanding illnesses; the number of days the respondent had to cut down on their usual activities in the previous two weeks because of illness or injury; and GHQ-12 score (a measure of psychosocial health on a scale from 0 to 12 with higher values indicating greater psychosocial disorder).

In terms of the calibration of the SAH variable, information on the empirical distribution of EQ-5D scores was not available in every year of the data; it is not included in the 1998-2002 rounds of HSE. When the variable was missing we used the distribution of EQ-5D scores from the nearest year of data available; SAH in 1998 and 1999 is calibrated using the 1996 HSE EQ-5D information; for 2000 to 2002, we used the information from the 2003 HSE. Pooling the data, the cumulative percentage of the respondents reporting very bad to very good health are 1.32%, 5.84%, 23.43%, 65.75% where the EQ-5D variable was available. The corresponding values from the empirical distribution of EQ-5D values are 0.016, 0.516, 0.796, and 1.000, respectively. These values are used to set the cut-points in the interval regression analysis for the whole pooled sample. Using these cut-points we then regressed the SAH variable against the indicators listed above and predicted health status for every individual observation in the pooled dataset.

Measurement of inequality

We use the health concentration index (CI) derived from the concentration curve as our measure of relative income-related health inequality [19]. With a continuous measure of health y_i the concentration curve $L(s)$ plots the cumulative proportion of the population ranked by income against the cumulative proportion of total health. If everyone enjoys the same health $L(s)$ coincides with the diagonal. If $L(s)$ lies below (above) the diagonal then inequalities in health exist and favour the rich (poor). The CI is defined as twice the area between the concentration curve and the diagonal. It takes a value of zero when $L(s)$ coincides with the diagonal and a positive (negative) value when there is income-related health inequality favouring the rich (poor). We use the convenient regression approach developed by Kakwani, Wagstaff and van Doorslaer (1997) to compute the CIs [17].

Decomposing inequality into between- and within- area components

We decompose the income-related health inequality measure into between- and within-area components. This provides insights into whether the national inequalities are mainly due to gaps between rich and poor areas in England, or are due to differences between rich and poor individuals within these areas, or both. The formula developed for the decomposition of the Gini coefficient [21], and applied to inequality measures of various health variables with respect to income [22; 23; 24, 25], is

$$CI = CI_b + \sum_j \alpha_j CI_j + R \qquad (1)$$

where CI_b is the between-area CI , α_j is the weight attached to the *jth* area (the product of its population share and health variable share), CI_j is the CI of the *jth* area and *R* is a residual term. We compute CI_b by replacing individual health with the mean health of the area where the individual lives in and applying the individual sample weight of the survey. (We also compute CI_b using a second approach by creating area level data with the mean health and mean income for each area and using the share of the population in the sample of each area as the weighting variable.) The residual term *R* takes a non-zero value when there is overlapping in the income variable of the richest individual in poorer areas and the poorest individual in richer areas.

To provide a ranking of areas that consider both the inequality as well as the level of health in an area, we also calculate the Health Achievement Index for every area, which is defined as $(1-CI)*\mu$, where μ is the mean health of the area, and based on an inequality aversion index equal to unity [26].

Decomposing inequality by factors – the contribution of smoking and obesity

Based on an additive linear regression model of health, the inequality indices can also be decomposed by the contribution of each explanatory factor [18]. We specify the following regression model

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i \quad (2)$$

where y_i is health status, x_k are a set of demographic factors, socioeconomic variables, and smoking and obesity indicators, and ε_i is an error term. The concentration index can be written as

$$CI = \sum_k \left(\frac{\beta_k \bar{x}_k}{\mu}\right) CI_k + \frac{GC_\varepsilon}{\mu} \quad (3)$$

where μ is the mean of *y*, $\overline{x_k}$ is the mean of x_k, CI_k is the concentration index of x_k (defined analogously to CI) and GC_ε is the generalized concentration index for the error term. This shows that the concentration index CI is equal to a weighted sum of the concentration indices of the *k* regressors, where the weight for regressor *k* is the elasticity of *y* with respect to x_k.

Using this approach we calculate the contribution of smoking and obesity to income-related health inequality. With this analysis it is possible to further disentangle the contribution of these factors into two separate parts and to quantify the importance of each part: first, the impact of smoking and obesity on health is measured by the health elasticity $(\frac{\beta_k \bar{x}_k}{\mu})$; second, the unequal distribution across income groups is measured by the (income) concentration index CI_k; the overall contribution to income related health inequality is the product of the two parts.

