Landman Economics

The impact of smoking, heavy drinking and obesity on employment prospects, earnings and productivity: analyses using UK data

Howard Reed, Landman Economics

November 2023

Acknowledgements

The author would like to thank members of the project steering group for comments on previous drafts, in particular Katherine Severi from the Institute of Alcohol Studies and Aalaa Jawad, Emily Reed and Hazel Cheeseman from ASH.

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

This research was funded by Cancer Research UK (Grant PICADV-Feb22\100004).

Contents

1 Introduction

It is well established that smoking is associated with lower productivity and losses in economic output due to increases in mortality and morbidity for the smoking population (Reed 2023; DHSC 2017). Previous work for ASH published in 2021 looked at the relationship between smoking and employment status and smoking and earnings using data from a British longitudinal dataset, Understanding Society (USoc) (Reed 2021). The aim of this technical paper is to extend and adapt the methodology used in the 2021 ASH research to look at the relationship between the labour market outcomes related to productivity (employment and earnings), and:

- Alcohol consumption (focusing in particular on heavy drinking as measured by the AUDIT-C test score);
- Obesity (measured by Body Mass Index, or BMI).

Most of the analysis in this report uses the Understanding Society panel dataset. At the time of Reed (2021) only Waves 1 through 8 of the survey were available, whereas an additional four waves (9 to 12) are now in the public domain. Therefore, this analysis uses the most recent wave of the data. Where possible, previous waves of the dataset are also used to take advantage of the panel structure of the data to improve the robustness of the estimates. The analysis also uses earlier waves of data to rest for the potential impacts of the Covid-19 pandemic on the estimated results where feasible.
The information available on obesity in the Understanding Society Data is more limited than for smoking or drinking, so this report also uses data from the Health Survey for England (HSE), a cross sectional survey dataset which contains additional data on obesity and overweightness in the survey sample, to supplement the analysis.

This research report is a part of the wider cross-risk-factor project and is designed to support the project narrative by providing new estimates of the relationship between the main labour market outcome indicators (employment and earnings) and smoking, drinking and obesity. The results are used to produce estimates of the aggregate productivity losses to the UK economy resulting from smoking, heavy drinking and obesity.

The cross-risk factor project looks at the economic impacts of high-risk consumption of tobacco, alcohol and unhealthy foods (see Jawad and Reed 2023 for the final report from the project, *Holding Us Back*). While smoking and drinking can be measured directly, in this report we have chosen to use obesity as a key indicator for the long-term harms of consuming unhealthy food.

The report is structured as follows.

Chapter 2 summarises previous evidence on the productivity impacts of smoking, heavy drinking and obesity. Chapter 3 gives details of the Understanding Society data and presents descriptive statistics from the USoc sample on employment rates for smokers and non-smokers and for drinkers (stratified by AUDIT-C score) and obese people according to characteristics such as gender, disability status and highest qualification. Chapter 4 gives details of the Health Survey for England data.

Chapter 5 presents descriptive statistics on the relationship between labour market outcomes and smoking, heavy drinking and obesity. Chapter 6 explains the methodology and specifications for the regressions for the relationship between smoking, employment and earnings, while Chapter 7 does the same for the regressions showing the relationship between heavy drinking and employmenr and earnings and Chapter 8 does the same for the obesity regressions. Chapters 9, 10 and 11 present the results for the smoking regressions, drinking regressions and obesity regressions respectively. Chapter 12 presents estimates for the overall cost of smoking, heavy drinking and obesity in terms of reduced productivity at a national level. Chapter 13 offers conclusions.

2 Recent research on the relationship between smoking, drinking and obesity and employment and earnings

This chapter outlines the results from previous research on the relationship between smoking, alcohol consumption and obesity and employment and earnings[1].

2.1 Smoking, employment and earnings

Reed (2021)

Previous work for ASH by this author found used data from Understanding Society Waves 1 to 8 (inclusive) to analyse the relationship between employment and earnings, and smoking. The regressions controlled for a range of other factors including age, ethnicity, disability, gender, age of youngest child, pregnancy, being a carer for another person in the household, highest qualification region of residence and housing tenure. The main findings were as follows:

- Controlling for other factors, the employment penalty for smoking was just under 5 percentage points (marginal effect of smoking in Wave 7 on employment at Wave 8: -0.0498).
- Controlling for other factors, the wage penalty for smoking was almost 7 per cent (coefficient on smoking variable in log earnings regression: -0.068).
- At the aggregate (UK-wide) level, smoking was associated with productivity losses of around £14.1 billion in 2019. This comprised:
    - £7.2 billion due lower earnings for smokers who were in work but earning less than non-smokers (controlling for other factors);
    - £6.9 billion due to smokers being less likely to be in employment than non-smokers (controlling for other factors).

Other UK research

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.

[1] There is an additional summary of international evidence on the impact of smoking on earnings and employment in Reed (2021).

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.

2.2 The economic costs associated with alcohol consumption

Overall UK economic costs

In general there is less evidence for the UK on the economic costs associated with alcohol consumption than there is for smoking. The overall costs are difficult to estimate accurately and there is significant debate around which types of cost to include.

