Landman Economics

The impact of smoking history on employment prospects, earnings and productivity: an analysis using UK panel data

Written by Howard Reed of Landman Economics for ASH
Revised September 2020

Acknowledgements

The author would like to thank Dr J Robert Branston, University of Bath and Ciaran Osborne of ASH for comments on previous drafts of this report, and Floe Foxon for additional comments and insights regarding Chapter 6.

The data from Understanding Society (USoc) used in this report were made available through the UK Data Archive.

This research was funded by the Department for Health and Social Care.

Contents

1 Introduction

Smoking is associated with lower productivity and losses in economic output due to increases in mortality and morbidity for the smoking population (Reed 2010; DHSC 2017). Previous work on the negative economic impacts of smoking in the UK has focused primarily on higher working-age morbidity and greater employee absenteeism due to smoking, resulting in lower economic output and lower tax receipts for the Exchequer. This report looks at the relationship between smoking and employment status and smoking and earnings using data from a British longitudinal dataset, Understanding Society (USoc). The primary aim of the research is to estimate the causal impact of previous smoking over a number of years on the probability of being employed, and on earnings for those in employment, at a recent point in time. The research estimates the impact of smoking in waves 2 through 7 of the survey – as well as information on whether people were smokers before wave 2 of the survey – on the probability of employment and the earnings from employment of respondents in Wave 8 of the survey, in 2016-17. This paper also examines the links between disability and smoking in an attempt to model the longer-term causal relationships between smoking, disability and employment. Finally, this report attempts to quantify the total productivity costs of smoking to the UK economy based on the most recent available data on smoking prevalence, employment and earnings among adults in Britain. These figures represent the most comprehensive evaluation of the productivity impacts of smoking in the UK to date.

The report is structured as follows. Section 2 summarises previous evidence on the relationship between smoking and employment and smoking and earnings. Section 3 gives details of the USoc data and presents descriptive statistics from the USoc sample on employment rates for smokers and non-smokers according to characteristics such as gender, disability status and highest qualification. Section 4 explains how the relationship between smoking and employment, smoking and earnings and smoking and disability is modelled in this report. Section 5 presents regression results for the relationship between smoking and employment and earnings, as well as smoking, employment and disability. Section 6 presents estimates for the overall cost of smoking in terms of reduced productivity at a national level. Section 7 offers conclusions.

2 Previous research on the relationship between smoking, employment and earnings

2.1 Smoking and employment

The relationship between smoking and employment is likely to be the result of a complex set of factors. It is certainly the case that people living in socially and economically deprived areas are less likely to be in work and are also more likely to smoke. For example, Semple (2015) uses data from the Scottish Index of Multiple Deprivation – a measure that looks at 38 indicators across income, housing, education, employment and health – and shows that those living in the 20% most socially deprived areas are about four times more likely to smoke than those in the 20% of most affluent areas.

A meta-analysis of 29 longitudinal or cohort studies of the relationship between smoking and work absenteeism concluded that smokers were 33% more likely to be absent from work with smokers taking an average of 2.74 additional days of sick leave per year compared to non-smokers. The overall productivity loss to the UK from smoking-related absences from work was calculated at £1.4 billion in 2011 (Weng *et al*, 2013). Having chronic health problems arising from smoking may impact on your employment record and make it more difficult to find a job when circumstances change.

Smoking may also be a barrier to gaining and staying in employment and recent research from France and the USA supports this theory. Even after controlling for demographic factors and other risk characteristics (e.g. obesity, binge drinking), current smoking among Californians was significantly associated with being unemployed and job seeking; the point estimate from a logistic regression showed that smokers were 23 per cent more likely to be unemployed than non-smokers, controlling for other factors (Prochaska *et al*, 2013). A longitudinal study of French workers found that heavy smoking was associated with becoming unemployed (Jusot *et al*, 2008) while a report examining longitudinal data from 1998 to 2008 from the German Socio-Economic Panel (SOEP) suggests that the unemployed were more likely to smoke and that smokers had a higher probability of becoming unemployed (Schunck and Rogge, 2012). A large, comprehensive study of employment and smoking among a cohort in New York state over a 29-year period found that, after adjusting for deprivation and other potential confounding variables, those who were continuous or occasional smokers were four times more likely to be unemployed at age 43 than those who were never smokers or who had quit smoking (Brook, 2014). The study concluded that intervention programmes designed to deal with unemployment should consider focusing on smoking as a potential barrier to employment.

2.2 Smoking and earnings

Smoking also appears to have a negative impact on wages. A study by Levine, Gustafson and Velenchik (1997) using data from the US National Longitudinal Survey of Youth examined the effect of smoking on wages controlling for differences

in individual characteristics that may be correlated with both smoking and wages, including unobservable person-specific characteristics that are constant over time, and unobservable characteristics that are constant within a family. Estimates indicate that smoking reduced wages by between 4 and 8 per cent.

Böckerman *et al* (2014) use a sample of twin data for Finnish men born between 1945 and 1957 (to remove shared environmental and genetic factors) to examine the long term effects of smoking on labour market outcomes. The results show that smokers have lower long-term income and earnings after controlling for shared environmental and genetic factors. The results suggest that an extra “cigarette-pack-year” of smoking (i.e. smoking one pack of 20 cigarettes per day, every day for a year) reduces lifetime incomes by approximately 1 per cent. This negative association is robust to the use of various covariates, including education, health indicators and extraversion.

3 The Understanding Society (USoc) data

3.1 The scope and sampling frame for USoc

USoc is a large-scale longitudinal panel survey operating in all four countries of the UK. The survey began in 2009 and eight waves of data have so far been released. The sampling period for each wave is two years, with each household in the survey being interviewed annually. This means that in any one year, interview fieldwork is being conducted for two waves simultaneously (for example in 2016, interviews were carried out for the first half of the wave 8 sample and the second half of the wave 7 sample). Table 3.1 gives details of the number of individual interviews achieved in each wave and when the fieldwork took place. The number of individual interviews falls between one wave and the next wave for most waves, with the exception of Waves 2 and 6. In both cases this is because the sample was boosted with the addition of additional households not in the USoc survey at Wave 1. In Wave 2, households from the British Household Panel Survey (the predecessor survey of USoc, which ran for 18 waves between 1991 and 2018) were added to USoc, while in Wave 6 an immigrant and ethnic minority boost sample (IEMBS) was introduced. Taking into account the boost samples, in Wave 8 just over 37,600 full individual interviews were conducted in USoc.

Table 3.1. Understanding Society Waves 1-8: fieldwork dates and sample sizes

Wave	Date of fieldwork	Number of successful full individual interviews (excluding proxy interviews)
1	January 2009-December 2010	47,732
2	January 2010-December 2011	50,684
3	January 2011-December 2012	45,860
4	January 2012-December 2013	43,132
5	January 2013-December 2014	40,970
6	January 2014-December 2015	41,859
7	January 2015-December 2016	39,332
8	January 2016-December 2017	37,606

Source: Institute for Social and Economic Research (2018).

3.2 Attrition across Waves 1 to 8

Attrition from panel surveys occurs when individuals interviewed for one wave of the survey are not able to be interviewed in subsequent waves. This can occur because individuals move house and the USoc administrators lose touch with them, or because individuals move outside the UK, or because individuals die, for example. Table 3.2 shows the attrition of the individuals who were interviewed in Wave 1 of USoc who dropped out in subsequent waves. Between Waves 1 and 2 there was a high rate of attrition – almost 20 percent (one in five) of the sample dropped out. The rate of attrition slowed in subsequent waves, but by wave 7 more than half the wave 1 sample were no longer in the survey, and by Wave 8 only just under 44 per cent of the sample from Wave 1 were left. This means that the subsample of individuals who were interviewed in all of Waves 1 to 8 inclusive could be a lot less representative of the general UK population than the sample from Wave 1. Similarly, the composition of the 20,803 interviewed adult sample members who were interviewed in all of Waves 1 to 8 may be quite different from the full sample of 37,606 adult interviews in Wave 8 – including the IEMB boost sample, the BHPS legacy sample and also other sample members who completed some, but not all, of the 8 USoc Wave interviews. We discuss the issue of sample representativeness and weighting further in Section 4.4 below.

Table 3.2. Understanding Society individual attrition: Number and proportion of interviewees from Wave 1 by last Wave interviewed

Wave	Number dropping out	Remaining sample members	% of original sample
1 (start)	n/a	47,732	100.0%
2	8,892	38,840	81.4%
3	4,894	33,946	71.1%
4	3,080	30,866	64.7%
5	2,183	28,683	60.1%
6	2,982	25,701	53.8%
7	2,185	23,516	49.3%
8 (most recent)	2,713	20,803	43.6%

Source: author's calculations using USoc data

3.3 Questions on smoking in the USoc data

Although Understanding Society records certain variables in each wave (e.g. region of residence, employment status, gender) and some variables are initially recorded and then fixed (e.g. date of birth), there are some questions which are in some waves but not other waves, or where the precise questions asked differ between waves. Smoking is one of these categories of question. Table 3.3 explains the structure of the questions asked about question in each wave of the USoc data and the sampling frame used. The main findings from Table 3.3 are:

- Current smoking status is available in all of the waves except for Wave 1 and Waves 3 and 4.
- In Waves 3 and 4, the smoking questions were only asked of interviewees who had entered the adult component of the USoc survey for the first year after previously being in the youth component of USoc. This means that it is not possible to estimate a smoking prevalence rate for the population in Waves 3 and 4, but it is possible to do so in wave 2, 5 and subsequent waves.
- Frequency of smoking is available in waves 5 to 8.
- In Wave 2 a question was asked about whether people who were current non-smokers had ever smoked before Wave 2.

Table 3.3. Smoking questions in Understanding Society Waves 1-8

Wave	Smoking questions	Sample asked smoking questions
1	none	none
2	• Whether ever smoked cigarettes • Whether smoke cigarettes now	All individuals
3	• Frequency of smoking	Young people who have moved from the USoc child sample to the adult sample only
4	As wave 3	As wave 3
5	• Whether ever smoked cigarettes • Whether smoke cigarettes now • Frequency of smoking (for current smokers) • Previous frequency of smoking (for ex-smokers who have now given up)	Whole sample
6	• Whether smoke cigarettes now • Smoking frequency	Whole sample
7	• Whether smoke cigarettes now • Smoking frequency • Whether use e-cigarettes	Whole sample
8	As wave 7	As wave 7

3.4 Other variables used in this research

This research also uses a range of other variables from the Understanding Society data. In most cases these are from wave 8 of the sample although some lags are also used in various specifications, as detailed in Chapter 4 below.

Employment status

USoc contains data on employment status for all adults in the survey. Because we are primarily interested in the distinction between working and not working rather than (for example) the relationship between different types of employment or non-employment, this paper models employment status as a binary variable (working vs not working) and so does not distinguish between self-employed and employee workers, or between unemployed, inactive and retired non-workers.

Earnings

USoc contains data on monthly and hourly earnings. For the earnings regressions in this paper we use monthly earnings data as we are most interested in the

relationship between smoking and overall labour market earnings per month (or year) as this is the most closely related earnings measure to aggregate productivity.

Disability, health status and life satisfaction

USoc contains a binary variable for whether sample members have a long-standing illness or disability. This is used as the main measure of disability in this report. The data also contain information on self-reported health status and self-reported satisfaction with life in general and with the respondent's level of earnings. Dummy variables for poor self-reported health and low levels of life satisfaction are included in some of the regression specifications estimated in this report as additional control variables.

Highest educational qualification

USoc contains detailed data on the highest qualifications attained by individuals in Wave 1 of the sample (or for newer entrants, the first wave in which they appear in the sample), and then subsequently records any further qualifications gained in future waves of the sample. We have used this information to construct a variable for highest educational qualification in Wave 8 of the sample which has six categories:

- Degree (first or higher);
- Other higher education qualifications (e.g. nursing qualifications, diploma);
- A Levels or equivalent;
- GCSEs or equivalent;
- Other qualifications (including non-UK qualifications);
- No qualifications.