Eq.(2) does not account for the endogeneity of smoking and obesity. While smoking and obesity may affect ill health, the reverse may also be true and/or there may be unmeasured characteristics that affect both smoking and obesity and health, such as the rate at which people discount the future. In these cases the coefficient β on the smoking and obesity variables in Eq.(2) will be biased and the contribution of these factors to income-related health inequality will be measured incorrectly. We use an IV regression model to measure the causal effect of smoking and obesity on health. Coefficient β for smoking and obesity in Eq.(2) and Eq.(3) is replaced by β^{IV}, which is not contaminated by endogeneity. In the IV model CI_k in Eq.(3) is the concentration index of predicted smoking and obesity estimated from the first stage of the IV model.

Standard errors of the estimates of the contributions in the decomposition analysis by factor are calculated using bootstrapping techniques computed using similar methods to Van Doorslaer and Koolman (2004) [19] and derived from 200 bootstrap replications.

Income

Income is included as a continuous variable based on the natural logarithm of annual equivalised household income estimated using a similar approach to Morris et al. [27]. Missing values were imputed using an interval regression model against a wide range of explanatory variables referring both to the respondent and the head of household. Observations with missing income values are not included in the computation of the CIs.

Regional indicators

The regional indicators used in our analysis are Government Office Regions (GORs), of which there were 9 in England during the period of our analysis.

Smoking and obesity

Smoking is measured as a binary variable taking the value unity if the individual is a current smoker and zero otherwise. Those who are not current smokers may have never smoked or may be ex-smokers. The obesity measure is based on BMI, measured as weight in kilogrammes divided by height in metres squared (kg/m^2). BMI is computed from the height and weight measures obtained during the interviewer visit in the HSE; it is not based on self reported height and weight, and the likelihood of systematic measurement error is reduced. Obesity is measured as a binary variable taking the value unity if the individual has BMI >30 kg/m^2 and zero otherwise.

Other covariates

In the decomposition of the income-related health inequality by factor we also include: social class of head of household (measured in seven categories); highest education qualification (seven categories); ethnicity (eight categories); housing tenure (five categories); marital status (five categories).

Accounting for endogeneity

To account for the endogeneity of smoking and obesity in the decomposition analysis we use IV regression to obtain the coefficients on the two lifestyle factors used in the health elasticities plus the other elements of the decomposition. We use a two stage approach. At the first stage we use separate OLS models to regress the smoking and obesity measures against a set of instrumental variables plus the covariates. We use the same set of three instruments in both models. We then compute the linear prediction of the smoking variable and the obesity variable for every observation using these models and at the second stage we use an OLS model to regress health status against the predicted smoking and obesity measures plus the covariates. The standard errors in the IV models are computed asymptotically. We instrument smoking using two variables delineating whether or not the respondent's mother smoked and whether or not their father smoked when the respondent was a child. We expect individual smoking to be positively correlated with parent smoking and we expect that, conditional on the large set of covariates also included in the model, parent smoking when individual was a child will only affect individual current health via its effect on individual smoking. We instrument obesity using the mean BMI among individuals in the HSE Primary Sampling Unit (PSU) in which the individual lives. This variable was constructed by collapsing individual level values of BMI in the HSE sample by PSU to produce a dataset of mean values at the PSU level. The mean number of sample observations per PSU (SD) is 21.9 (6.67) and we dropped individuals who lived in PSUs with less than 9 other individuals in the sample (2.16% of the sample). Conditional on the covariates, we expect that PSU mean BMI is positively correlated with individual obesity because the area measure summarises local obesity-affecting behaviors and reflects local social norms. In the IV models we also include three further PSU level variables: the % of the sample in the PSU with equivalised household income in the bottom quintile of the national distribution; the % with no education qualifications; and, the % with a limiting longstanding illness. PSU mean BMI is unlikely to be correlated with the error terms in the health equations given the large set of covariates also included in the model.

We test for exogeneity using auxiliary regressions [28]. We regress predicted health against the smoking and obesity measures and the covariates, also including the residual terms from the first stage smoking and obesity equations described above. The null hypothesis is that the coefficients on the residual terms equal zero. If they are significantly different from zero then, under the maintained assumption that the instruments are valid, the smoking and obesity measures are endogenous and the IV estimates are preferred. If we fail to reject the null hypothesis then, assuming the instruments are valid, it is not possible to identify any endogeneity problems with respect to the smoking and obesity variables and the non-IV estimates are preferred because they have lower standard errors [29]. We tested whether smoking and obesity were endogenous separately. In the event that one variable was endogenous but the other was not we reran the IV models only adjusting for the endogenous variable.