A 2020 briefing by the Institute of Alcohol Studies (IAS) on the costs of alcohol to society identifies three major pieces of work in the last twenty years:

1. A 2003 Cabinet Office Strategy Unit report which surveyed the evidence on the cost of alcohol to society in England and Wales, and estimated the 'external' cost of alcohol to society in England and Wales to be £21 billion – this external cost was defined as the costs imposed by drinkers on others, excluding any personal impact (Cabinet Office, 2003).
2. A 2007 study by researchers at the University of York (commissioned by The Scottish Government) found that the total cost of alcohol consumption for Scotland in 2007 was £3.6 billion, which comprised roughly £2.1 billion of external costs and £1.1 billion of private costs (York Health Economics Consortium 2010). This study included costs to the social care system in the calculation (whereas the 2003 Cabinet Office Study for England and Wales excluded them).
3. A cost of alcohol study from the Northern Ireland calculated the external costs of alcohol misuse to Northern Ireland in 2008/09 to be just under £700 million. This comprised the costs to the NHS, social work, fire and police services, courts and prison services and the wider economic costs. The costs to the wider economy were estimated at just over £200 million (Department of Health, Social Services and Public Safety, 2010).

It is important to note that lost productivity is only a small element of these estimates of social cost. For example the York research for the Scottish Government estimated that on average people in employment turned up for work with a hangover 2.5 days a year, and were 27% less productive than normal on those days. All of these estimates look at a much wider range of cost than productivity and so the results produced in this report will not be like-for-like comparable with the earlier figures for

UK economic costs presented here. Below we present some evidence from other studies (mainly for other countries) which are more comparable.

Evidence on the relationship between alcohol and employment

The recent evidence on the relationship between alcohol and employment status from other countries is mixed. Mangot-Sala, Smidt and Liefbroer (2022) find no significant impact of alcohol consumption on unemployment using data for the Netherlands. De Sio et al (2020) find a negative relationship between work and alcohol consumption using data from Italy but interpret this as a causal relationship running the opposite way (from work to alcohol consumption) on the grounds that employed people are less likely to drink alcohol than the unemployed controlling for other factors.

Evidence on the relationship between alcohol and earnings

There is some evidence on the relationship between alcohol consumption and earnings for the US. Heien (1996) uses regression analysis to estimate the effect of alcohol consumption on earnings controlling for other explanatory variables such as age, education and socioeconomic status. The results show a concave quadratic relationship between earnings and alcohol consumption – in other words, an “upside-down U shape”. Moderate drinkers have higher earnings (controlling for other factors) than either non-drinkers or heavy drinkers. The non-linear relationship between alcohol consumption and earnings helps explain apparent contradictions in the previous literature on earnings and alcohol consumption, which often assumed a “straight-line” relationship.

Alcohol consumption and impaired work performance

Thørissen et al (2019) conduct a systematic review of the relationship between alcohol consumption and impaired work performance (presenteeism) comprising 26 studies, containing 132 tested associations. The majority of associations (77%) indicated that higher levels of alcohol consumption were associated with higher levels of impaired work performance. The authors report that “evidence does provide some support for the notion of alcohol-related presenteeism. However, due to low research quality and lack of longitudinal designs, evidence should be characterised as somewhat inconclusive”.

2.3 The economic costs associated with obesity

Overall UK economic costs

As with the costs associated with alcohol consumption analysed in Section 2.2 above, most of the existing work on the economic costs of obesity looks at the wider economic costs rather than just the productivity costs. As with alcohol, the results for overall economic costs in this section are not like-for-like comparable with the earlier research as our focus is narrower, looking at just the productivity costs due to lower employment and earnings for people with obesity relative to the non-obese (controlling for other factors)

Research by Public Health England (2017) estimated the overall costs of obesity to wider society at £27 billion. More recently, the consultancy Frontier Economics (2022) found that the annual social cost of obesity in the UK was around £58 billion in 2020, equivalent to around 3% of 2020 UK GDP. As part of this measure, Frontier calculated the loss of workplace productivity through the following methodology:

1. Collecting evidence on the additional number of sick leave days taken by individuals with obesity every year.
2. Calculating the average daily wage in the UK.
3. Multiplying the results from steps 1 and 2 by the population with obesity in employment in England and the devolved nations.

Frontier Economics' estimate of the productivity costs of obesity in the UK is around £2 billion[2].

The relationship between obesity and incomes

Kim and van dem Knesebeck (2018) analyse the relationship between income and obesity using a meta-analysis of studies from the UK and US and find some evidence of negative causal links running both ways. In other words, lower income is associated with higher risks for subsequent obesity, but obesity is also associated with higher risks for subsequent lower income. Note however that this study looks at the relationship between obesity and *income* from all sources rather than just earnings.

[2] This is calculated using the findings on page 21 of the Frontier Economics report (2022).

Obesity and employment

Monsivias et al (2015) analyse data from the British Household Panel Study (the predecessor of the Understanding Society dataset used for most of the results in this report) and the European Prospective Investigation of Cancer (EPIC)-Norfolk study who were followed up over 26 months and 43 months, respectively. In both samples, changes in weight were computed for each participant and assessed in relation to three employment transitions: maintaining paid employment, retirement and job loss. In both studies, a significant positive relationship between losing work and subsequent weight gain was found. The authors report that “two UK-based samples of working adults reveal strong associations between job-loss and excess weight gain. The mediating behaviours are so far unclear but psychosocial mechanisms and sleep-loss may contribute to the excess weight gain among individuals who become unemployed”.