The first five of these categories are entered into the regressions as binary variables, with "no qualifications" being the base category.

Age group

USoc includes an age variable for everybody in the survey. The regressions include age dummies for under-25s and then five year age categories for 25-29, 30-34 and so on, all the way up to 65-69. The age range used in the regressions is 21 to 69 inclusive. People aged under 21 are not included because a high proportion of them are still in full-time education, while people aged over 69 are not included because most of them are retired.

Ethnicity

USoc contains a detailed ethnicity variable which is used to divide respondents into 9 ethnic groups using binary variables as follows:

- White British;
- White non-British;
- Mixed ethnicity;
- Indian;
- Pakistani;
- Bangladeshi;
- Other Asian (e.g. Chinese);
- Black (African, Caribbean, Black British);
- Other ethnic group.

Youngest child in household

Binary variables are included for whether there are any children aged 16 or under in the household, broken down according to the age of the youngest child in the household:

- Aged under 2;
- Aged 2 to 4;
- Aged 5 to 10;
- Aged 11 to 16.

The youngest child variable is interacted with gender of the (adult) interviewee as statistics show that the relationship between having children in the household and being in paid employment is very different for women than for men. Women with children, and especially young children, have lower rates of employment than women without children whereas men with children have slightly higher rates of employment than men without children, on average.

Other individual covariates

Other variables from the individual USoc data records which might affect employment status are included as covariates, specifically the following:

- Gender
- Currently pregnant
- Caring full-time for a disabled adult in the household

Other household covariates

Other variables from the household USoc data records which are correlated with employment levels are included as sets of binary variables, namely:

- Housing tenure (Local authority or social tenant, private sector tenant, own own home outright or with mortgage)
- Region (9 English regions, Scotland, Wales, Northern Ireland)

3.5 Descriptive statistics

This subsection presents some descriptive statistics from the USoc data on smoking and employment rates.

Smoking prevalence over time

Table 3.4 shows smoking prevalence and employment rates for adults aged 21 to 69 in the USoc data using the panel weights in the data for individuals with a full set of interviews over each of the 8 USoc waves. Only Waves 2, 5, 6 7 and 8 are included in the Table as the ‘current smoking’ question was only asked for all adult sample members in these waves.

Table 3.4 shows that smoking prevalence has declined over time in the USoc data from 20.2% of adults aged 21-69 in Wave 2 to 17.6% of adults in Wave 8. The average rate of decline in smoking prevalence across Waves 2 through 8 is just under 0.4 percentage points per year. At the same time, employment rates in the weighted USoc sample have risen slightly, from 66.2% in Wave 2 to 67.4% by Wave 7 (with a slight fall in Wave 8).

Table 3.4 Smoking prevalence and employment rates over time in Understanding Society

Wave	Smoking prevalence (% of sample, weighted)	Employment rate (% of sample, weighted)
2	20.2%	66.2%
5	18.7%	67.0%
6	18.3%	67.1%
7	17.8%	67.4%
8	17.6%	67.3%

Sample: adults aged 21-69, balanced panel with full data for waves 1 to 8 inclusive.

Employment rates by smoker status and other characteristics

Table 3.5 shows the employment rate for adults aged 21-69 inclusive in the USoc Wave 8 data, separately for current smokers and non-smokers, for the population as a whole and also broken down by a number of other variables (gender, disability status and highest qualification). The left-hand column shows the employment rate for non-smokers in the relevant group, the middle column shows the employment rate for smokers and the right-hand column is the difference between smokers' and non-smokers' employment.

Overall, employment for smokers is 11 percentage points lower than for non-smokers. The employment gap is larger for men than for women, and much larger for disabled adults than for non-disabled adults (there is a gap of over 16 percentage points for disabled smokers compared to non-smokers, whereas the equivalent gap for non-disabled adults is only just over 4 percentage points). By age group, the biggest gaps in employment rate between smokers and non-smokers are for 45-49 year olds (23 percentage points) and 55-59 year olds (18.5 percentage points). In contrast, the employment rate for 65-69 year olds is very similar for smokers and non-smokers at between 15 and 16 per cent for both groups.

By highest qualification, the biggest gap in employment rates between smokers and non-smokers is for A-levels (around 10 percentage points) and those with no qualifications (around 8 percentage points). For adults with 'other' qualifications, the gap is *negative*, i.e. smokers are more likely to be in employment than non-smokers. However, this is a relatively small group (less than 2 percent of the sample).

Table 3.5 Employment rates in Understanding Society Wave 8

Sample characteristic	Employment rate		
	Non-smokers (%)	Current smokers (%)	Difference (% pts)
Gender:			
Male	74.7	62.1	12.6
Female	64.3	54.4	9.9
Disability:			
Non-disabled	76.5	72.1	4.4
Disabled	53.3	36.8	16.5
Age group:			
21-24	72.9	65.9	7.0
25-29	80.9	65.9	15.0
30-34	80.5	65.7	14.8
35-39	79.4	69.3	10.1
40-44	82.6	69.9	12.7
45-49	84.0	61.0	23.0
50-54	82.3	66.5	15.8
55-59	71.8	53.3	18.5
60-64	50.3	37.1	13.2
65-69	15.8	15.2	0.6
Highest educational qualification:			
Degree	78.0	74.8	3.2
Other HE	71.0	69.4	1.6
A Level	78.0	68.3	9.7
GCSE	70.7	63.1	7.6
Other	48.7	52.1	-3.4
None	46.4	38.0	8.4
Overall sample	69.2	58.2	11.0

Average earnings by smoker status and other characteristics

Table 3.6 shows median monthly earnings for employed adults aged 21-69 inclusive in the USoc Wave data, using the same set of breakdowns as for Table 3.5 above. The left-hand column shows median earnings for non-smokers in the relevant group, the middle column shows earnings for smokers and the right-hand column is the earnings premium for non-smokers compared to smokers in each group (in percent). Overall, median monthly earnings for non-smokers are 25 percent higher than for smokers. The earnings gap is larger for men than for women (31 percent compared to 25 percent); it is very similar for disabled and non-disabled people (26 percent compared to 27 percent). By age group, the biggest earning gaps are for 40-44 year olds (44 per cent) and 50-54 year olds (35 per cent) whereas the gap is smallest for workers aged 60 or over. By highest qualification, there is a 41 percent gap between median earnings for non-smokers and smokers for degree holders, and smaller gaps

for other qualifications holders, but no difference in median earnings between smokers and non-smokers for workers with no qualifications.

Table 3.6 Median monthly earnings for people in work in Understanding Society Wave 8

Sample characteristic	Median monthly earnings		
	Non-smokers (%)	Current smokers (%)	% gap (non-smokers compared to smokers)
Gender:			
Male	2,300	1,750	31
Female	1,500	1,200	25
Disability:			
Non-disabled	1,929	1,517	27
Disabled	1,750	1,387	26
Age group:			
21-24	1,400	1,300	8
25-29	1,775	1,370	30
30-34	1,916	1,586	21
35-39	2,167	1,654	31
40-44	2,153	1,500	44
45-49	2,150	1,731	24
50-54	2,100	1,560	35
55-59	1,900	1,463	30
60-64	1,560	1,500	4
65-69	894	850	5
Highest educational qualification:			
Degree	3,100	2,200	41
Other HE	2,528	2,300	10
A Level	1,817	1,600	14
GCSE	1,587	1,332	19
Other	1,428	1,387	3
None	1,300	1,300	0
Overall sample	1,877	1,500	25

4 Methodology

This section explains the methodology used in the employment and earnings regressions as well as the additional regressions used to model disability.

4.1 Main model specification

Employment regression

The main employment model specification uses a logistic regression for individuals in Wave 8 of the USoc Survey who were also in Waves 1 through 7 of the survey (i.e. a complete set of panel data for all existing waves). The specification uses a binary dependent variable for employment at Wave 8 of USoc. This is regressed against two sets of explanatory variables:

1) **Smoking status in previous waves of USoc.** Smoking status in previous waves, rather than Wave 8, is used to help to control for possible mis-specification of the regression, arising because it is possible that the causal link between smoking and employment runs both ways. In other words, people might be more likely to smoke because they are not working, as well as people being less likely to be in work because they smoke. Using lagged smoking variables helps isolate the causal impact of smoking on employment, on the grounds that previous smoking behaviour may affect employment status. We discuss the exact specification of the previous smoking variable(s) in Section 4.2 below.
2) **Control variables.** A range of other control variables which might affect employment is also included in the regression, for example age group, gender, age of youngest child in the household, highest qualification, ethnicity, disability, housing tenure and region. The complete set of control variables is listed in Section 4.3 below.

The employment regression sample consists of all adults in USoc Wave 8 aged 21 to 69 inclusive. Although the state pension age is currently in the process of rising from 65 to 66, it is useful to include men and women aged between 66 and 69 in the sample as the employment rate for this group is significantly greater than zero[1] and it is possible that smoking may have a particular impact on retirement ages which should be taken into account in this research.

Earnings regression

The earnings regression is an Ordinary Least Squares (OLS) specification using the subsample of individuals who were in employment (either employees or self-employed) during Wave 8 and for whom there also exists a complete set of interview data for Waves 1 to 7. The dependent variable is the log of earnings in the most recent month before interview. Using a log earnings measure means that the

[1] For example, in USoc Wave 8 the employment rate for adults aged 66 to 69 (inclusive) is around 15 per cent.

coefficients from the earnings regression can be interpreted (approximately) as percentage effects on earnings of a unit change in each explanatory variable. However it also means that the sample has to be restricted to individuals with positive monthly earnings only; we exclude the small proportion of the employed Wave 8 sample (around 0.2 per cent) who are voluntary workers with zero earnings, or self-employed people making losses.

The set of previous smoking variables and other control variables used in the earnings regression is the same as the variables used in the employment regression covered in points (1) and (2) above, with the addition of a labour market history binary variable to capture the impacts of being unemployed or inactive in the waves of Understanding Society prior to Wave 8. This variable is set to 1 for individuals who are unemployed and inactive in two or more of Waves 5, 6 and 7, and zero for individuals who are unemployed or inactive in just one of those waves, or employed in all three waves.

Disability regression

As well as modelling the relationship between employment and previous smoking and earnings and previous smoking, this report also estimates models of the relationship between disability status and smoking for individuals. In particular, we are interested in the correlation between having a limiting long-standing illness or disability, and a long-run history of smoking. The sample used for the disability regressions is the same as for the employment regressions.

4.2 Smoking variables used in the regressions

The USoc data contains smoking prevalence data across the whole sample for Wave 2 and Waves 5 to 8 of the survey. Estimating regressions using employment or earnings in Wave 8 as the dependent variable, four different lags of smoking (smoking status in Wave 7, Wave 6, Wave 6 and Wave 2) are available to use as explanatory variables. After some experimentation with different lags, the results presented in this paper use two lagged smoking variables in the main regression specification: Smoking in Wave 7 (a one year lagged variable) and smoking in Wave 2 (a six-year lagged variable). The Wave 7 variable measures the ‘short-run’ correlation between smoking and employment or earnings, whereas the Wave 2 variable measures whether there are any ‘medium-term’ effects of smoking on employment or earnings which may manifest themselves over and above the short-term effects.

In addition to this, a variable is also included for whether people have *ever* smoked (including smokers before Wave 2). This is based on a question asked at Wave 2 about whether individuals in the sample had ever smoked. The inclusion of this variable is meant to capture longer-run correlations between smoking and employment or earnings.

4.3 Other control variables

As discussed in Section 3.3, the regression specifications also include control variables for other factors that may be correlated with employment, earnings or disability. These comprise the following:

- Gender;
- Age group;
- Age of youngest child (interacted with adult gender);
- Limiting long standing illness or disability;
- Ethnicity;
- Highest educational qualification;
- Pregnancy;
- Caring for a disabled adult in the household;
- Region of residence;
- Housing tenure.