Sampling issues

We applied survey weights reported in the HSE to each observation and we controlled for PSU-level clustering in every regression.

Results

The results of the interval regression model of SAH against age, gender and the health indicators for the pooled sample are in Table 1. Most variables have the expected sign. Controlling for all the other variables, females report better health than males and the effect of age is non-linear. Higher GHQ-12 score, more days cut down on usual activities, limiting longstanding illness, and comorbidities are all negatively associated with health status. Most longstanding illnesses have a negative effect on reported health, with the exception of eye and ear problems. Neoplasm and benign growths, and infectious diseases have the largest negative impact on health.

Concentration indices for income-related inequality in health by year and by GOR are in Table 2. All the measures show that health is concentrated among the rich and all the indices are significantly different from zero. Inequality is lowest in London, the East and the South of

England and highest in the North and West. In most areas the level of inequality is higher in 2006 compared with 1998. The areas with highest CIs (North East, North West and West Midlands) experienced the biggest increases in inequality over time, together with the South East. The last three columns of Table 2 show the results of the decomposition of inequality into within- and between-area components by GOR by year. Within-area inequalities are positive (CI_w = 0.0027 for the pooled data), which suggests that within each GOR poorer individuals have worse health. The between-area inequality component shows that richer areas tend to have better health than poor areas. The between-area component is larger than the within-area component (CI_b = 0.0063 for the pooled data). Similar results were obtained using the second approach described above to estimate the between-area component. Most inequality is associated with the overlapping of the level of income across GORs (R = 0.0152 for pooled data), which has both within- and between-area inequality aspects. The positive value of R can be interpreted as evidence that richer individuals have better level of health irrespective of where they live. The ordering of GORs by increasing value of the Health Achievement Index produces a similar ordering to that achieved by decreasing value of the CI, and it indicates that areas with less inequality tended to have higher mean health (this is also shown directly in Table 2).

Table 3 shows raw prevalence data for current smoking and obesity (BMI>30 kg/m^2) by GOR and by year. Across most areas the prevalence of smoking has declined over time and the prevalence of obesity has increased. There is geographical variation in the distribution of both problems.

Table 4 shows the results of the non-IV OLS regression model on pooled data (all years and areas combined) used to conduct the decomposition analysis by factor on the pooled data. Socioeconomic status is positively correlated with health status while smoking and obesity are negatively correlated with health status. Table 5 shows analogous results for smoking and obesity only using the IV model. The coefficients on the instruments in the first stage smoking and obesity equations are highly significant and, as expected, parental smoking is positively correlated with individual smoking and PSU mean BMI is positively correlated with individual BMI. The coefficients on the smoking and obesity measures in the second stage health equation are negative and significant. The exogeneity tests reject the exogeneity of smoking and obesity and hence the IV results are preferred to the non-IV results.

The contribution of smoking and obesity to income-related inequality in health estimated using the non-IV and IV pooled models is shown in Table 6. In all cases the elasticities of health with respect to smoking and obesity are negative and smoking and obesity are both concentrated among the poor. Overall, in the non-IV models, smoking explains 1.4% of the socioeconomic inequality in health and obesity explains 1.3%. In the IV models the percentage contributions are higher for smoking than in the non-IV models (9.4%) and lower for obesity (0.8%). All the contributions are significantly different from zero. The contribution of smoking and obesity to pro-rich inequality in health is in part due to the negative effect of these variables on health and in part due to the concentration of these variables among the poor. In terms of the other covariates, income makes the largest contribution to income-related health inequality (35%), followed by age and gender (31%). The next largest contributors are smoking and obesity (10%), followed by housing tenure (8%) and PSU level area characteristics (3%).

The decomposition results by area are in Table 7. As with the national model, the general trend across areas is that for smoking the contributions using the IV models are greater than those from the non-IV models, whereas the opposite is true for obesity. Table 8 selects the most appropriate contributions derived from the coefficients in the non-IV and IV models, depending on the results of the exogeneity tests. There is considerable variation between areas in the contribution of smoking to income-related health inequality, while there is much less variation in the contribution of obesity. Smoking has the largest impact in the North East, accounting for 26% of income-related inequality. Obesity makes the largest contribution in the South West (1.6%).

Discussion

In this paper we have measured income-related health inequality using a representative dataset for England from 1998 to 2006, and across the nine Government Office Regions. We found that health is significantly concentrated among the rich in every year of the analysis and in every area of England. In most areas inequality is higher in 2006 compared with 1998, and especially so in the areas with overall larger level of inequality (the North East, the North West and West Midlands). London, the East of England and the South have the lowest level of income-related inequality in health. Health was found to be concentrated among rich individuals in each area, and also in rich areas as compared with poorer ones. The degree of overlapping in the income of rich individuals in poor areas and poor individuals in rich areas make difficult to disentangle into between- and within-inequality across GORs.