3 The Understanding Society (USoc) data

This chapter looks at the Understanding Society data used for most of the analysis in this report – updating the information on smoking variables and explanatory (control) variables given in Reed (2021) and also including detailed information on measures of alcohol consumption and obesity in the 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 twelve 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 2020, interviews were carried out for the first half of the wave 12 sample and the second half of the wave 11 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. Due to attrition, the number of full individual interviews had fallen to just over 29,000 by Wave 12.

Table 3.1. Understanding Society Waves 1-12: 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	50,688
3	January 2011-December 2012	45,862
4	January 2012-December 2013	43,136
5	January 2013-December 2014	40,975
6	January 2014-December 2015	41,865
7	January 2015-December 2016	39,337
8	January 2016-December 2017	37,610
9	January 2017-December 2018	34,959
10	January 2018-December 2019	33,514
11	January 2019-December 2020	31,542
12	January 2020-December 2021	29,072

Source: Institute for Social and Economic Research (2022).

3.2 Attrition across Waves 1 to 12

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 – over 25 per cent (1 in 4) of the sample dropped out. The rate of attrition slowed in subsequent waves, but by wave 6 more than half the wave 1 sample were no longer in the survey, and by Wave 12 only just over 30 per cent (3 in 10) of the sample from Wave 1 were left. This means that the subsample of individuals who were interviewed in all of Waves 1 to 12 inclusive could be a lot less representative of the general UK population than the sample from Wave 1. Similarly, the composition of the 15,062 interviewed adult sample members who were interviewed in all of Waves 1 to 12 may be quite different from the full sample of 31,542 adult interviews in Wave 12 – including the IEMB boost sample, the BHPS legacy sample and also other sample members who completed some, but not all, of the 12 USoc Wave interviews.

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	Percentage of original sample
1 (start)	n/a	47,732	100.0%
2	12,268	35,464	74.3%
3	4,764	30,700	64.3%
4	2,781	27,919	58.5%
5	2,035	25,884	54.2%
6	2,917	22,967	48.1%
7	1,263	21,704	45.5%
8	1,288	20,416	42.8%
9	1,687	18,729	39.2%
10	937	17,792	37.3%
11	1,346	16,446	34.5%
12 (most recent)	1,384	15,062	31.6%

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 12.
- 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 1-12

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
9	As wave 7	As wave 7
10	As wave 7	As wave 7
11	As wave 7	As wave 7
12	As wave 7	As wave 7

3.4 Questions on drinking in the USoc data

Table 3.4 shows which questions about drinking are asked in which wave of the Understanding Society survey.

Table 3.4. Drinking questions in Understanding Society Waves 1-12

Wave	Smoking questions	Sample asked smoking questions
1	none	none
2	• Ever had an alcoholic drink • How old were you the first time you had an alcoholic drink? • On how many days did you have an alcoholic drink? • How often have you had an alcoholic drink over the past 12 months? • Drank alcohol in the last 7 days	Whole sample
3	• Frequency of five or more alcoholic drinks • Ever had an alcoholic drink • How many times intoxicated in last 4 weeks	Whole sample
4	• Frequency of five or more alcoholic drinks • How many times intoxicated in last 4 weeks	Whole sample
5	As wave 2	Whole sample
6	• Frequency of five or more alcoholic drinks • Ever had an alcoholic drink • How many times intoxicated in last 4 weeks	Whole sample
7	• Past 12 months alcoholic drink • Always been non-drinker • Alcohol frequency past 12 months • Drinks on typical day • Frequency of having six or more drinks	Whole sample
8	• Frequency of five or more alcoholic drinks • Ever had an alcoholic drink • How many times intoxicated in last 4 weeks	Whole sample
9	As wave 7	Whole sample

10	• Frequency of five or more alcoholic drinks • Ever had an alcoholic drink • How many times intoxicated in last 4 weeks	Whole sample
11	As wave 7	Whole sample
12	As wave 7, plus: • Frequency of five or more alcoholic drinks • Ever had an alcoholic drink • How many times intoxicated in last 4 weeks	Whole sample

Source: Understanding Society variable guide

Although every wave (except Wave 1) includes some drinking questions, the most useful for our purposes are the three questions which make up the AUDIT-C questionnaire (NIDA CTN, 2023). AUDIT-C (Alcohol Use Disorders Identification Test-Concise) is a brief alcohol screening instrument that is intended to identify people who are hazardous drinkers or have active alcohol use disorders (including alcohol abuse or dependence). The AUDIT-C is a condensed version of the 10-question AUDIT (Alcohol Use Disorders Identification Test) questionnaire which was developed by the World Health Organisation (WHO, 2001).

The AUDIT-C has 3 questions and in USoc, is scored on a scale of 5-15. Each AUDIT-C question has 5 answer choices valued from 1 points to 5 points[3]. On this scale, A score of 7 or more for men, or 6 or more for women, is considered positive for identifying hazardous drinking or active alcohol use disorders.

The exact questions and scoring are shown in Table 3.5 below.

[3] The version of AUDIT-C outlined in the US National Institute on Drug Abuse guidance document (NIDA CTN, 2023) is scored from 0 to 12; the USoc version is the same except that each answer is coded from 1 to 5 rather than 0 to 4 (i.e. the overall AUDIT-C score is always 3 points higher in USoc than the original NIDA version).