4.4 Variations on the main regression specification

As well as the main specification, a number of other specifications are estimated to explore the robustness of the results. The details of specifications are as follows:

Main specification: Smoking lags in Wave 7, Wave 2 and the "ever smoked" variable, plus control variables listed in Section 3.3.

Variant 1: Smoking lags only (with no other controls). This variant is designed to look at the 'raw' correlations between the outcome variables and smoking without taking any other controls into account.

Variant 2: Smoking lag in Wave 7 only, plus other controls.

Variant 3: Smoking lags in Wave 7 and Wave 2 only (no "ever smoked" variable), plus other controls. Variants 2 and 3 are simplified versions of the main specification which assess the impact of introducing lagged smoking variables and the "ever smoked" variable sequentially.

Variant 4: Current smoking (instead of lagged smoking variables) plus other controls. This variant is a comparison to show what happens to the coefficients on smoking if smoking data from Wave 8 instead of previous waves is used.

Variant 5: as variant 4 but using whole Wave 8 sample (including sample members who do not have complete data from earlier waves) as well as other controls. This is a comparison to show what difference it makes to the results if the regression includes all individuals with complete interviews in USoc Wave 8 who were asked about their current smoking status[2], including those individuals with incomplete data for USoc Waves 1 to 7.

[2] Note that Variant 5 does not include members of the Immigrant and Ethnic Minority Boost sample introduced in USoc Wave 6, as these sample members are not asked about smoking behaviour.

Variant 6: a random effects panel specification for Waves 5 to 8 using the lag of smoking in previous wave and the "ever smoked" variable plus current and lagged explanatory and control variables for all waves. This variant exploits information on employment and the other control variables across the four most recent waves of USoc which contain on smoking behaviour and the other covariates. This increases the number of individual observations being used by a factor of around 3 (all individual observations in Waves 6, 7 and 8 where data from the previous period exists in the survey as well) and may improve the accuracy of the estimates for the effect of smoking lagged one wave. The trade-off is that we are unable to include smoking lagged 6 waves in this specification as we only have this information for wave 8 (i.e. the wave 2 smoking variable).

Variant 7: additional variables for self-reported health and life satisfaction
As main specification but with additional binary variables included for the following responses to the self-reported health and life satisfaction questions in the USoc Wave 8 survey:

- Self-reported health "fair" or "poor";
- "completely dissatisfied" or "mostly dissatisfied" with life in general;
- "completely dissatisfied" or "mostly dissatisfied" with level of income.

These binary variables are included in variant 8 to control for potential health and motivational factors which could affect labour market outcomes. The health and life satisfaction indicators are probably endogenous as it is likely that causality runs from being non-employed or from low earnings to life satisfaction (and perhaps poor health as well) but they are included as an additional robustness check.

Variant 8: as main specification but with binary variable for smoking in Wave 7 interacted with various wave 8 characteristics as follows:

8a: age dummies
8b: gender
8c: disability
8d: highest qualifications

These variants assess whether the effect of smoking on employment (and earnings) varies according to other characteristics – for example, whether there is a larger effect for disabled people than non-disabled people. To the extent that variation in the smoking effects by other characteristics exists, it may provide more information about how the impacts are caused[3].

[3] It should be noted that not every interaction with other explanatory variables is explored in this paper because for some sample characteristics – e.g. ethnicity – many of the sample sizes in the USoc data for some categories (e.g. the BAME categories in this case) are too small to enable statistically significant breakdown effects of smoking to be estimated.

4.5 Separating the employment variable into unemployment and inactivity

As shown in Section 2 above, most previous research on the impact of smoking on labour market outcomes looks at the impact of smoking on *unemployment* rather than the overall impact of smoking on employment or inactivity. As a check on whether the relationship between smoking and employment (controlling for other factors) is strongest for unemployment, inactivity, or both, we estimate two additional versions of the main specification in this report:

a) a specification that uses unemployment in Wave 8 as the dependent variable (estimated on the sample of people in the labour force in Wave 8, i.e. the employed and the unemployed);
b) a specification that uses inactivity in Wave 8 as the dependent variable (estimated on the whole sample of people aged 21-69 in Wave 8, as with the employment regression).

4.6 Instrumental variable specifications

The regression specifications shown in Sections 4.2, 4.3 and 4.4 above attempt to control for the possible endogeneity of the smoking variable by including lags of the variable and a large set of other explanatory variables. In addition to this, a commonly-used econometric technique for controlling for causality is the **instrumental variables (IV)** approach. This involves using an additional 'instrumental' variable which is a determinant of smoker status, but which does not have a causal impact on the dependent variable in the employment or earnings regressions.

The variable used in the instrumental variable specification for these regressions is a binary variable for whether there are any other adults in the USoc member's household who smoke. This variable is only defined for households with more than one adult in them, and so we narrow the sample for these IV regressions to include only sample members s in households with more than one adult in them. IV regressions are estimated using Variant 5 (where current smoking is included as a regressor and the whole USoc Wave 8 sample is included) and instrumenting the current smoking variable using the "other household members who smoke" binary variable. Results from these IV regressions are reported for the employment and earnings regressions (together with a comparison specification which uses the same sample and specification (Variant 4 specification for the subsample of adults in USoc Wave 8 with at least one other adult in their household) but *without* the additional instrumental variable correction, for comparison.

Note that the IV specifications are estimated by the 2-Stage Least Squares (2SLS) technique which requires the main regression to be an OLS regression rather than a logistic regression, so the IV employment regression is estimated as a linear probability model instead of a logistic. This means that the coefficients from the IV employment model will not be strictly comparable with the other logistic specifications estimated for employment and earnings regressions in this report, but comparison of the coefficients from the IV model with the equivalent OLS model is

useful to assess whether the relationship between smoking and labour market outcomes is weaker when the IV technique is used to control for the endogeneity of the smoking variable.

5 Regression Results

5.1 Employment regressions

Coefficients on smoking variables from main specification and variants

Table 5.1 shows the coefficients on the smoking variable in the employment regressions for the main specification as well as the variants explained in Section 4.4 above.

Because the dependent variable (employment) is binary, the regressions use a logistic specification with the coefficients for each variable expressed as marginal effects – i.e. the change in employment rate associated with a unit change in the explanatory variable. In the case of the smoking dummy variables this corresponds to the change in employment associated with being a smoker in Wave 7 (or in Wave 2, or ever having smoked, or Wave 8, depending on the precise specification being estimated).

In Table 5.1 (and in subsequent tables in this chapter), coefficients that are statistically significant at the 5% level are shaded in grey. This significance level corresponds to an absolute value of the z-statistic of more than 1.96.

In the main specification, smokers in Wave 7 of USoc are around 5 per cent less likely to be in employment in Wave 8 than people who have never smoked, controlling for other factors. People who smoked in Wave 2 (but who gave up smoking by Wave 7) are around 2.5 per cent less likely to be in employment in Wave 8 than never-smokers. For people who were smokers in Wave 2 *and* Wave 7, the effect of smoking on employment is calculated by adding the Wave 2 and Wave 7 coefficients, giving a total reduction in the employment probability of around 7.5 per cent. That is, people who smoked in Wave 2 and Wave 7 are around 7.5 per cent less likely to be in employment than those who had never smoked. The coefficients for smoking in Wave 7 and Wave 2 are both statistically significant at the 5% level, although the z-statistic for the Wave 7 coefficient is higher than for the Wave 2 coefficient, suggesting that the Wave 7 result is more robust. The coefficient on the variable for ever having smoked is positive but is not statistically significant. Taken together, a Chi-squared test of the joint significance of the three smoking variables in the main specification suggests that they are significant at the 1% level (P = 0.000) while the same is true for a Wald test of the significance of the employment regression as a whole.

Overall, the main specification suggests a significant negative relationship between current smoking status in previous waves of the USoc survey and the probability of being in employment in Wave 8.

The variant specifications explore the relationship between smoking status and employment status in USoc in more detail. Variant 1 shows the ‘raw’ relationship between smoking and employment status without any control variables (I discuss the relationship between other control variables and smoking in more detail later in this

section). The coefficients for smoking in Wave 7 and Wave 2 are very similar to the main specification, but the coefficient for people who had ever smoked (but who were not smokers in Wave 2 or Wave 7) is different from the main specification; negative (at around -0.0028, or minus 2.8 percent) and statistically significant. This suggests that there is a negative relationship between being an ex-smoker in Wave 2 of USoc and being employed at Wave 8 which disappears once we control for other variables. (One of the most important variables to control for is disability status, which is discussed in more detail in Section 5.3 below).

Variant 2 shows that when a variable for smoking status in Wave 7 is included, but not the Wave 2 smoking or 'ever smoked' variables, the coefficient on the Wave 7 variable is -0.0665 – suggesting that smokers in Wave 7 are just under 7 per cent less likely to be in employment than non-smokers at Wave 7. This is similar to a population-weighted average of the Wave 7 and Wave 2 smoking variables in the main specification. Variant 3 shows that if the 'ever smoked' variable is omitted, the coefficients on the Wave 7 and Wave 2 smoking variables are similar to those in the main specification.

Variant 4 shows that when contemporaneous smoking status in Wave 8 is used in the regression (instead of lagged smoking status in Wave 7), the coefficient on smoking status shows a slightly bigger negative impact of smoking: around 7.5 per cent compared to 6.7 per cent in Variant 2. However, the difference between the two coefficients is not statistically significant. Variant 5 shows that when the full sample of available observations in Wave 8 is used (rather than the sample of individuals with complete data for Waves 1 through 8), the sample size expands from 16,515 to 20,949 individuals. The coefficient on smoking in Wave 7 shows slightly smaller effects with the larger sample (-0.069 compared to -0.0741) but the difference is not statistically significant.

Variant 6 shows the coefficients on smoking in the previous wave ("Wave t-1") when a random effects logistic specification is used, incorporating dependent variables from waves 6-8 and regressors from waves 5-7. This increases the sample size to 78,253 observations (multiplying the number of individuals by the number of time periods used). In this specification the coefficient on smoking in the previous wave is -0.0589, implying that smokers in each Wave of USoc are just under 6 per cent less likely to be employed than non-smokers in the next wave of the survey. This is a larger result than for Variant 2, which is the closest analogous specification to the Wave 8 logistic regression model.

Finally, in Variant 7, which includes additional explanatory variables for self-reported health status and life satisfaction, the employment penalty from smoking in Wave 7 is smaller (a coefficient of -0.037, meaning that Wave 7 smokers are just under 4 per cent less likely to be in employment, controlling for other factors), but is still statistically significant. The coefficient on smoking at Wave 2 is -0.0191 but is no longer significant at the 5% level. This suggests that most of the negative effect of smoking on employment is robust to the insertion of extra control variables into the regression.

The Wald tests for the significance of the whole regression show that the regression is significant at the 5% level for all of the variant specifications in Table 5.1.