When exploring the contributions of smoking and obesity to income-related health inequality, we found that both factors deteriorate health significantly, and both were mainly concentrated among the poor. In the pooled model we found that smoking and obesity were endogenous, and that accounting for endogeneity changed the contribution of smoking and obesity to income-related health inequality. In the case of smoking, controlling for endogeneity increased the contribution of smoking. One explanation for this might be reverse causality. Poor health status reduces smoking and we observe this positive correlation between health status and smoking in our non-IV estimates. Once this effect is removed in the IV models the negative effect of smoking on health is recovered more clearly and the contribution of smoking to health inequality increases. We found the opposite effect for obesity in that the contribution of obesity to health inequality was higher in the non-IV models compared with the IV models. This suggests that there are unobserved variables that are positively correlated with obesity and negatively correlated with health (e.g. the discount rate). By controlling for these variables in the IV models the impact of obesity on health becomes less negative and its contribution to health inequality diminishes. Our findings have resonance with earlier studies. In particular, Balia and Jones (2008) found that before correcting for endogeneity, smoking and obesity accounted for 1.4% and 0.3% of total inequality in mortality. After correcting for endogeneity the figures were 9.6% and 0.06%. Our figures for the impact of these variables on income-related health inequality are 1.4% and 1.3% before correcting for endogeneity and 9.4% and 0.8% afterwards.

The results of this paper have important implications for policies aimed to reduce avoidable inequalities in health through the reduction of smoking and obesity and the variation in these factors across income groups. Our paper shows that both smoking and obesity are significant components of income related inequality in health even after controlling for a wide set of other socioeconomic and demographic indicators and endogeneity. We also show that their relative contributions vary across areas. While both smoking and obesity make a statistically significant contribution to income-related health inequality smoking makes a larger contribution, suggesting that policies aimed at reducing health inequality should focus more on this health issue.

In summary, we show that in England health continues to be concentrated among the rich and that in some areas that concentration has increased over time. Smoking and obesity explain part of the observed socioeconomic inequality in health, with smoking playing the larger role, and their contribution varying by area. We also show that smoking and obesity are endogenous and that it is important account for endogeneity when investigating the impact of these lifestyle factors on income-related health inequality. Considering both the prevalence of these factors and their role in explaining socioeconomic health inequality is necessary in order to inform policies aimed at reducing avoidable inequalities in health.

References

[1] Department of Health. Health Inequalities: Progress and Next Steps. June 2008.

[2] National Institute for Health and Clinical Excellence (2008). *NICE Public Health Guidance 10: Smoking cessation services in primary care, pharmacies, local authorities and workplaces, particularly for manual working groups, pregnant women and hard to reach groups.* NICE, London.

[3] Jarvis M, Wardle J (1999) Social patterning of individual health behaviours: the case of cigarette smoking. In: Marmot M, Wilkinson R, editors. Social determinants of health. Oxford: Oxford University Press.

[4] Twigg L, Moon G, Walker S (2004). The smoking epidemic in England. London: Health Development Agency

[5] Department of health. Consultation on the future of tobacco control. May 2008.

[6] Goddard E. (2008). *General Household Survey 2006: Smoking and drinking among adults 2006*. Office for National Statistics, Newport.

[7] National Centre for Social Research and Department of Epidemiology and Public Health at the Royal Free and University College Medical School. *Health Survey for England 2005: health of older people*. London: Information Centre for Health and Social Care, 2007.

[8] National Heart, Lung and Blood Institute [NHLBI]. *Clinical guidelines on the identification, evaluation and treatment of overweight and obesity in adults, NIH publication number 98-4083*. New York: National Institutes of Health, 1998.

[9] House of Commons Health Committee. Obesity*: third report of session 2003-4, Volume I, report, together with formal minutes*. London: The Stationery Office, 2004.

[10] The Information Centre for Health and Social Care (2008) *The Health Survey for England 2006*

[11] Department of Health. Tackling Health Inequalities – A Programme for Action. June 2003.

[12] Borg, V., Kristensen, T., (2000). Social class and self-rated health: can the gradient be explained by differences in life style or work environment? Social Science and Medicine 51, 1019–1030.

[13] Lantz, P.M., Lynch, J.W., House, J.S., Lepkowsky, J.M., Mero, R.P., Musick, M.A.,Williams, D.R. (2001). Socioeconomic disparities in healht change in a longitudinal study of US adults: the role of health-risk behaviors. Social Science and Medicine 53, 29–40.