Table 3.5. Scoring of AUDIT-C questionnaire responses in Understanding Society

Question	How often do you have a drink containing alcohol?	How many standard drinks containing alcohol do you have on a typical day?	How often do you have six or more drinks on one occasion?
Scoring:			
1 point	Never	1 or 2	Never
2 points	Monthly or less	3 to 4	Less than monthly
3 points	2-4 times a month	5 to 6	Monthly
4 points	2-3 times a week	7 to 9	Weekly
5 points	4 or more times a week	10 or more	Daily or almost daily

3.5 Questions on obesity in the USoc data

The information on obesity in the Understanding Society dataset is much less detailed than for smoking or drinking. There are only two waves of USoc that contain any information on obesity:

- Wave 1 has a BMI (body mass index) continuous variable.
- Wave 12 has a variable for whether the respondent is very overweight (has a BMI of 40 or above). This variable was introduced as a control variable for underlying health conditions for analysing the impact of Covid-19 in USoc, but it is asked of all respondents in Wave 12 (not just those who have Covid currently or previously).

3.6 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 12 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 12 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 20 to 69 inclusive. People aged under 20 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)

4 The Health Survey for England (HSE) data

This chapter examines the Health Survey for England (HSE) data, which is used for some of the analysis of the relationship between obesity and employment in this report because it has more detailed information on obesity than the USoc data.

4.1 The scope and sample size of the 2019 HSE

The Health Survey for England is an annual repeated cross-sectional dataset with detailed information on health status plus a number of other explanatory variables (described below). This report uses the 2019 HSE data, which was the most recent wave of data at the time of writing. The 2019 HSE consists of 8,204 interviews with people aged 16 and over plus 2,095 children aged under 16. For this report, only adults aged 20 to 69 were used (6,241 interviews in total).

4.2 Obesity questions in the HSE

The HSE has 2 variables which can be used to identify obesity:

i) a **continuous BMI (Body Mass Index)** variable.
ii) A **grouped BMI** variable which divides the population into four groups:
   a. underweight (BMI less than 18.5)
   b. normal weight (BMI between 18.5 and 25)
   c. overweight (BMI between 25 and 30)
   d. obese (BMI between 30 and 40)
   e. morbidly obese (BMI over 40).

This variable allows us to replicate the same dummy variable as for the obesity analysis in the USoc data (BMI more than 40) as well as including dummy variables for overweight and obese (BMI between 25 and 30 and BMI between 30 and 40 respectively).

4.3 Labour market status in the HSE

The HSE has an employment variable but no earnings variable. This means that it can be used to run employment regressions to look at the relationship between obesity and employment, but it cannot be used to run earnings regressions.

4.4 Other variables in the HSE

The HSE has a range of other covariates that we use in the regressions for obesity and employment as explanatory variables, as follows:

- gender;
- age dummies (five-year age categories from 20 to 69, similarly to USoc);
- number of children in the household (dummies for 1, 2, 3 or more);
- highest educational qualification (defined as for USoc);
- receipt of Personal Independence Payment, Disability Living Allowance or Attendance Allowance (used as a proxy for disability, which is not separately recorded in HSE);
- long standing illness (dummy variable)
- caring full-time for a disabled adult in the household (defined similarly to USoc);
- ethnicity (defined as in USoc);
- currently pregnant;
- region (9 English regions);
- housing tenure (defined as for USoc).

In general, our aim in designing the HSE regressions for employment and obesity was to match the set of USoc control variables as closely as possible.

5 Descriptive statistics

This subsection presents some descriptive statistics from the USoc data on smoking, drinking and obesity and earnings and employment rates. We also present some descriptive statistics on obesity from the HSE data.

5.1 Smoking prevalence over time in USoc

Table 5.1 shows smoking prevalence rates for adults aged 20 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 12 USoc waves. Only Waves 2 and 5 to 12 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 24.2% of adults aged 20-69 in Wave 2 to 14.0% of adults in Wave 12. The average rate of decline in smoking prevalence across Waves 2 through 12 is just over 0.9 percentage points per year.

Table 5.1. Smoking prevalence over time in Understanding Society, Waves 2 and 5 to 12

Wave	Smoking prevalence (% of sample, weighted)
2	24.8
5	21.3
6	20.2
7	18.5
8	17.4
9	16.3
10	15.8
11	14.9
12	14.0

Sample: adults aged 20-69, balanced panel with full data for waves 1 to 12 inclusive.

5.2 Prevalence of drinking (and problem drinking) over time in USoc

Table 5.2 shows three statistics relating to alcohol consumption in the Usoc data for Waves 7, 9, 11 and 12. These are:

- the proportion of adults in the sample drinking any alcohol at all (left-hand column);
- the proportion of adults with an AUDIT-C score indicating possible problem drinking (score of more than 6 for men, more than 5 for women);

- the proportion of adults with an AUDIT-C score of 11 or more, which we define as “heavy” drinkers. This is the definition of heavy drinking that we use for the regression analyses of alcohol consumption, employment and earnings in later chapters.

Table 5.2. Percentage of respondents drinking alcohol and with AUDIT-C scores indicating possible problem drinking and heavy drinking in Understanding Society, Waves 7, 9, 11 and 12

Wave	% who drank at all in last 12 months	% with AUDIT-C score>6 (men), Score>5 (women)	% with AUDIT-C score>10
7	85.0	35.5	8.1
9	84.4	34.7	8.3
11	82.5	31.8	7.3
12	79.2	30.4	6.7

Sample: adults aged 20-69, balanced panel with full data for waves 1 to 12 inclusive.