Table 5.1 Coefficients on smoking variables in employment regressions: Main regression and variants

	Main specification		**Variant 1**		**Variant 2**		**Variant 3**		**Variant 4**		**Variant 5**		**Variant 6**		**Variant 7**	
Smoking in wave:	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]
Wave 8	-		-		-		-		-.0741	7.93	-.0649	7.62	-			
Wave 7/t-1	-.0498	3.74	-.0621	4.11	-.0665	7.28	-.0489	3.69	-		-		-0.0589	10.74	-.0370	2.82
Wave 2	-.0257	2.00	-.0298	2.06	-		-.0230	1.88	-		-		-		-.0191	1.51
ever	.0058	0.70	-.0280	3.05	-		-		-		-		0.0083	0.36		
Number of obs	16,515		16,515		16,515		16,515		16,515		20,949		78,253		16,515	
Pseudo R^2	0.2436		0.0066		0.2443		0.2435		0.2429		0.2320		N/A		0.2710	
Chi-sq test on combined smoking vars with P-value	Chi-2 (3) = 5055.90 (P = 0.000)		Chi-2 (3) = 105.05 (P = 0.000)		n/a		Chi-2 (2) = 55.28 (P = 0.000)		n/a		n/a		Chi-2 (2) = 122.99 (P = 0.0000)		Chi-2 (2) = 28.87 (P = 0.0000)	
Wald test for significance of regression with P-value	Chi-2 (50) = 2804.23 (P = 0.0000)		Chi-2 (3) = 105.05 (P = 0.0000		Chi-2 (48) = 2865.24 (P = 0.0000)		Chi-2 (49) = 2978.72 (P = 0.0000)		Chi-2 (48) = 2844.12 (P = 0.0000)		Chi-2 (48) = 3240.58 (P = 0.0000)		Chi-2 (48) = 6911.33 (P = 0.0000)		Chi-2 (53) = 3031.83 (P = 0.0000)	

*Wave 7 in main specification and Variants 1-5; t-1 in Variant 6
Note: grey shading indicates statistically significant result at the 5% level.
m.e. – marginal effect

Relationship between employment and other explanatory variables

Table A.1 in the appendix to this report shows a full set of results from the main specification of the employment regression, including the coefficients and z-statistics from all the control variables included in the regression. To summarise, the main relationships between employment status and control variables are as follows:

- The age groups most likely to be employed are 25-29 year olds and 30-34 year olds. The age groups least likely to be in work are the over-60s.
- Women are less likely to be in work than men overall, especially women in households where the youngest children is aged 10 or younger.
- Men with children are more likely to be in work than men without children.
- Pregnancy is negatively associated with employment.
- People who care for disabled relatives in their household are less likely to be in work than non-carers.
- Most differences in employment propensity between ethnic groups are not statistically significant. The only significant differences are that Indian, Pakistani and Bangladeshi individuals are less likely to be in work than other ethnic groups.
- Disabled people are less likely to be in work than non-disabled people.
- People with qualifications are significantly more likely to be in work than those without qualifications. This is especially the case for people with degrees or A-levels as their highest qualification.
- The regions with the highest employment rates, controlling for other factors, are the East Midlands, East of England, the South West and Wales. The North East and Northern Ireland have the lowest employment rates.
- People in homeowning households are more likely to be in employment than private tenants, who are in turn more likely to be in employment than social (local authority or housing association) tenants.

These findings are consistent with regression analysis of recent data from other UK cross-sectional micro data sets which have data on employment status and other explanatory variables (such as the Labour Force Survey) and are a useful check on the quality of the USoc panel data.

Interacting smoking behaviour at Wave 7 with other control variables

Table 4.2 shows the results from employment regression specifications which are identical to the main specification except that the ‘smoker in Wave 7’ variable is interacted with other control variables. There are four different specifications, listed as (i), (ii), (iii) and (iv) in the Table. The results from each specification are as follows:

(i) **Interaction with age group:** the negative correlation between smoking and employment is largest (and statistically significant) for individuals aged 30-39, 40-49 and 50-59 years. In each of these groups, smokers at Wave 7 are around 6 percent less likely to be in employment than non-smokers. The coefficients for the under-30 and 60-69 age groups are not significant.

(ii) **Interaction with gender:** while the coefficient on the smoking variable is statistically significant for both men and women, the negative impacts are much larger for men than for women. Male smokers at Wave 7 are around 7

per cent less likely to be in employment than non-smokers (conditional on other factors) whereas the equivalent figure for women is around 3.5 per cent.

(iii) **Interaction with disability:** While there is essentially no correlation between employment rates and smoking status for non-disabled adults, there is a substantial negative impact of smoking on employment for disabled adults (the coefficient implies that disabled smokers are around 12.5 per cent less likely to be in work than disabled non-smokers, controlling for other factors). Thus, the entire relationship between smoking status and employment is driven by a lower probability of employment for disabled smokers than non-disabled smokers. This is an important finding and one which we discuss further in Section 4.3 below on the relationship between disability and smoking status controlling for other factors.

(iv) **Interaction with highest educational qualifications.** The largest negative correlations between smoker status and highest educational qualification (and the only categories for which the coefficient on smoker status in Wave 7 is statistically significant) are for individuals whose highest qualification is A-level, and for those with no qualifications at all.

Finally, it should be noted that the chi-squared tests of the smoking interaction variables considered jointly in each of specifications (i), (ii), (iii) and (iv) are all significant.

Table 5.2. Coefficients on interaction of Wave 7 smoking variable with other explanatory variables: employment regressions

Specification (variant 8)	Marginal effect	|z|
i. Age group:		
21-29 years	-.0380	1.37
30-39 years	-.0617	2.80
40-49 years	-.0629	3.29
50-59 years	-.0590	3.19
60-69 years	-.0095	0.42
Joint test of age-smoking interactions		Chi-2 (5) = 20.59 (P = 0.0010)
ii. Gender:		
Male	-.0699	4.04
Female	-.0344	2.31
Chi-sq test of gender-smoking interactions)		Chi-2 (2) = 16.98 (P = 0.0002)
iii. Disability status:		
Disabled	-.1244	7.51
Non-disabled	.0008	0.05
Chi-sq test of disability-smoking interactions		Chi-2 (2) = 70.00 (P = 0.0000)
iv. Highest qualification:		
Degree	-.0119	0.28
Other HE	.0125	0.37
A Level	-.0726	2.10
GCSE	-.0260	1.45
Other	-.0356	0.47
None	-.0920	4.81
Chi-sq test of highest qualification-smoking interactions		Chi-2 (6) = 26.78 (P = 0.0002)

Note: grey shading indicates statistically significant result at the 5% level.

Specifications using unemployment and inactivity as the dependent variables

Table 5.3 shows the coefficients on alternative versions of the main specification from Table 5.1 which split the (inverse of the) dependent variable “in employment” into two separate variables – unemployed (in the left hand column) and inactive (in the right hand column). In the unemployment regression, none of the smoker variables is significant, and the smoker variables are also jointly insignificant. Therefore, there does not seem to be any identifiable relationship between previous smoking and unemployment in the USoc Wave 8 sample. By contrast, when inactivity is used as as the dependent variable there is a significant coefficient on the “smoker at Wave 7” variable (but not the other smoker status variables). The coefficient is 0.0193, suggesting that smokers at Wave 7 are just under 2 per cent more likely to be inactive in Wave 8 than non-smokers, controlling for other factors. This implies that the significant relationship between smoking and employment is

being driven by the impact of smoking on labour market inactivity rather than the impact of smoking on the probability of being unemployed.

Table 5.3. Coefficients on smoking variables using unemployment and inactivity as separate dependent variables

Dependent variable	Unemployed, wave 8		Inactive, Wave 8)	
	Marginal effect	|z|	Marginal effect	|z|
Smoker at Wave 7	-.0069	1.33	.0193	3.46
Smoker at Wave 2	-.0028	0.53	.0090	1.56
Ever smoked	.0038	1.04	-.0033	0.74
Number of obs	11,314		16,515	
Wave 8 sample (aged 21-69)	Employed + unemployed		Employed + unemployed + inactive	
R^2	0.2756		0.3492	
Joint test of smoking variables	Chi-2 (3) = 3.18 (P = 0.3640)		Chi-2 (3) = 45.41 (P = 0.0000)	
Test of regression significance	Chi-2 (47) = 389.10 (P = 0.0000)		Chi-2 (47) = 1072.86 (P = 0.0000)	

Instrumental variables regression

Table 5.4 shows the coefficients on the smoking variables from the instrumental variables (IV) employment regression using the presence of other smokers in the household as an instrument for smoker status in Wave 8 of USoc. The middle column of Table 5.3 reports the IV results, with the results from the equivalent OLS regression (without using the instrument) reported in the right hand column for comparison. The sample in these regressions is restricted to USoc sample members living in households with two or more adults in them; this reduces the sample size from 20,949 (in Variant 5 of Table 5.1) to 15,989 here. Table A.2 in the appendix gives a full set of coefficients from the IV employment regression.

Table 5.4 shows that the coefficient on the smoking variable is around -0.09 in the IV specification (suggesting that smokers are around 9 per cent less likely than non-smokers to be in employment). The comparable coefficient in the non-IV specification is around -0.07. Both coefficients are significant at the 5% confidence level but the difference between the two coefficients is not significant. The size of the employment effects of smoking estimated using the IV regression is similar to the marginal effects from the the logistic specifications shown in Table 5.1. A comparison of the coefficients in the two columns of table 5.3 suggests that controlling for endogeneity of smoker status using instrumental variables methods does not lead to a significant change in the size of the relationship between smoking and employment, controlling for other factors. This is futher evidence that the estimates in our main specification in Table 5.1 are valid.

Table 5.4. Coefficients on smoking variables from instrumental variables employment regression and OLS comparison

Specification (linear probability model, variant 5)	Instrumental variables (2SLS)		OLS comparison (no instrument)	
	Coeff	|t|	Coeff	|t|
Smoker at Wave 8	-0.0919	2.61	-0.0681	5.90
Number of obs		15,989		15,989
R2		0.2659		0.2663
Test of regression significance	Chi-2 (48) = 6317.90 (P = 0.0000)		F (48, 15940) = 132.70 (P = 0.0000)	

Note: grey shading indicates statistically significant result at the 5% level.

5.2 Earnings regressions

Coefficients on smoking variables from main specification and variants

Table 5.5 shows the coefficients on the smoking variable in the earnings regressions for the main specification as well as the variant specifications. Because the regressions here are OLS (linear) regressions using a continuous dependent variable (log weekly earnings) the coefficients can be interpreted as percentage impacts of the smoking variable on earnings.

The main specification shows a negative coefficient of -0.068 on log weekly earnings at Wave 8 for smoking at Wave 7. This implies an earnings penalty for smokers of just under 7 per cent. The coefficient on smoking status at Wave 2 is also negative but is a lot smaller in absolute terms (-0.022) and is not statistically significant. The coefficient on the “ever smoked” variable is also small and statistically insignificant. An F-test of the joint significance of the smoking variables shows that they are jointly significant.

Variant 1, with just the smoking indicators and no other control variables, shows a larger negative relationship between Wave 7 smoking and earnings (an earnings penalty of around 18 per cent) and a negative relationship between Wave 2 smoking and earnings (a penalty of just over 8 per cent). The variable for “ever smoked” is also associated with a small *positive* earnings effect (of around 5 per cent). All three coefficients are statistically significant at the 5% level. In other words the ‘raw’ correlations between smoking in Wave 2 and the “ever smoked” variable are significant but these correlations both become insignificant when other control variables are added. By contrast, the correlation between earnings and smoking in Wave 7 is robust to the inclusion of other control variables.

Variant 2 shows that when the Wave 7 smoking variable is included, but no other smoking variables, the earnings penalty for smokers is 8.5 per cent, while Variant 3 shows a smaller earnings penalty for Wave 7 smokers of 6.6 per cent when the

Wave 2 smoking variable is also included (although once again, the coefficient on the Wave 2 smoking variable is not statistically significant).

Variant 4 shows that when the Wave 8 smoking measure is used instead of the Wave 7 smoking measure, the earnings penalty to smoking is almost unchanged from Variant 2, at 8.7 per cent. Variant 5 shows that when the full Wave 8 sample is included rather than the balanced 8-wave sample (expanding the number of individual observations from 10,581 to 13,728) the coefficient on smoking is slightly lower (-0.070 rather than -0.087). However, the difference between the two coefficients is not statistically significant.

Variant 6 shows that in a random effects panel specification, the number of individual observations in the regression (including Waves 6 and 7 as well as Wave 8) expands to 51,259. The estimate for the earnings penalty from smoking at Wave t-1 is 7.2 per cent, which is very similar to the Wave 7 estimate from the main specification. Finally, Variant 7 shows that including additional variables for self-reported health status and life satisfaction in the earnings regression makes the coefficient on smoking status in Wave 7 smaller – to the point where it is no longer significant at the 5% level. The F-test of the three smoking variables shows that they are also jointly insignificant in this regression. However, the additional variables in Variant 7 are likely to be endogenous to earnings and so the results from Variant 7 should be treated as indicative only.