[14] Contoyannis, P., Jones, A.M. (2004). Socio-economic status, health and lifestyle. Journal of Health Economics 23 (5), 965–995.

[15] Balia S. and Jones A.J. (2008). Mortality, lifestyle and socio-economic status. Journal of Health Economics, 27 (1); 1-26.

[16] Wagstaff A, Paci P, van Doorslaer E.(1991). On the measurement of inequalities in health. Soc Sci Med, 33; 545-557.

[17] Kakwani, Wagstaff and van Doorslaer (1997). Socioeconomic inequalities in health: Measurement, computation and statistical inference. Journal of Econometrics 77(1): 87-104.

[18] Wagstaff, van Doorslaer, Watanabe N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal Econon; 112(1):207-223.

[19] Van Doorslaer, E. and Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries, *Health Economics* 13(7): 609-628.

[20] van Doorslaer E, Jones AM. (2003). Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics, 22(1): 61-78.
[21] Lambert P, Aronson JR. (1993). Inequality decomposition analysis and the Gini coefficient revisited. Economics Journal. 103(420): 1221-1227.
[22] Clarke PM, Gerdtham U, Connelly LB. (2003). A note on the decomposition of the health concentration index. Health Economics. 12: 511-516.
[23] Wagtaff and van Doorslaer. (2004). Overall versus socio-economic health inequality; a measurement framework and two empirical illustration. Health Economics. 13: 297-301.
[24] Wagtaff (2005). Inequality decomposition and geographic targeting with applications to China and Vietnam. Health Economics 14: 649-653.
[25] Jimenez-Rubio, Smith & van Doorslaer (2008). Equity in Elath care in a decentralised context: evidence from Canada. Health Economics. 17: 377-392.
[26] Wagstaff (2002). Inequality aversion, health inequality and health achievement. Journal of Health Economics 21: 627-641.
[27] Morris, S., Sutton, M., Gravelle, H. (2005). Inequity and inequality in the use of health care in England: an empirical investigation. Social Science & Medicine; 60, 1251-66.
[28] Davidson R, MacKinnon G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.
[29] Baum CF, Schaffer ME, Stillman S. (2003). *Instrumental variables and GMM: estimation and testing*. Boston College Department of Economics, Working paper 545. Boston: Boston College Department of Economics.

Table 1. Results of interval regression model of self-assessed general health (pooled data)

Variable	Coeff	t	Variable	Coeff	t	Variable	Coeff	t
Age and gender			***Longstanding illness***			***Number of longstanding illnesses***		
Female	0.410	4.34	Infectious disease	-0.070	-2.51	0 or 1	Base category	
Age	-0.934	-4.45	Neoplasms & benign growths	-0.089	-10.70	2	-0.030	-8.13
Age-squared	0.541	3.77	Endocrine & metabolic	-0.047	-12.63	3	-0.086	-13.17
Age-cubed	-0.001	-0.03	Blood & related organs	-0.037	-3.44	4 or more	-0.137	-12.76
Female*age	0.014	0.11	Mental disorders	-0.049	-8.59	***Limiting longstanding illness***		
Female*age-squared	0.038	0.14	Nervous system	-0.047	-9.42	Yes	-0.130	-53.40
Female*age-cubed	-0.039	-0.21	Eye complaints	0.016	2.46	***Missing values***		
GHQ-12 score			Ear complaints	0.025	4.58	GHQ-12	-0.061	-18.56
0	Base category		Heart & circulatory	-0.057	-19.66	Acute ill	-0.015	-0.14
1	-0.020	-11.36	Respiratory system	-0.041	-14.35	Number LIs	-0.064	-1.57
2	-0.035	-13.43	Digestive system	-0.040	-9.62			
3	-0.048	-13.41	Genito-urinary system	-0.031	-4.90			
4	-0.057	-12.94	Skin complaints	0.008	1.21			
5	-0.075	-13.71	Musculo-skeletal	-0.013	-5.02			
6	-0.086	-13.31	Other complains	-0.062	-3.14			
7	-0.098	-13.17	***Acute ill health (days cut down)***					
8	-0.105	-12.17	0 days	Base category				
9	-0.121	-13.58	1-3 days	-0.009	-3.32			
10	-0.156	-14.42	4-6 days	-0.047	-9.03			
11	-0.154	-13.61	7-13 days	-0.068	-12.68			
12	-0.211	-16.85	14 days	-0.127	-27.27			
Observations				100,317				
Adjusted R-squared				0.332				