Table 5.2 shows that the percentage of respondents drinking any alcohol at all has fallen steadily in the USoc data – from 85 per cent in Wave 7 to just over 79 per cent in Wave 12. Meanwhile, the proportion of the sample with an AUDIT-C score indicating potential problem drinking has also fallen – from 35.5 per cent in Wave 7 to 30.4 per cent in Wave 12. The proportion of the sample with an AUDIT-C score of more than 10 (“heavy” drinkers) has fallen from 8.1 percent in Wave 7 to 6.7 per cent in Wave 12 (although there was a slight increase between Wave 7 and Wave 9).

5.3 Prevalence of obesity in Wave 12 of USoc and 2019 HSE

As discussed earlier, the obesity measure in Understanding Society is only available in Wave 12 so it is not possible to produce a table of results over time for obesity as for smoking prevalence or problem drinking. In Wave 12, 4.5% of the USoc adult sample aged between 20 and 69 reported a BMI of 40 or above (using cross-sectional wave 12 weights). In wave 12 of the balanced 12-wave panel used for the other descriptive analyses in this section, the proportion of adults with a BMI of 40 or above was also 4.5%.

Table 5.3 shows the proportion of the 2019 HSE sample in the ‘overweight’, ‘obese’ and ‘morbidly obese’ categories (the latter category corresponds exactly with the USoc obesity dummy variable).

Table 5.3. Percentage of respondents in each BMI category in Health Survey for England 2019

Category	% in category (weighted)
Underweight/normal (BMI<25)	35.8
Overweight (BMI 25-30)	36.2
Obese (BMI 30-40)	24.7
Morbidly obese (BMI 40+)	3.3

Sample: adults aged 20-69, cross-sectional sample

Table 5.3 shows that the proportion of adults in the 2019 HSE sample with a BMI of 40 or over is 3.3 per cent, which is lower than the equivalent proportion in Wave 12 of USoc. In total, 28 per cent of the sample have a BMI of 30 or above, and more than 64 per cent – almost two-thirds – have a BMI of 25 or above, and are therefore classified as overweight or obese.

5.4 Employment rates by smoker, drinker and obesity status and other characteristics, Understanding Society Wave 12

Smoker status

Table 5.4 shows the employment rate for adults aged 20-69 inclusive in the USoc Wave 12 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 almost 10 percentage points lower than for non-smokers. The employment gap is slightly larger for men than for women, and much larger for disabled adults than for non-disabled adults (there is a gap of almost 15 percentage points for disabled smokers compared to non-smokers, whereas the equivalent gap for non-disabled adults is only just under 2 percentage points). By age group, the biggest gaps in employment rate between smokers and non-smokers are for 40-49 year olds (18 percentage points) and 50-59 year olds (20 percentage points). In contrast, the employment rate for 20-29 year olds is almost identical for smokers and non-smokers at between 60 and 61 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). For most other

qualification groups the gap between smokers and non smokers is between 5 and 6 percentage points.

Table 5.4 Employment rates for current smokers and non-smokers in Understanding Society Wave 12

Sample characteristic	Employment rate		
	Non-smokers (%)	Current smokers (%)	Difference (% pts)
Overall sample	**62.3**	**52.4**	**9.9**
Gender:			
Male	65.6	55.3	10.3
Female	59.2	49.4	9.8
Disability:			
Disabled	49.4	34.6	14.8
Non-disabled	67.9	66.1	1.8
Age group:			
20-29	60.2	60.7	-0.5
30-39	72.6	62.7	9.9
40-49	76.5	58.4	18.1
50-59	71.9	51.9	20.0
60-64	46.9	38.7	8.2
65-69	19.3	11.2	8.1
Highest educational qualification:			
Degree	71.0	65.8	5.2
Other HE	68.7	58.5	10.2
A Level	67.0	61.6	5.4
GCSE	61.0	55.5	5.5
Other	50.3	46.4	3.9
None	42.2	36.5	5.7

Source: Landman Economics calculations using USoc Wave 12 data

Problem drinkers

Table 5.5 compares the employment rates for problem drinkers (as defined in Section 5.2 above) with the rest of the USoc Wave 12 sample. Across the board, the results show that employment rates for problem drinkers are *higher* than for the rest of the sample. Across the sample as a whole, problem drinkers have an employment rate almost 9 percentage points higher than the rest of the sample. This "employment premium" for problem drinkers is larger for women (just over 9 percentage points) than for men (just over 6 percentage points), and larger for disabled people than for non-disabled people. By age group, the gap between problem drinkers and the rest of the sample is largest for people aged 20-29 and 40-

49. By educational qualification the gap is largest for the “other” qualifications category (almost 28 percentage points) and GCSE (over 9 percentage points).