The F-test of the significance of the whole regression is significant in all eight regression specifications in Table 5.5.

Table 5.5 Coefficients on smoking variables in earnings regressions: Main regression and variants

	Main specification		**Variant 1**		**Variant 2**		**Variant 3**		**Variant 4**		**Variant 5**		**Variant 6**		**Variant 7**	
Smoking in wave:	Coeff	[t]	Coeff	[t]	Coeff	[t]	Coeff	[t]	Coeff	[t]	Coeff	[t]	Coeff	[t]	Coeff	[t]
Wave 8	-		-		-		-		-0.087	3.61	-0.070	3.32	-			
Wave 7/t-1	-0.068	2.26	-0.183	5.52	-0.085	3.49	-0.066	2.22	-		-		-0.072	6.54	-0.058	1.94
Wave 2	-0.022	0.79	-0.084	2.68	-		-0.017	0.64	-		-		-		-0.019	0.66
ever	0.011	0.57	0.051	2.41	-		-		-		-		-0.003	0.29	0.011	0.56
Number of obs	10,581		10,581		10,581		10,581		10,581		13,728		51,259		10,581	
R^2	0.2045		0.0095		0.2058		0.2435		0.2064		0.2108		N/A		0.2095	
Test of combined smoking vars with P-value	F (3, 10529) = 3.50 (P = 0.0149)		F (3, 10577) = 25.18 (P = 0.0000)		n/a		F (2, 10530) = 5.10 (P = 0.0061)		n/a		n/a		Chi-2 (2) = 46.42 (P=0.0000)		F (3, 10525) = 2.49 (P = 0.0586)	
Test of whole regression with P-value	F (51, 10529) = 52.123 (P = 0.000)		F (3, 10577) = 25.18 (P = 0.0000)		F (49, 10531) = 57.04 (P = 0.0000)		F (50, 10530) = 53.16 (P = 0.0000)		F (49, 10531) = 57.72 (P = 0.0000)		F (49, 13678) = 65.75 (P = 0.0000)		Chi-2 (49) = 6050.46 (P=0.0000)		F (55, 10525) = 51.41 (P = 0.0000)	

*Wave 7 in main specification and Variants 1-5; t-1 in Variant 6
Note: grey shading indicates statistically significant result at the 5% level.

Relationship between employment and other explanatory variables

As with the employment regressions discussed above, Table A.3 in this paper contains a complete set of coefficients and t-statistics for the relationship between the other explanatory variables (besides smoking) and weekly earning, for the preferred specification of the earnings regression. The main findings are as follows:

- Labour market history has a strong correlation with earnings. Workers who were not in work for two or more of the previous three waves of USoc have hourly earnings that are around 47 per cent lower than those who were in employment for two or more of the previous three waves.
- Weekly earnings are highest for individuals in the 35-39, 40-44 and 45-49 age groups, and lowest in the 65-69 age group.
- Compared to white British adults, Indian, Pakistani and Bangladeshi workers have significantly lower earnings, while the earnings of other ethnic groups are not significantly different.
- Disabled workers have lower earnings than non-disabled workers: the weekly earnings penalty to being disabled is around 7 per cent.
- Women have lower weekly earnings than men. The earnings penalty is bigger for women with children than those without children. Meanwhile, men with children have higher weekly earnings than men without children.
- People who care for a disabled person in their household have lower earnings than non-carers.
- There is an earnings premium for a degree or other form of higher education, but not for other qualifications (relative to a base of having no qualifications).
- Employed people in accommodation which they own outright or with a mortgage have higher earnings than private tenants, who in turn earn more than social tenants.
- The regions with the highest wages are London, the West Midlands and the East and South East of England.

Interacting smoking behaviour at Wave 7 with other control variables

Table 5.6 shows the results from earnings regression specifications where the ‘smoker in Wave 7’ variable is interacted with other control variables. The results from each specification are as follows:

(i) **Interaction with age group:** the only statistically significant negative correlation between earnings and Wave 7 smoking is for 40-49 year olds (an earnings penalty of around 15 per cent).

(ii) **Interaction with gender:** the negative impact of smoking on weekly earnings is bigger for women than men – this is the opposite of the finding for the effect of smoking on employment. For women, the earnings penalty for smokers is just over 9 percent (and statistically significant) whereas for men it is 4.5 per cent (and not significant).

(iii) **Interaction with disability:** The negative correlation between smoking and weekly earnings is much larger for disabled people (with an earnings penalty of just over 13 per cent) than for non-disabled people. For non-disabled people the earnings penalty is just under 5 per cent and is not statistically significant at the 5% level.

(iv) **Interaction with highest educational qualifications.** The two qualifications categories with significant negative relationships between smoking and weekly earnings are degree level (a very large earnings penalty of 43 per cent) and no qualifications (an earning penalty of 15 per cent).

The joint tests of the interaction variables are significant at the 5% level in the case of the disability-smoking interactions and the highest qualification-smoking interactions but not for the other interactions.

Table 5.6. Coefficients on interaction of Wave 7 smoking variable with other explanatory variables: earnings regressions

Specification (variant 8)	Coeff	|z|
i. Age group:		
21-29 years	-0.081	1.25
30-39 years	-0.042	0.86
40-49 years	-0.146	2.88
50-59 years	-0.030	0.60
60-69 years	0.071	0.76
Joint test of age-smoking interactions		F (5,10525) = 2.21 (P = 0.0503)
ii. Gender:		
Male	-0.045	1.26
Female	-0.091	2.29
Joint test of gender-smoking interactions		F (2, 10528) = 2.91 (P = 0.0546)
iii. Disability status:		
Disabled	-0.132	2.20
Non-disabled	-0.048	1.54
Joint test of disability-smoking interactions		F (2, 10528) = 3.04 (P = 0.0478)
iv. Highest qualification:		
Degree	-0.430	3.12
Other HE	-0.067	0.89
A Level	-0.041	0.55
GCSE	-0.025	0.61
Other	0.066	0.28
None	-0.150	2.06
Joint test of highest qualification-smoking interactions		F (6,10524) = 2.42 (P = 0.0245)

Note: grey shading indicates statistically significant result at the 5% level.

Instrumental variables regression

Table 5.7 shows the coefficients on the smoking variables from the instrumental variables earnings regression using the same instrument (presence of other smokers in the household) as in the employment regression. The coefficient on the smoking variable is around -0.22 in the IV specification compared to around -0.07 in the non-IV specification. This suggests that controlling for the endogeneity of smoking using instrumental variables methods actually leads to a *larger* estimated negative impact of smoking.

Table 5.7. Coefficients on smoking variables from instrumental variables earnings regression: Variant 4 specification, subsample of adults living in households with at least one other adult

	Instrumental variables (2SLS)		OLS comparison (no instrument)	
Specification (variant 4)	Coeff	|z|	Coeff	|t|
Smoker at Wave 8	-0.2199	2.65	-0.0668	2.79
Number of obs	10,730		10,730	
R2	0.2032		0.2068	
Test of regression significance	Chi-2 (49) = 2638.02 (P = 0.0000)		F (49, 10680) = 53.99 (P = 0.0000)	

Note: grey shading indicates statistically significant result at the 5% level.

5.3 Disability regressions

This report also estimates logistic regressions using disability status as the dependent variable and including the smoking variables plus the same set of explanatory variables as the employment regressions and earnings regressions in Sections 4.1 and 4.2 above (except of course for the disability variable, which is now the dependent variable rather than an explanatory variable). The reason for estimating these disability regressions is to dig deeper into the relationship between smoking and disability in the USoc data, controlling for other factors.

Table 5.8 below presents the marginal effects from disability regressions using a similar main specification and set of variants to the employment and earnings regressions presented in Tables 4.1 and 4.3 respectively. Overall, most of these results show that the relationship between 'ever smoked' and disability seems to be stronger than for recent smoking and disability. This is consistent with the hypothesis that smoking at an earlier stage in the life course has a particularly strong relationship to disability and health status, controlling for other factors. It is also consistent with the finding from the employment regressions where smoker status was interacted with disability; smoking status had a negative association with employment for disabled people but not for non-disabled people (controlling for other factors).

The detailed results from Table 5.8 confirm that the relationship between 'ever smoked' and disability is stronger than between recent smoking and disability. In the main specification, people who have ever smoked are 2.35 per cent more likely to be disabled than those who have never smoked, and this relationship is statistically significant at the 5% level. Smoking in wave 7 has a coefficient of a similar size to the "ever smoked" variable (0.0235), but the coefficient is not statistically significant ($|z| = 1.53$). The coefficient on smoking in Wave 2 is very close to zero, and not statistically significant.

In Variant 1 (with no controls), ever smoking *and* the Wave 7 smoking variable (but not the Wave 2 smoking variable) are both significantly and positively associated with disability. However, once other factors are controlled for, the additional impact of Wave 7 smoking on disability is no longer significant.

In Variant 2 (with just the Wave 7 control), Wave 7 smokers are just under 4 per cent more likely to be disabled than non-smokers at Wave 7. Including the Wave 2 control (in Variant 3) means that neither individual coefficient is significant (although an F-test across the two coefficients indicates they are jointly significant).

Variant 4 (using the contemporaneous Wave 8 smoking variable instead of Wave 7) shows that smokers are around 3.3 per cent more likely to be disabled, while variant 5 (on the full wave 8 sample) shows a marginal effect of similar magnitude (around 3 per cent).
Finally, Variant 6 (random effects logistic) shows strong positive associations between disability and both smoking variables (a 2.2 per cent higher probability of being disabled for those who ever smoked; with those who were also smokers in the previous wave being over 6 per cent more likely to be disabled in total). There is thus

a powerful association between disability in both the long and the short run, with bigger impacts than in the Wave 8 logistic model.

Table 5.8 Coefficients on smoking variables in disability regressions: Main regression and variants

	Main specification		Variant 1		Variant 2		Variant 3		Variant 4		Variant 5		Variant 6	
Smoking in wave:	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]	m.e.	[z]
Wave 8	-		-		-		-		.0330	3.10	.0305	3.14	-	
Wave 7/t-1*	.0235	1.53	.0427	2.82	.0390	3.71	.0269	1.74	-		-		.0393	6.71
Wave 2	.0033	0.23	.0228	1.57	-		.0139	0.98	-		-		-	
ever	.0235	2.62	.0492	5.31	-		-		-		-		.0224	6.15
Number of obs	16,515		16,515		16,515		16,515		16,515		20,949		78,253	
Pseudo R^2	0.0793		0.0063		0.0788		0.0789		0.0788		0.0777		N/A	
Chi-sq test on combined smoking vars with P-value	Chi-2 (3) = 20.16 (P=0.0002)		Chi-2 (3) = 100.12 (P=0.0000)		n/a		Chi-2 (2) = 13.69 (P=0.0011)		n/a		n/a		Chi-2 (2) = 112.95 (P=0.0000)	
Wald test for significance of regression with P-value	Chi-2 (49) = 1128.97 (P=0.0000)		Chi-2 (3) = 100.12 (P=0.0000)		Chi-2 (47) = 1150.79 (P=0.0000)		Chi-2 (48) = 1123.65 (P=0.0000)		Chi-2 (47) = 1159.85 (P=0.0000)		Chi-2 (47) = 1354.06 (P=0.0000)		Chi-2 (48) = 3534.24 (P=0.0000)	

Note: m.e. – marginal effect

6 Estimating the overall productivity losses from smoking

This section of the paper uses the results from the regression specifications estimated in Section 4, combined with some other aggregate statistics from the ONS relating to the UK labour market, to derive an estimate for the overall productivity losses to the UK economy arising from smoking. Table 5.1 shows the calculations involved in deriving a figure for the total productivity losses, which are estimated at around £14 billion – around 0.5% of current UK GDP.