Table 2. Income-related inequality in health: CIs by area and year (ordered by pooled CI)

	NE	NW	W Mid	Yorks	E Mid	Lon	East	SE	SW	Nat.	Mean Health	CI_w	CI_b	*R*
1998	0.0255	0.0208	0.0232	0.0278	0.0220	0.0209	0.0190	0.0152	0.0170	0.0206	0.8483	0.0023	0.0058	0.0125
1999	0.0344	0.0279	0.0271	0.0213	0.0163	0.0191	0.0181	0.0153	0.0138	0.0205	0.8501	0.0024	0.0062	0.0119
2000	0.0369	0.0320	0.0230	0.0234	0.0317	0.0255	0.0263	0.0211	0.0230	0.0271	0.8376	0.0030	0.0059	0.0182
2001	0.0316	0.0287	0.0281	0.0255	0.0297	0.0305	0.0206	0.0264	0.0203	0.0269	0.8490	0.0031	0.0056	0.0182
2002	0.0287	0.0305	0.0466	0.0340	0.0275	0.0200	0.0234	0.0185	0.0227	0.0271	0.8463	0.0031	0.0083	0.0158
2003	0.0354	0.0257	0.0255	0.0243	0.0214	0.0215	0.0214	0.0174	0.0192	0.0226	0.8549	0.0026	0.0039	0.0161
2004	0.0490	0.0242	0.0325	0.0330	0.0273	0.0259	0.0213	0.0199	0.0142	0.0245	0.8535	0.0030	0.0032	0.0183
2005	0.0376	0.0366	0.0212	0.0315	0.0190	0.0264	0.0193	0.0215	0.0211	0.0251	0.8492	0.0030	0.0051	0.0170
2006	0.0414	0.0238	0.0275	0.0276	0.0243	0.0175	0.0218	0.0199	0.0112	0.0230	0.8547	0.0027	0.0074	0.0129
Pooled	*0.0340*	*0.0274*	*0.0271*	*0.0261*	*0.0234*	*0.0231*	*0.0207*	*0.0190*	*0.0173*	*0.0242*	*0.8487*	*0.0027*	*0.0063*	*0.0152*
Mean Health	0.801	0.835	0.836	0.835	0.843	0.851	0.867	0.874	0.864	0.8487				
Mean Income	20,216.4	20,598.7	22,518.6	22,643.2	23,133.2	30,507.7	27,887.8	31,047.9	24,207.5	25,655.6				
HAI	0.7736	0.8125	0.8135	0.8129	0.8237	0.8317	0.8486	0.8569	0.8492					
Observations	5,068	10,861	8,611	8,190	7,367	8,861	9,639	12,861	8,611	80,069				

NE: North East, NW: North West, Yorks: Yorkshire, E Mid: East Midlands, Lon: London, W Mid: West Midlands, East: East of England, SE: South East, SW: South West, Nat: National, CI: Concentration index, HAI: Health Achievement Index, CI_w: Within areas concentration index, CI_b: Between areas concentration index, *R*: Residual

Table 3. Smoking obesity prevalence by area and year (ordered by pooled results)

	Smoking (current smoker, %)										Obesity (BMI >30kg/m^2, %)									
	NE	NW	Yorks	E Mid	Lon	W Mid	East	SE	SW	Nat	W Mid	NE	E Mid	Yorks	NW	East	SW	SE	Lon	Nat
1998	31.2	29.5	29.6	28.7	28.1	26.2	25.4	26.2	24.7	27.5	19.4	18.4	18.9	18.4	17.6	17.8	16.6	17.1	15.8	17.6
1999	26.8	30.7	28.2	23.6	28.8	26.6	23.7	25.5	23.2	26.4	19.7	19.5	18.2	16.7	20.5	17.0	18.4	15	16.6	17.8
2000	27.7	25.4	26.7	24.4	23.0	19.9	23.2	20.3	21.9	23.3	19.7	20.8	21.2	19.1	19.8	18.0	16.7	18.9	17.5	18.9
2001	29.7	25.4	27.4	26.8	27.0	23.7	25.2	22.4	23.8	25.3	22.8	23.4	21.6	21.5	18.2	20.0	19.2	18.1	16.2	19.8
2002	31.0	27.8	27.4	26.5	26.0	26.2	22.4	24.9	25.2	26.1	20.9	22.2	24.7	23.4	21.2	18.7	15.9	18.7	17.7	20.2
2003	27.4	26.4	27.0	23.4	24.8	25.0	25.8	25.0	24.5	25.4	22.7	22.1	21.9	21.5	21.7	21.9	20.4	17.7	15.9	20.2
2004	26.4	24.7	25.4	25.8	21.1	24.8	23.4	19.9	19.9	23.0	22.0	21.6	23.0	21.4	17.1	21.5	18.7	19.3	14.4	19.4
2005	33.1	28.0	26.2	27.1	24.6	24.1	23.1	23.3	22.6	25.2	23.9	22.3	22.0	19.0	20.9	21.8	20.0	18.9	12.0	19.6
2006	27.9	24.0	23.3	26.3	22.1	23.0	19.7	20.3	20.0	22.5	26.1	23.6	23.1	22.1	19.9	20.0	22.8	19.4	14.8	20.7
Pooled	29.3	26.8	26.8	26.0	25.0	24.3	23.5	23.2	23.0	25.0	22.1	21.7	21.7	20.6	19.7	19.6	18.9	18.2	15.6	19.5