Table 5.5 Employment rates for problem drinkers and rest of sample in Understanding Society Wave 12

Sample characteristic	Employment rate		
	Not problem drinkers (%)	Problem drinkers (%)	Difference (% pts)
Overall sample	58.7	67.6	-8.9
Gender:			
Male	61.9	68.1	-6.2
Female	56.8	66.0	-9.2
Disability:			
Disabled	44.2	53.7	-9.5
Non-disabled	65.9	72.9	-7.0
Age group:			
20-29	58.9	71.9	-13.0
30-39	69.9	77.7	-7.8
40-49	70.2	82.2	-12.0
50-59	66.1	75.8	-9.7
60-64	45.0	47.4	-2.4
65-69	15.6	22.9	-7.3
Highest educational qualification:			
Degree	69.8	72.3	-2.5
Other HE	66.3	70.0	-3.7
A Level	63.6	70.0	-6.4
GCSE	58.8	68.1	-9.3
Other	46.1	73.8	-27.7
None	38.6	45.4	-6.8

Source: as Table 5.4

Heavy drinkers

To analyse the relationship between high levels of alcohol consumption and employment in more detail, Table 5.6 compares employment rates for *heavy* drinkers (those with AUDIT-C score of 11 or more) and the rest of the sample. The results are much more mixed than for Table 5.5, implying that when the relatively small group of heavy drinkers is isolated, their employment rates are much closer to the rest of the sample than for the problem drinkers. While heavy drinkers are still slightly more likely to be employed than the rest of the sample (1.6 percentage points), for men, disabled people, people aged 30-39 and 50-59 and people whose highest qualification is “other HE”, A levels or other qualifications, heavy drinkers are *less* likely to be in employment than the rest of the sample.

Table 5.6. Employment rates for heavy drinkers (AUDIT-C score >10) and rest of sample, Understanding Society Wave 12

Sample characteristic	Employment rate		
	Not heavy drinkers (%)	Heavy drinkers (%)	Difference (% pts)
Overall sample	60.7	62.3	-1.6
Gender:			
Male	64.2	63.7	0.5
Female	57.7	59.2	-1.5
Disability:			
Disabled	46.6	45.7	0.9
Non-disabled	67.5	70.7	-3.2
Age group:			
20-29	60.1	65.1	-5.0
30-39	71.9	64.3	7.6
40-49	73.9	76.2	-2.3
50-59	69.0	66.4	2.6
60-64	45.4	52.9	-7.5
65-69	18.3	22.9	-4.6
Highest educational qualification:			
Degree	70.8	70.8	0.0
Other HE	68.5	61.3	7.2
A Level	66.4	65.1	1.3
GCSE	59.7	63.4	-3.7
Other	50.2	14.5	35.7
None	40.5	42.7	-2.2

Source: as Table 5.4

People with BMI more than 40

Table 5.7 compares employment rates for people with BMI equal to or more than 40 in the USoc Wave 12 data, and the rest of the sample. The results show that across the sample as a whole, people with BMI of 40 or above are substantially less likely to be employed than the rest of the sample (a gap of almost 12 percentage points). The "obesity gap" is largest for men (20 percentage points), people aged 21-29 (23 percentage points), people aged 40-49 (almost 14 percentage points), people aged 60-64 (almost 17 percentage points), people whose highest qualification is GCSE level (14 percentage points) and people with no qualifications (15 percentage points). The only group for whom a BMI of 40 or above is associated with *better* outcomes than the rest of the sample is people whose highest qualification is "other" but this is a fairly small proportion of the sample (about 2 per cent).

Table 5.7. Employment rates for very overweight people (BMI >=40) and rest of sample, Understanding Society Wave 12

Sample characteristic	Employment rate		
	BMI < 40 (%)	BMI >=40 (%)	Difference (% pts)
Overall sample	61.8	50.2	11.6
Gender:			
Male	65.5	45.4	20.1
Female	58.5	52.8	5.7
Disability:			
Disabled	47.0	43.3	3.7
Non-disabled	68.2	64.5	3.7
Age group:			
20-29	62.9	39.9	23.0
30-39	72.2	67.1	5.1
40-49	74.7	61.1	13.6
50-59	69.6	61.5	8.1
60-64	46.4	29.8	16.6
65-69	18.6	8.7	9.9
Highest educational qualification:			
Degree	70.8	70.9	-0.1
Other HE	67.7	59.5	8.2
A Level	65.7	62.8	2.9
GCSE	62.5	48.3	14.2
Other	50.9	66.5	-15.6
None	41.1	26.0	15.1

Source: as Table 5.4

5.4 Employment rates by smoker, drinker and obesity status and other characteristics, Understanding Society Wave 12

Smoker status

Table 5.8 shows median monthly earnings for employed adults aged 20-69 inclusive in the USoc Wave data, using the same set of breakdowns as for Table 5.5 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 19 percent higher than for smokers. The earnings gap is larger for men than for women (23 percent compared to 18 percent); it is larger for disabled than non-disabled people (21 percent compared to 17 percent). By age group, the biggest earning gaps are for 30-39 year olds (24.5 per cent) and 40-49 year olds (25 per cent) whereas the gap is smallest for workers aged 65 or over. By highest qualification, there is a gap of more than 23

percent between median earnings for non-smokers and smokers for degree holders, more than 17 per cent for people whose highest qualification is A levels, and 19 per cent for people with no qualfiications, but almost no difference in median earnings between smokers and non-smokers for workers whose highest qualification is GCSE.