6.1 Methodology

The calculations for the overall productivity losses from smoking use two results from the regressions in this report in particular. One is the impact of smoking on employment (from Table 5.1), and the other is the impact of smoking on earnings (Table 5.3). The calculations here use the coefficient on smoking in the previous wave of USoc (Wave 7) from the main specification of the employment and earnings regressions. The coefficient on smoking at Wave 2, and the "ever smoked" variable, are not included in the calculations. This means that the calculations give the (approximately) immediate impact on productivity if all smokers in the UK were to stop smoking immediately[4].

The calculation proceeds in four stages as follows. First, the marginal effect of smoking on employment (calculated from the employment regression as -0.0498) is combined with data on smoking prevalence among adults currently in work in the UK and statistics from ONS for total current employment in the UK to calculate the number of additional people employed if UK smoking prevalence were zero. Second, average annual earnings are calculated for current smokers in the USoc data (just under £21,000 per year). Third, the coefficient from the earnings regression in Variant 2 of Table 5.3 is used to estimate what the average earnings of smokers in the UK economy would be if they had never smoked (an increase of just over £1,400 per year)[5]. Finally, the overall productivity loss to the UK economy from smoking is calculated as the sum of two components:

a) The increase in overall earnings for current smokers already in employment if they had never smoked;
b) The additional earnings for smokers who are not currently in employment but would be if they had never smoked.

[4] The latest wave of USoc is lagged at least one year compared to current statistics from ONS on employment and earnings across the UK economy.
[5] Note that this calculation assumes that wages are a function of productivity, and therefore that higher wages for non-smokers are funded out of increased productivity for non-smokers compared to smokers.

6.2 Results and discussion

Table 6.1 shows the detailed calculations involved in these estimates. Component (a) is estimated at £7.2bn and component (b) at £6.9bn, leading to an overall estimate for the productivity losses arising from smoking in the UK economy of £14.1bn.

This analysis has revealed that smoking has a far greater impact on employment and earnings than was previously understood. In 2017 the Department for Health and Social Care (DHSC) estimated the output loss due to economic inactivity and unemployment, absenteeism and smoking breaks to be £6.3 bn for England (DHSC, 2017) and in 2019 it was estimated by ASH to be £6bn (ASH, 2019)

The £14.1bn includes previously uncalculated costs of under-employment linked to smoking, not just economic inactivity. In addition, previous analyses of the impact of economic inactivity only included smokers who had applied for incapacity benefit while this analysis includes all unemployed smokers. Less significantly, previous DHSC and ASH analyses were not UK-wide as this is, but only applied to England. There may be other reasons too, and further research is needed.

However, there are also additional impacts of smoking on economic output caused by early deaths among smokers of £3bn for England alone (ASH, 2019) which have not been included in this analysis.

Table 6.1. Calculations of overall productivity gains to the UK economy from immediate cessation of smoking in the population

Additional employment if UK smoking prevalence were zero		
Statistic	**Value**	**Source**
1: Total population aged 21-69 in UK	34.78m	ONS (2019a)
2: Total smoking prevalence in the age group 21-69	17.86%	Usoc Wave 8
3: marginal effect of smoking on employment	-0.0498	Marginal effect calculation based on smoking (t-1) variable in Table 4.1 main specification
4: Amount of additional employment for smokers aged 21-69 if they did not smoke	309,000	(1) x (2) x (3)
Wage levels for current smokers if UK smoking prevalence were zero		
5: Average annual earnings across UK economy	£26,104	ONS (2019b)
6: Average earnings for smokers as percentage of average earnings across whole economy	80.20%	Usoc Wave 8
7: Average annual earnings for smokers in UK economy	£20,935	(5) x (6)
8: earnings premium for non-smokers compared to smokers controlling for other factors	6.8%	Inverse of coefficient on smoking (t-1) variable in Table 4.3 main specification
9: increase in earnings for current smokers if they had never smoked	£1,424	(7) x (8)
10: total earnings for current smokers if they had never smoked	£22,359	(7) + (9)
11: total smoking prevalence among people in work	15.41%	USoc Wave 8
12: total UK employment	32.75m	ONS (2019c)
13: total number of smokers in employment	5.05m	(11) x (12)
Increase in UK productivity arising from:		
14: Increase in earnings for current smokers already in work	£7.2bn	(9) x (13)
15: Earnings of people who are not currently in work but would be if smoking prevalence were zero	£6.9bn	(4) x (10)
Total loss in UK productivity	**£14.1bn**	**(14) + (15)**

6 Conclusions

This results from the analysis in this research report suggest that smoking is negatively associated with the probability of being employed in the UK, and that there is an earnings penalty associated with smoking. The results from the preferred specification of the employment regression indicate that for adults in Wave 8 of the Understanding Society survey, smoking in Wave 7 was associated with being around 5 per cent less likely to be in employment, controlling for other factors. Smoking in Wave 2 – six years prior to Wave 8 – was associated with being a further 2.5 per cent less likely to be in employment. Meanwhile, smokers have weekly earnings that are on average 6.8 per cent lower than non-smokers.

The estimated associations from the employment and earnings regressions imply sizeable productivity losses from smoking. A simple calculation of the overall productivity losses arising from smokers being less likely to be in employment, and earning less than, non-smokers suggests that the total cost of smoking to the UK economy in terms of reduced productivity is just over £14 billion – around 0.5 per cent of total UK Gross Domestic Product.

For individuals who smoke and who are in work, these results imply that they face a double financial penalty from smoking. Their disposable income is reduced firstly because their wage is lower because they are a smoker (the wage penalty) but is then also further reduced by the need to spend considerable sums buying tobacco (the tobacco purchase penalty). Based on recent statistics from HMRC on total tobacco expenditure in the UK in 2016-17 (£8.91 billion)[6], smoking prevalence for the employed sample in USoc Wave 8 (16.2%), and a total UK population in the relevant age group in mid-2017 of 41.2 million (ONS, 2019), total tobacco purchase costs per smoker in employment can be estimated as £8.91 billion / (41.2m x 16.2%), which gives a result of £1,335 per year.

Thus, the average cost of smoking for smokers in employment is equal to £1,424 (in lower earnings) plus £1,335 (in tobacco purchase costs), giving total costs of £2,759 per year.

In addition, the results from regressions for the determinants of disability in the USoc sample suggest that there is a longer-term positive association between smoking and disability. Because the Understanding Society panel only currently features eight waves of data, it does not have the full life-course data necessary to examine longer-term relationships between smoking and health status in a high level of detail. While the British Household Panel Survey (which ran from 1991 to 2008 before being subsumed into USoc) has a longer run of data, it has a much smaller sample size than USoc and is unlikely to produce statistically significant findings for the smoking-disability relationship. However, other UK data sources such as the National Child Development Survey (which follows a cohort of everybody born in a particular week in 1958, reinterviewing cohort members every few years throughout their lifetimes) and the British Cohort Survey (which follows a cohort born in 1970) have relatively large sample sizes as well as detailed data on health status and disability, employment, earnings and smoking behaviour over the sample members' lifetimes.

[6] HMRC. Tobacco bulletin. November 2019.

Exploiting the data from cohort studies to look at the longer-term relationships between smoking and disability in the UK should be a high priority for future research in this field.

References

Action on Smoking and Health (2019), Ready Reckoner 2019 edition. November 2019. https://ash.org.uk/ash-ready-reckoner/

Böckerman P, Hyytinen A and Kaprio J (2014), "Smoking and long-term labour market outcomes", *Tobacco Control*, February 2014.

Brook J, Zhang C, Burke L *et al* (2014), "Trajectories of cigarette smoking from adolescence to adulthood as predictors of unemployment status. *Nicotine Tob Res*, Vol 16, pp 1559-1566.

Department for Health and Social Care [DHSC] (2017), *Towards a Smokefree Generation: A Tobacco Control Plan for England*: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630217/Towards_a_Smoke_free_Generation_-_A_Tobacco_Control_Plan_for_England_2017-2022__2_.pdf

Institute for Social and Economic Research (2018), *Understanding Society: The UK Household Longitudinal Study Waves 1-8 User Guide*. November 2018.

Jusot F, Khlat M, Rocherau T and Serment C (2008), "Job loss from poor health, smoking and obesity: a national prospective survey in France", *Journal of Epidemiology and Community Health* 62 (April), 332-337.

Levine P, Gustafson T and Velenchik A (1997), "More bad news for smokers? The effects of cigarette smoking on wages", *ILR Review*, Vol 50 No 3, pp 493-509.

Office for National Statistics [ONS] (2019), "Population estimates: Persons by single year of age and sex for local authorities": https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland

Office for National Statistics [ONS] (2019b), "Average weekly earnings in Great Britain: November 2019": https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/averageweeklyearningsingreatbritain/november2019

Prochaska J, Shi Y and Rogers A (2013), "Tobacco use among the job-seeking unemployed in California, *Prev Med* , Vol 56, pp 329-332.

Reed H (2010), *The Effects of Increasing Tobacco Taxation: A Cost Benefit and Public Finances Analysis*. London, ASH. http://www.ash.org.uk/tax/analysis

Schunck R and Rogge B (2012), "Unemployment and smoking: causation, selection, or common cause? Evidence from longitudinal data." SOEP Papers on Multidisciplinary Panel Data Research 491. Berlin, Germany: German Socio-Economic Panel Study (SOEP).

Semple S (2015), “Employment, Smoking, and Health: The Role of the Hygienist”, *Annals of Occupational Hygiene*, Vol 59 Issue 5, June 2015, pp 529-533. https://academic.oup.com/annweh/article/59/5/529/2196197

Weng S, Ali S and Leonardi-Bee J (2013), “Smoking and absence from work: systematic review and meta-analysis of occupational studies”, *Addiction*, Vol 108, pp 307-319.

Appendix

This appendix shows the detailed results from the main specification in Table 5.1 (employment regression), Table 5.3 (earnings regression) and Table 5.5 (disability regression).

Guide to variable names

The variables featured in the regression results shown in this appendix are labelled as follows:

Name of variable in tables	Description of variable
Dependent variables	
employed	Employed in USoc Wave 8
logearn	Log of weekly earnings, USoc Wave 8
disab	Long standing illness or disability, USoc Wave 8
Smoking dummy variables	
cursmoke L1.	Smoker – USoc Wave 7
cursmoke L6.	Smoker – USoc Wave 2
smoke_bf2	Ever smoked
Explanatory variables	**(all Wave 8):**
reg_NW	Region: North West England
reg_YH	Region: Yorkshire and the Humber
reg_EM	Region: East Midlands
reg_WM	Region: West Midlands
reg_EE	Region: East of England
reg_LO	Region: London
reg_SE	Region: South East England
reg_SW	Region: South West England
reg_WA	Country: Wales
reg_SC	Country: Scotland
reg_NI	Country: Northern Ireland
ten_own	Housing tenure: owner occupier
ten_la	Housing tenure: social tenant
age2529	Age: 25-29
age3034	Age: 30-34
age3539	Age: 35-39
age4044	Age: 40-44
age4549	Age: 45-49
age5054	Age: 50-54
age5559	Age: 55-59
age6064	Age: 60-64
age6569	Age: 65-69
eth_ow	Ethnicity: White, non-British/Irish

eth_mx	Ethnicity: Mixed
eth_in	Ethnicity: Indian
eth_pa	Ethnicity: Pakistani
eth_ba	Ethnicity: Bangladeshi
eth_oa	Ethnicity: Other Asian (e.g. Chinese)
eth_bl	Ethnicity: Black
eth_ot	Ethnicity: Other (e.g. Arab)
disab	Long standing illness or disability
female	Female
f_c0_1	Female, youngest child aged 0-1 years
f_c2_4	Female, youngest child aged 2-4 years
f_c5_10	Female, youngest child aged 5-10 years
f_c11_16	Female, youngest child aged 11-16 years
m_c0_1	Male, youngest child aged 0-1 years
m_c2_4	Male, youngest child aged 2-4 years
m_c5_10	Male, youngest child aged 5-10 years
m_c11_16	Male, youngest child aged 11-16 years
pregnant	Currently pregnant
carer	Caring for a disabled relative
hq_deg	Highest qualification: first/higher degree
hq_ohe	Highest qualification: other higher education
hq_al	Highest qualification: A Levels
hq_gcse	Highest qualification: GCSE
hq_oth	Highest qualification: other qualifications
_cons	Constant