NE: North East, NW: North West, Yorks: Yorkshire, E Mid: East Midlands, Lon: London, W Mid: West Midlands, East: East of England, SE: South East, SW: South West, Nat: National

Table 4. Regression results for non-IV predicted health model (pooled data)

Variable	Coeff	t
Age and gender		
Female	0.449	5.85
Age	-0.016	-1.21
Age-squared	-1.543	-9.16
Age-cubed	0.967	8.44
Female*age	0.030	0.3
Female*age-squared	0.046	0.21
Female*age-cubed	-0.046	-0.31
Ln(income)	0.018	21.11
Social class of head of household		
Professional	Base category	
Managerial/technical	-0.002	-1.07
Skilled non-manual	0.000	0.12
Skilled manual	0.000	0.07
Semi-skilled manual	-0.002	-0.87
Unskilled manual	-0.002	-0.63
Other	-0.007	-1.68
Education		
Degree	Base category	
Higher education less than	0.0002	0.15
A level or equivalent	-0.002	-1.43
GCSE or equivalent	0.002	1.41
CSE or equivalent	-0.001	-0.4
Other qualification	0.002	0.64
No qualification	-0.010	-5.47
Housing tenure		
Own	Base category	
Mortgage	-0.003	-2.07
Part mortgage	-0.018	-2.41
Rent	-0.030	-16.14
Free rent	-0.014	-2.48

Variable	Coeff	t
Ethnic group		
White	Base category	
Black Caribbean	-0.002	-0.33
Black African	0.024	4.83
Indian	-0.006	-1.36
Pakistani	-0.019	-3.33
Bangladeshi	-0.006	-0.75
Chinese	0.032	5.06
Other	-0.008	-2.1
Marital status		
Married	Base category	
Single	-0.013	-8.84
Separated	-0.020	-5.79
Divorced	-0.019	-9.08
Widowed	-0.008	-3.35
Smoking and obesity		
Smoker	-0.009	-8.23
BMI >30	-0.026	-20.63
Year		
2006	Base category	
2005	-0.004	-1.86
2004	-0.005	-2.03
2003	-0.004	-2.28
2002	-0.006	-2.69
2001	-0.001	-0.42
2000	0.001	0.29
1999	-0.004	-1.47
1998	0.001	0.56

Variable	Coeff	t
GOR		
London	Base category	
North East	-0.026	-8.81
North West	-0.019	-8.47
Yorkshire	-0.015	-6.11
East Midlands	-0.015	-0.015
West Midlands	-0.009	-0.009
East of England	-0.001	-0.66
South East	-0.004	-0.004
South West	0.004	1.78
Missing		
Social class	0.012	0.75
Education	-0.018	-1.37
Ethnic	0.029	2.18
Marital	-0.027	-0.7
Tenure	-0.025	-0.69
Smoking	-0.018	-3.16
BMI	-0.039	-21.7
Income	0.054	3.58
GOR	-0.016	-1.33

Observations	79,688
Adjusted R-squared	0.208

Table 5. Regression results for IV predicted health model (pooled data, impact of smoking and obesity only)

	Coef.	t-value
Impact of smoking and obesity on health status (IV model, second stage)		
Smoking	-0.067	-4.92
Obesity	-0.017	-2.96
Observations	78,156	
Exogeneity test: χ^2 [*p* value]		
Smoking	20.73 [<0.0001]	
Obesity	6.31 [0.012]	
Impact of the instruments on smoking (IV model, first stage)		
Mother smoker	0.047	13.33
Father smoker	0.043	12.26
PSU mean BMI	-0.008	-6.10
Observations	78,156	
F test $\text{Coef.}_{\text{instruments}}$=0 [*p* value]	142.34 [<0.0001]	
Impact of the instruments on obesity (IV model, first stage)		
Mother smoker	0.017	5.41
Father smoker	0.005	1.63
PSU mean BMI	0.052	52.94
Observations	78,156	
F test $\text{Coef.}_{\text{instruments}}$=0 [*p* value]	951.1 [<0.0001]	

Controls are also included but not reported for age and gender, income, social class of head of household, education, housing tenure, ethnic group, marital status, year, GOR, missing values, % of the sample in the PSU in the lowest national income quintile, % of the sample in the PSU with no qualifications and % of the sample in the PSU with a limiting longstanding illness.