Table 5.8 Median earnings for current smokers and non-smokers in Understanding Society Wave 12

Sample characteristic	Median monthly earnings		
	Non-smokers (£)	Current smokers (£)	Difference (%)
Overall sample	2,073	1,677	19.1%
Gender:			
Male	2,500	1,920	23.2%
Female	1,732	1,419	18.1%
Disability:			
Disabled	1,907	1,504	21.1%
Non-disabled	2,106	1,746	17.1%
Age group:			
20-29	1,691	1,500	11.3%
30-39	2,300	1,737	24.5%
40-49	2,473	1,858	24.9%
50-59	2,187	1,700	22.3%
60-64	1,733	1,646	5.0%
65-69	1,250	1,240	0.8%
Highest educational qualification:			
Degree	3,265	2,500	23.4%
Other HE	2,925	2,979	-1.8%
A Level	2,083	1,725	17.2%
GCSE	1,693	1,676	1.0%
Other	1,646	1,800	-9.4%
None	1,542	1,390	9.9%

Source: as Table 5.4

Heavy drinkers

Table 5.9 uses the same heavy drinker definition as in Table 5.7 to look at the difference in median earnings for heavy drinkers and the rest of the sample in USoc Wave 12. Across the sample as a whole, median earnings for heavy drinkers are around 10 percent higher than for the rest of the sample. The “earnings premium” for heavy drinkers is larger for women than for men, and larger for non-disabled than disabled people. The largest gaps in earnings by age are for people aged 60 to 64 (a 21 per cent advantage for heavy drinkers compared to the rest of the sample) and people aged 50-59 (a difference of 14 per cent). By highest educational qualification

the “earnings premium” for heavy drinkers is largest for other HE, other qualifications and no qualifications.

Table 5.9. Median monthly earnings for heavy drinkers and rest of sample in Understanding Society Wave 12

Sample characteristic	Median monthly earnings		
	Rest of sample (£)	Heavy drinkers (£)	Difference (£)
Overall sample	2,000	2,195	-9.7%
Gender:			
Male	2,380	2,400	-0.8%
Female	1,680	1,750	-4.2%
Disability:			
Disabled	1,827	1,950	-6.7%
Non-disabled	2,057	2,253	-9.5%
Age group:			
21-29	1,666	1,604	3.7%
30-39	2,180	2,200	-0.9%
40-49	2,380	2,655	-11.6%
50-59	2,099	2,396	-14.1%
60-64	1,700	2,058	-21.1%
65-69	1,240	1,127	9.1%
Highest educational qualification:			
Degree	3,224	3,500	-8.6%
Other HE	2,900	3,750	-29.3%
A Level	2,042	2,200	-7.7%
GCSE	1,680	1,815	-8.0%
Other	1,647	1,907	-15.8%
None	1,515	1,666	-10.0%

Source: as Table 5.4

People with BMI of 40 or over

Finally in this chapter, Table 5.10 shows median monthly earnings for people with BMI of 40 or above in USoc Wave 12 compared to the rest of the sample. The results here are mixed. Overall, people with BMI of 40 or above have slightly lower median earnings than the rest of the sample (a 2 percent difference). The groups for whom median earnings are lowest for very overweight people in USoc compared to the rest of the sample are people aged 30-39 (a difference of almost 23 per cent), people aged 65-69 (a difference of almost 42 per cent), people whose highest qualification is “other” (24 per cent) and people with no qualifications (16.5 per cent).

For women, disabled people, people aged 60-64, and people whose highest qualification is “other HE” or A levels, median earnings are higher for people with BMI over 40 than for the rest of the population.

Table 5.10. Median monthly earnings for very overweight people (BMI >=40) and rest of sample, Understanding Society Wave 12

Sample characteristic	Median monthly earnings		
	BMI < 40 (%)	BMI >=40 (%)	Difference (% pts)
Overall sample	2,000	1,959	2.1%
Gender:			
Male	2,383	2,383	0.0%
Female	1,666	1,700	-2.0%
Disability:			
Disabled	1,850	1,957	-5.8%
Non-disabled	2,064	2,000	3.1%
Age group:			
21-29	1,666	1,625	2.5%
30-39	2,200	1,703	22.6%
40-49	2,400	2,400	0.0%
50-59	2,058	1,916	6.9%
60-64	1,690	2,300	-36.1%
65-69	1,455	850	41.6%
Highest educational qualification:			
Degree	3,240	2,667	17.7%
Other HE	2,916	3,166	-8.6%
A Level	2,023	2,300	-13.7%
GCSE	1,720	1,625	5.5%
Other	1,647	1,252	24.0%
None	1,500	1,252	16.5%

[blurb discussion here]

6 Methodology: smoking regressions

This section explains the methodology used in the employment and earnings regressions for smoking. Subsequent sections explain the methodology for the heavy drinking regressions (Chapter 7) and the obesity regressions (Chapter 8).

6.1 Employment regression: main specification

The main employment model specification for smoking uses a logistic regression for individuals in Wave 12 of the USoc Survey who were also in Waves 1 through 11 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 12 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 12, 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) later in this section under the subheading "smoking variables used in the regressions".
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 5.2 below.

The employment regression sample consists of all adults in USoc Wave 8 aged 20 to 69 inclusive. Although the state pension age is currently 66, it is useful to include men and women aged between 67 and 69 in the sample as the employment rate for this group is significantly greater than zero[4] and it is possible that smoking may have a particular impact on retirement ages which should be taken into account in this research.

6.2 Earnings regression: main specification

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 12 and for whom there also exists a complete set of interview data for all 12 Waves. The dependent variable is the log of earnings in the

[4] For example, in USoc Wave 12 the employment rate for adults aged 66 to 69 (inclusive) is around x per cent.

most recent month before interview. Using a log earnings measure means that the 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 12 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 12. This variable is set to 1 for individuals who are unemployed and inactive in two or more of Waves 9, 10 and 11, and zero for individuals who are unemployed or inactive in just one of those waves, or employed in all three waves.