Table A.1: Full results from employment regression, main specification

```
Logistic regression                                 Number of obs     =     16,515
                                                    Wald chi2(50)     =    2804.23
                                                    Prob > chi2       =     0.0000
Log pseudolikelihood = -8034.0747                   Pseudo R2         =     0.2436

------------------------------------------------------------------------------
             |               Robust
    employed | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cursmoke |
         L1. |   .7256308   .0622624    -3.74   0.000     .6133082    .8585245
         L6. |   .8471896   .0703447    -2.00   0.046     .7199507    .9969158
             |
   smoke_bf2 |   1.038042   .0550877     0.70   0.482     .9354972    1.151827
      reg_NW |     1.3759   .1656835     2.65   0.008     1.086642    1.742156
      reg_YH |   1.153909   .1443397     1.14   0.252     .9030189    1.474505
      reg_EM |   1.542829   .1911172     3.50   0.000     1.210252    1.966799
      reg_WM |    1.29652   .1642389     2.05   0.040     1.011467    1.661907
      reg_EE |   1.573838   .1910688     3.74   0.000      1.24057    1.996634
      reg_LO |   1.353318   .1780832     2.30   0.021      1.04566    1.751498
      reg_SE |   1.392652   .1628565     2.83   0.005     1.107396    1.751386
      reg_SW |    1.46277   .1803164     3.09   0.002     1.148809    1.862533
      reg_WA |   1.651769   .2157459     3.84   0.000     1.278703    2.133679
      reg_SC |   1.308282    .162605     2.16   0.031      1.02543    1.669153
      reg_NI |   .8715333   .1127622    -1.06   0.288     .6763203    1.123093
     ten_own |   1.192274   .0862526     2.43   0.015      1.03466    1.373898
      ten_la |   .4980944   .0422764    -8.21   0.000     .4217595    .5882452
     age2529 |   1.836956   .2865069     3.90   0.000     1.353126    2.493786
     age3034 |    1.85657    .278764     4.12   0.000     1.383259    2.491835
     age3539 |   1.696767   .2405993     3.73   0.000     1.285058    2.240379
     age4044 |     1.6693   .2311094     3.70   0.000     1.272589     2.18968
     age4549 |   1.609113   .2101143     3.64   0.000     1.245772    2.078426
     age5054 |   1.529558   .1987717     3.27   0.001      1.18563    1.973253
     age5559 |   .8519665   .1087516    -1.26   0.209     .6633894    1.094149
     age6064 |   .3316471   .0423258    -8.65   0.000     .2582517    .4259015
     age6569 |   .0598782   .0080952   -20.83   0.000     .0459399    .0780452
      eth_ow |   .9634857   .1231057    -0.29   0.771     .7500431    1.237668
      eth_mx |   .8094981   .1708889    -1.00   0.317     .5352087    1.224358
      eth_in |   .5448059   .0816575    -4.05   0.000     .4061257    .7308412
      eth_pa |    .320505    .048551    -7.51   0.000     .2381729    .4312978
      eth_ba |   .6315589    .128625    -2.26   0.024     .4236975     .941395
      eth_oa |   .9532803   .1964214    -0.23   0.816     .6365497    1.427608
      eth_bl |   1.027162   .1511688     0.18   0.856       .76978    1.370602
      eth_ot |   .5718765   .2218015    -1.44   0.150     .2674022    1.223037
       disab |    .405473   .0190497   -19.21   0.000     .3698037    .4445828
      female |   .7583183   .0406715    -5.16   0.000     .6826503    .8423736
        f_c0_1 |   .2759279   .0414687    -8.57   0.000     .2055275    .3704429
        f_c2_4 |   .5795755   .0692582    -4.56   0.000     .4585566    .7325328
      f_c5_10 |   .7399562   .0710153    -3.14   0.002     .6130758    .8930953
     f_c11_16 |   1.000367   .1035945     0.00   0.997     .8166042    1.225481
        m_c0_1 |   1.538105   .3630299     1.82   0.068     .9684602    2.442812
        m_c2_4 |   1.616942   .3261728     2.38   0.017       1.0889    2.401046
      m_c5_10 |   1.299928   .1867336     1.83   0.068     .9809455    1.722637
     m_c11_16 |   1.694878   .2699339     3.31   0.001      1.24043    2.315819
     pregnant |   .3090213   .0471877    -7.69   0.000     .2290918    .4168379
        carer |   .4888658   .0424171    -8.25   0.000     .4124146    .5794891
       hq_deg |   1.356648    .118037     3.51   0.000     1.143951    1.608893
       hq_ohe |   1.152586    .075951     2.16   0.031     1.012937    1.311488
        hq_al |    1.20338   .1053568     2.11   0.034      1.01363    1.428651
      hq_gcse |   1.109713   .0645657     1.79   0.074     .9901149    1.243757
       hq_oth |   .7523009   .1474122    -1.45   0.146     .5123892    1.104544
        _cons |   3.585389   .5730785     7.99   0.000     2.621094    4.904446
------------------------------------------------------------------------------
```

Table A.2: Full results from employment regression, instrumental variables specification

Instrumental variables (2SLS) regression

Number of obs	15,989
Wald chi2(48)	6317.90
Prob > chi2	0.0000
R-squared	0.2659
Root MSE	.39593

employed	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cursmoke	-.0919639	.0351798	-2.61	0.009	-.160915	-.0230127
reg_NW	.0203307	.0201687	1.01	0.313	-.0191993	.0598607
reg_YH	.0078081	.0207967	0.38	0.707	-.0329526	.0485688
reg_EM	.04714	.0208998	2.26	0.024	.0061771	.0881028
reg_WM	.0175819	.0211043	0.83	0.405	-.0237818	.0589455
reg_EE	.0411339	.0204038	2.02	0.044	.0011433	.0811245
reg_LO	.0235563	.0215699	1.09	0.275	-.01872	.0658325
reg_SE	.0194045	.0195297	0.99	0.320	-.0188731	.057682
reg_SW	.0464227	.0204201	2.27	0.023	.0064001	.0864453
reg_WA	.0401562	.0223882	1.79	0.073	-.0037238	.0840361
reg_SC	.0180304	.0206752	0.87	0.383	-.0224923	.0585531
reg_NI	-.0278406	.0227758	-1.22	0.222	-.0724804	.0167993
ten_own	.0060103	.0123359	0.49	0.626	-.0181677	.0301882
ten_la	-.1275466	.0166452	-7.66	0.000	-.1601706	-.0949227
age2529	.072385	.0211417	3.42	0.001	.0309479	.1138221
age3034	.1047257	.0201682	5.19	0.000	.0651967	.1442547
age3539	.1003499	.0195794	5.13	0.000	.0619749	.1387248
age4044	.0959795	.0191176	5.02	0.000	.0585097	.1334492
age4549	.0959656	.0182374	5.26	0.000	.060221	.1317102
age5054	.0897423	.018051	4.97	0.000	.054363	.1251217
age5559	.0057696	.0190085	0.30	0.761	-.0314863	.0430256
age6064	-.2149313	.0206725	-10.40	0.000	-.2554486	-.174414
age6569	-.5468717	.0190096	-28.77	0.000	-.5841298	-.5096136
eth_ow	-.0084577	.0191133	-0.44	0.658	-.045919	.0290037
eth_mx	-.0283569	.0366287	-0.77	0.439	-.100148	.0434341
eth_in	-.0678591	.0207778	-3.27	0.001	-.1085828	-.0271354
eth_pa	-.2029971	.0268243	-7.57	0.000	-.2555717	-.1504225
eth_ba	-.1023108	.0318492	-3.21	0.001	-.164734	-.0398875
eth_oa	-.0337256	.0336146	-1.00	0.316	-.099609	.0321579
eth_bl	.0209235	.0246093	0.85	0.395	-.0273098	.0691568
eth_ot	-.0499641	.0649577	-0.77	0.442	-.1772788	.0773507
disab	-.143688	.0086555	-16.60	0.000	-.1606524	-.1267235
female	-.0601964	.0091546	-6.58	0.000	-.0781391	-.0422536
f_c0_1	-.2367138	.0307365	-7.70	0.000	-.2969562	-.1764714
f_c2_4	-.0815455	.0219704	-3.71	0.000	-.1246067	-.0384843
f_c5_10	-.0425493	.0172354	-2.47	0.014	-.07633	-.0087687
f_c11_16	.0043794	.017644	0.25	0.804	-.0302022	.0389611
m_c0_1	-.0105515	.0226652	-0.47	0.642	-.0549745	.0338714
m_c2_4	.0380438	.0164981	2.31	0.021	.0057082	.0703794
m_c5_10	.0209955	.0152022	1.38	0.167	-.0088003	.0507913
m_c11_16	.0316688	.0175646	1.80	0.071	-.0027571	.0660948
pregnant	-.1990833	.0335321	-5.94	0.000	-.264805	-.1333615
carer	-.1387123	.0157919	-8.78	0.000	-.1696638	-.1077608
hq_deg	.0340562	.0131149	2.60	0.009	.0083515	.0597609
hq_ohe	.0246399	.010889	2.26	0.024	.0032978	.0459819
hq_al	.0374407	.0133277	2.81	0.005	.011319	.0635624
hq_gcse	.0059893	.0099531	0.60	0.547	-.0135185	.025497
hq_oth	-.0191212	.0326633	-0.59	0.558	-.0831401	.0448978
_cons	.8025056	.0262956	30.52	0.000	.7509672	.854044

Instrumented: cursmoke
Instrument: othsm

Table A.3: Full results from earnings regression, main specification

Linear regression

Number of obs	10,581
F(51, 10529)	52.13
Prob > F	0.0000
R-squared	0.2045
Root MSE	.77787