Table 6. Contribution of smoking and obesity to income-related inequality in health (pooled data)

	Coeff	Elasticity	CI_k	Contribution (Elasticity * CI_k)	Contribution /SE	Percentage
Non IV Models						
Smoker	-0.009	-0.003	-0.122	**0.00035**	8.20	**1.4**
BMI >30	-0.026	-0.006	-0.054	**0.00033**	10.32	**1.3**
Smoking + Obesity				**0.00068**	13.02	**2.8**
IV Models						
Smoker	-0.067	-0.020	-0.113	**0.0023**	4.40	**9.4**
BMI >30	-0.017	-0.004	-0.050	**0.0002**	2.38	**0.8**
Smoking + Obesity				**0.0025**	4.65	**10.2**

Coeff: Coefficient, SE: Bootstrapped standard error of the contribution based on 200 replications.
All the models also include controls for age and gender, income, social class of head of household, education, housing tenure, ethnic group, marital status, year, GOR and missing values. In addition, the IV models also control for % of the sample in the PSU in the lowest national income quintile, % of the sample in the PSU with no qualifications and % of the sample in the PSU with a limiting longstanding illness. Contributions in bold are statistically significant at the 10% level.

Table 7. Percentage contribution of smoking and obesity to income-related health inequality by GOR

	Smoking		Obesity		Smoking + Obesity	
	Non-IV	IV	Non-IV	IV	Non-IV	IV
North East	**2.9**	**38.0**	**1.2**	2.4	**4.1**	**40.4**
North West	**1.1**	1.4	**1.3**	0.4	**2.4**	1.7
West Midlands	**1.6**	8.2	**1.5**	1.1	**3.1**	9.3
Yorkshire	**2.5**	14.1	**0.8**	-0.1	**3.3**	14.0
East Midlands	**1.2**	**16.2**	**0.8**	-0.1	**2.1**	16.1
London	**1.2**	4.6	**1.5**	2.0	**2.7**	**6.6**
East of England	**0.9**	**10.4**	**1.1**	0.6	**1.9**	**11.0**
South East	0.4	7.9	**1.4**	1.4	**1.8**	**9.3**
South West	**1.9**	**15.7**	**1.6**	**2.8**	**3.6**	**18.5**
National	**1.4**	**9.4**	**1.3**	**0.8**	**2.8**	**10.2**

All the models also include controls for age and gender, income, social class of head of household, education, housing tenure, ethnic group, marital status, year and missing values. The national model also controls for GORs. In addition, the IV models also control for % of the sample in the PSU in the lowest national income quintile, % of the sample in the PSU with no qualifications and % of the sample in the PSU with a limiting longstanding illness. Contributions in bold are statistically significant at the 10% level.

Table 8. Percentage contribution of smoking and obesity to income-related health inequality by GOR (from preferred model)

	Smoking	Obesity	Smoking + Obesity
North East[c]	**26.3**	**1.1**	**27.4**
North West[a]	**1.1**	**1.3**	**2.4**
West Midlands[a]	**1.6**	**1.5**	**3.1**
Yorkshire[a]	**2.5**	**0.8**	**3.3**
East Midlands[b]	**16.2**	-0.1	16.1
London[a]	**1.2**	**1.5**	**2.7**
East of England[c]	**10.4**	**1.0**	**11.4**
South East[c]	**8.0**	**1.4**	**9.3**
South West[c]	**13.7**	**1.6**	**15.3**
National[b]	**9.4**	**0.8**	**10.2**

a = smoking and obesity are both exogenous; b = smoking and obesity are both endogenous; c = smoking is endogenous and obesity is exogenous

All the models also include controls for age and gender, income, social class of head of household, education, housing tenure, ethnic group, marital status, year and missing values. The national model also controls for GORs. In addition, the IV models also control for % of the sample in the PSU in the lowest national income quintile, % of the sample in the PSU with no qualifications and % of the sample in the PSU with a limiting longstanding illness. Contributions in bold are statistically significant at the 10% level.