logearn	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
outlab	-.4658085	.0474868	-9.81	0.000	-.5588916	-.3727254
cursmoke						
L1.	-.0678948	.0300073	-2.26	0.024	-.1267148	-.0090749
L6.	-.022388	.0282677	-0.79	0.428	-.0777981	.0330221
smoke_bf2	.011194	.0195264	0.57	0.566	-.0270815	.0494695
reg_NW	.0178394	.0400793	0.45	0.656	-.0607237	.0964025
reg_YH	-.0174838	.0414771	-0.42	0.673	-.0987867	.0638191
reg_EM	-.0379852	.0420136	-0.90	0.366	-.1203398	.0443693
reg_WM	.0798413	.0402456	1.98	0.047	.0009522	.1587304
reg_EE	.0681813	.0418687	1.63	0.103	-.0138893	.1502519
reg_LO	.2255331	.0464744	4.85	0.000	.1344344	.3166317
reg_SE	.0693876	.0406611	1.71	0.088	-.0103159	.1490911
reg_SW	-.0448603	.041116	-1.09	0.275	-.1254554	.0357349
reg_WA	-.0485388	.0465478	-1.04	0.297	-.1397813	.0427036
reg_SC	.0244079	.0490651	0.50	0.619	-.0717689	.1205847
reg_NI	-.0763823	.048145	-1.59	0.113	-.1707557	.0179911
ten_own	.1539631	.0296134	5.20	0.000	.0959152	.2120109
ten_la	-.0975176	.0355237	-2.75	0.006	-.1671508	-.0278845
age2529	.1819469	.0499216	3.64	0.000	.0840911	.2798027
age3034	.1839546	.0457602	4.02	0.000	.0942559	.2736532
age3539	.3001957	.0427419	7.02	0.000	.2164135	.3839778
age4044	.299167	.0424958	7.04	0.000	.2158673	.3824668
age4549	.3215081	.0405844	7.92	0.000	.2419549	.4010613
age5054	.2665152	.041599	6.41	0.000	.1849733	.3480572
age5559	.1973237	.0428499	4.60	0.000	.1133298	.2813176
age6064	.0043279	.0470335	0.09	0.927	-.0878666	.0965224
age6569	-.4493381	.0795529	-5.65	0.000	-.6052768	-.2933994
eth_ow	-.0989948	.0479001	-2.07	0.039	-.192888	-.0051016
eth_mx	-.0438692	.0704204	-0.62	0.533	-.1819065	.0941681
eth_in	-.1154166	.0392282	-2.94	0.003	-.1923113	-.0385218
eth_pa	-.3518084	.0684516	-5.14	0.000	-.4859866	-.2176302
eth_ba	-.4937635	.1519122	-3.25	0.001	-.7915401	-.1959869
eth_oa	-.0827857	.0573309	-1.44	0.149	-.1951651	.0295937
eth_bl	-.1056849	.0538621	-1.96	0.050	-.2112648	-.000105
eth_ot	.0895202	.1676709	0.53	0.593	-.2391465	.4181868
disab	-.0665283	.0203951	-3.26	0.001	-.1065067	-.02655
female	-.3003524	.0213309	-14.08	0.000	-.342165	-.2585398
f_c0_1	-.2415455	.057781	-4.18	0.000	-.3548073	-.1282837
f_c2_4	-.1640489	.0421668	-3.89	0.000	-.2467039	-.081394
f_c5_10	-.2688236	.0353966	-7.59	0.000	-.3382076	-.1994395
f_c11_16	-.180814	.0359371	-5.03	0.000	-.2512575	-.1103704
m_c0_1	.1153813	.0505094	2.28	0.022	.0163734	.2143892
m_c2_4	.0634127	.0424225	1.49	0.135	-.0197433	.1465688
m_c5_10	.1039693	.031171	3.34	0.001	.0428682	.1650703
m_c11_16	.1011784	.0445794	2.27	0.023	.0137943	.1885625
pregnant	.0060855	.07597	0.08	0.936	-.14283	.155001
carer	-.1639777	.0399534	-4.10	0.000	-.2422939	-.0856614
hq_deg	.4198777	.0325871	12.88	0.000	.3560008	.4837546
hq_ohe	.2959147	.024272	12.19	0.000	.2483369	.3434925
hq_al	-.0115594	.0338535	-0.34	0.733	-.0779187	.0547999
hq_gcse	-.1056188	.0218294	-4.84	0.000	-.1484086	-.062829
hq_oth	-.0567327	.1088311	-0.52	0.602	-.2700624	.156597
_cons	7.268687	.0518999	140.05	0.000	7.166954	7.370421

Table A.4: Full results from earnings regression, instrumental variables specification

```
Instrumental variables (2SLS) regression          Number of obs   =     10,730
                                                  Wald chi2(49)   =    2638.02
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2032
                                                  Root MSE        =     .77437
```

logearn	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cursmoke	-.219884	.0829317	-2.65	0.008	-.3824272	-.0573408
outlab	-.4322369	.0469893	-9.20	0.000	-.5243342	-.3401396
reg_NW	.0190765	.0394392	0.48	0.629	-.0582229	.0963759
reg_YH	-.0008887	.0379612	-0.02	0.981	-.0752914	.0735139
reg_EM	-.0050759	.0401937	-0.13	0.900	-.0838542	.0737024
reg_WM	.0367735	.0403912	0.91	0.363	-.0423918	.1159388
reg_EE	.0278093	.0416191	0.67	0.504	-.0537627	.1093812
reg_LO	.1976442	.0454738	4.35	0.000	.1085172	.2867713
reg_SE	.0748023	.0382727	1.95	0.051	-.0002108	.1498154
reg_SW	-.0409558	.0402624	-1.02	0.309	-.1198687	.0379571
reg_WA	-.0198214	.0427553	-0.46	0.643	-.1036202	.0639773
reg_SC	.0403781	.0399145	1.01	0.312	-.0378529	.1186092
reg_NI	-.0282003	.0434991	-0.65	0.517	-.113457	.0570563
ten_own	.1344851	.0317217	4.24	0.000	.0723117	.1966584
ten_la	-.0878391	.0368855	-2.38	0.017	-.1601334	-.0155447
age2529	.2460464	.037283	6.60	0.000	.172973	.3191197
age3034	.298748	.0382396	7.81	0.000	.2237996	.3736963
age3539	.3928561	.0375507	10.46	0.000	.3192582	.466454
age4044	.3818154	.0375494	10.17	0.000	.30822	.4554108
age4549	.4205843	.0342588	12.28	0.000	.3534382	.4877304
age5054	.3597613	.0373983	9.62	0.000	.2864621	.4330606
age5559	.3151303	.0370378	8.51	0.000	.2425375	.3877231
age6064	.0844862	.0478419	1.77	0.077	-.0092822	.1782546
age6569	-.3210449	.0854253	-3.76	0.000	-.4884753	-.1536144
eth_ow	-.1033945	.0584656	-1.77	0.077	-.217985	.0111959
eth_mx	-.0616614	.0651439	-0.95	0.344	-.189341	.0660183
eth_in	-.1552098	.0400011	-3.88	0.000	-.2336106	-.076809
eth_pa	-.4339286	.0653159	-6.64	0.000	-.5619455	-.3059117
eth_ba	-.4813108	.1400022	-3.44	0.001	-.75571	-.2069116
eth_oa	-.1162258	.0554448	-2.10	0.036	-.2248956	-.0075559
eth_bl	-.1252052	.0545839	-2.29	0.022	-.2321878	-.0182227
eth_ot	.0571834	.1631509	0.35	0.726	-.2625864	.3769533
disab	-.088465	.0212716	-4.16	0.000	-.1301566	-.0467734
female	-.3561308	.021658	-16.44	0.000	-.3985797	-.3136819
f_c0_1	-.2282134	.0617635	-3.69	0.000	-.3492677	-.1071592
f_c2_4	-.1572101	.046795	-3.36	0.001	-.2489266	-.0654936
f_c5_10	-.2164426	.0379559	-5.70	0.000	-.2908349	-.1420503
f_c11_16	-.1178546	.0396183	-2.97	0.003	-.195505	-.0402041
m_c0_1	.080161	.0421218	1.90	0.057	-.0023961	.1627182
m_c2_4	.0256999	.0393245	0.65	0.513	-.0513748	.1027746
m_c5_10	.0897066	.032074	2.80	0.005	.0268427	.1525705
m_c11_16	.1009154	.0378218	2.67	0.008	.026786	.1750447
pregnant	-.0290753	.0812358	-0.36	0.720	-.1882946	.1301439
carer	-.1450586	.0387623	-3.74	0.000	-.2210313	-.0690859
hq_deg	.3981325	.0354655	11.23	0.000	.3286215	.4676435
hq_ohe	.3070371	.0271806	11.30	0.000	.2537642	.3603101
hq_al	.0027381	.0308791	0.09	0.929	-.0577839	.0632601
hq_gcse	-.0670341	.021953	-3.05	0.002	-.1100612	-.024007
hq_oth	-.104849	.0823865	-1.27	0.203	-.2663236	.0566255
_cons	7.226805	.0501028	144.24	0.000	7.128605	7.325005

```
Instrumented:  cursmoke
Instruments:   othsm
```

Table A.5: Full results from disability regression, main specification

```
Logistic regression                                     Number of obs     =     16,515
                                                        Wald chi2(49)     =    1128.97
                                                        Prob > chi2       =     0.0000
Log pseudolikelihood = -9807.5346                       Pseudo R2         =     0.0793

------------------------------------------------------------------------------
             |               Robust
       disab | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cursmoke |
          L1. |   1.124879   .0866478     1.53   0.127     .9672507    1.308196
          L6. |   1.016809   .0749022     0.23   0.821     .8801091    1.174741
              |
   smoke_bf2 |   1.125056   .0506854     2.62   0.009     1.029974    1.228915
      reg_NW |   .9478437   .0956285    -0.53   0.595     .7777831    1.155088
      reg_YH |   .8488601   .0902873    -1.54   0.123     .6891278    1.045616
      reg_EM |   .9122747   .0965692    -0.87   0.386     .7413467    1.122613
      reg_WM |   .9164178   .0986066    -0.81   0.417     .7421713    1.131574
      reg_EE |   .7698568   .0801756    -2.51   0.012     .6277154    .9441849
      reg_LO |   .6440371    .073511    -3.85   0.000     .5149369    .8055042
      reg_SE |   .7534733   .0743736    -2.87   0.004     .6209372    .9142984
      reg_SW |   .9319316   .0945125    -0.70   0.487     .7639393    1.136866
      reg_WA |   .9393906   .1042136    -0.56   0.573      .755816    1.167552
      reg_SC |    .763319   .0799661    -2.58   0.010     .6216319    .9373004
      reg_NI |   .7205669    .081692    -2.89   0.004     .5769949    .8998634
     ten_own |   .7265378   .0469326    -4.95   0.000     .6401366    .8246008
      ten_la |   1.868215   .1445373     8.08   0.000     1.605359    2.174109
     age2529 |   .9669524   .1565925    -0.21   0.836     .7039758    1.328166
     age3034 |   1.215591   .1823838     1.30   0.193     .9058899     1.63117
     age3539 |   1.474004   .2109112     2.71   0.007     1.113532    1.951169
     age4044 |   1.638668   .2296107     3.52   0.000     1.245146     2.15656
     age4549 |   1.893187   .2517274     4.80   0.000      1.45886    2.456821
     age5054 |   2.388083   .3142709     6.61   0.000      1.84515    3.090774
     age5559 |   3.000544   .3934613     8.38   0.000     2.320502    3.879878
     age6064 |   3.448117   .4547953     9.38   0.000     2.662633    4.465321
     age6569 |   3.951885   .5203023    10.44   0.000     3.053062     5.11532
      eth_ow |   .9936293   .1019859    -0.06   0.950     .8125633    1.215043
      eth_mx |   1.165111   .2134395     0.83   0.404     .8136422    1.668403
      eth_in |   .7464281   .1001615    -2.18   0.029     .5738087    .9709767
      eth_pa |   1.149985   .1765801     0.91   0.363     .8511188    1.553796
      eth_ba |   .8629026   .1833364    -0.69   0.488     .5689989    1.308616
      eth_oa |   .6093879    .126734    -2.38   0.017     .4053872    .9160469
      eth_bl |   .8048194   .1142677    -1.53   0.126       .60932    1.063045
      eth_ot |   .9467104   .3475112    -0.15   0.881     .4610676    1.943881
      female |   1.162832   .0534158     3.28   0.001     1.062713    1.272382
       f_c0_1 |   .7038604   .1267728    -1.95   0.051     .4945111    1.001837
       f_c2_4 |   .7272871   .0875012    -2.65   0.008     .5745082    .9206944
      f_c5_10 |    .791904    .069964    -2.64   0.008     .6659931    .9416193
     f_c11_16 |   .8545615   .0766308    -1.75   0.080     .7168262    1.018762
       m_c0_1 |   .5515131   .1130592    -2.90   0.004     .3690294    .8242343
       m_c2_4 |   .7089457   .1012555    -2.41   0.016     .5358457    .9379641
      m_c5_10 |   .5857319   .0683133    -4.59   0.000     .4660411    .7361624
     m_c11_16 |   .7503397    .084236    -2.56   0.011     .6021418    .9350118
     pregnant |   .6820806   .1418597    -1.84   0.066     .4537348    1.025343
        carer |   1.512163     .11299     5.53   0.000     1.306159    1.750657
       hq_deg |   .8133227   .0629645    -2.67   0.008      .698821    .9465855
       hq_ohe |   .8654275     .04945    -2.53   0.011     .7737373    .9679833
        hq_al |   .8940025   .0679297    -1.47   0.140     .7703022    1.037567
      hq_gcse |   .9124728   .0466202    -1.79   0.073     .8255249    1.008578
       hq_oth |   1.089975   .1858345     0.51   0.613     .7803538    1.522445
        _cons |    .310465   .0485199    -7.48   0.000     .2285523    .4217349
------------------------------------------------------------------------------
```