6.3 Smoking variables used in the regressions

The USoc data contains smoking prevalence data across the whole sample for Wave 2 and Waves 5 to 12 of the survey. Estimating regressions using employment or earnings in Wave 8 as the dependent variable, eight different lags of smoking (smoking status in Waves 11, 10, 9, 8, 7, 6, 5 and 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 11 (a one year lagged variable) and smoking in Wave 2 (a ten-year lagged variable). The Wave 11 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.

6.4 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 11, 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 11 only, plus other controls.

Variant 3: Smoking lags in Wave 11 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 12 instead of previous waves is used.

Variant 5: as variant 4 but using whole Wave 12 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 12 who were asked about their current smoking status[5], including those individuals with incomplete data for USoc Waves 1 to 11.

Variant 6: a random effects panel specification for Waves 5 to 12 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 7 (all individual observations in Waves 6 to 12 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 10 waves in this specification as we only have this information for wave 12 (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 12 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 7 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.

[5] 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 8: pre-Covid estimates

The interview period for Waves 11 and 12 of the Understanding Society data overlapped with the Covid-19 pandemic which began in March 2020. As discussed in Institute for Social and Economic Research (2022b), the pandemic affected the Understanding Society data in at least three ways which are relevant for this report:

i) During the initial period of lockdown in March-July 2020, the interview method for USoc, which was previously a mix of in-person and online interviews, shifted to online only.
ii) Employment rates were lower during 2020 than in 2019. The reduced employment rate because of increased unemployment during the early stages of the pandemic because some workers were laid off or could not find work. Workers covered by the furlough scheme for employees or the Self-Employed Income Scheme (SEISS) were still recorded as employees and self-employed respectively rather than being unemployed or inactive, and so the furlough and SEISS should not have affected recorded employment rates.
iii) Earnings were lower during 2020 than 2019. This is partly because of the furlough and SEISS schemes (which only covered 80% of earnings up to a maximum of £30,000 per year) and partly because of the steep decline in economic activity during lockdown, which affected the earnings and self-employed income of people whose incomes were sensitive to economic conditions in the short run (e.g. self-employed people outside SEISS and people on zero hours contracts).

To check whether the Covid pandemic affected the estimated results from the Wave 12 analysis, we have estimated versions of the main specification which use employment and earnings in Wave 10 (the final pre-Covid wave, with interviewing finishing in December 2019) as the dependent variables and regress on smoking variables in Waves 9 and 2, plus control variables as recorded in Wave 10.

6.5 Control variables used in the regressions

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 (used in the specifications for smoking and drinking, but not for the main obesity specification – see Section 5.x below).
- Ethnicity;
- Highest educational qualification;
- Pregnancy;
- Caring for a disabled adult in the household;
- Region of residence;
- Housing tenure.

7 Methodology: Drinking regressions

7.1 Employment regression: main specification

As with the smoking regressions, the main employment model specification for drinking (alcohol consumption) uses a logistic regression. Whereas the smoking regression used data for Waves 1 to 12, the main employment regression for alcohol consumption only uses data for waves 7 and 12 of USoc so the sample size is larger in this case. The binary dependent variable for employment at Wave 12 of USoc is regressed against three sets of explanatory variables:

1) **Drinking behaviour (measured by AUDIT-C score) at Waves 12 and 7 of USoc.** Unlike with the smoking variable, the results from the drinking regressions seem to be fairly sensitive to whether the variables measuring drinking behaviour are from the same wave as the dependent variable (Wave 12) or the previous wave (Wave 11). The measured effects (at least in the main logistic specification) seem to be much weaker when drinking status in the previous wave is used. This might be due to reverse causality (i.e. people might be more likely to drink heavily because they are not working) although the descriptive statistics in Section 3.5 (which show much higher rates of problem drinking for people in employment than non-working people) do not support this hypothesis. On the other hand it may be that problem drinking status is more volatile from wave to wave than smoker status. However, as with the smoking regression (discussed in Section 6.3), we also include alcohol consumption at an earlier wave in the USoc panel to control for longer-term drinking behaviour. The earliest USoc wave at which AUDIT-C questions were asked is Wave 7 so we include AUDIT variables from that wave. More details of the drinking variables are given in Section 7.3 below.
2) **Control variables** – exactly the same as listed in Section 6.5 above for the smoking regressions.

Because the alcohol consumption regressions do not use any data from waves before Wave 7, we use the data from Waves 7 and 12 only in the employment regression (to increase the available sample size).

7.2 Earnings regression: main specification

As with the employment regression, the earnings regression for examining the relationship between drinking behaviour and employment is an Ordinary Least Squares (OLS) specification using the subsample of individuals who were in employment (either employees or self-employed) during Wave 12 and for whom there also exists a complete set of interview data for Waves 7 to 12. The dependent variable is the log of earnings in the most recent month before interview.

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 (as with the earnings regression for smoking behaviour discussed in Section 6.3).

7.3 Alcohol consumption variables used in the regressions

To measure drinking behaviour in Wave 12 and Wave 7, the AUDIT-C score is entered into the regression using three dummy variables for each wave:

i) Score 1 to 5;
ii) Score 6 to 10;
iii) Score 11 to 15.

The base category is people who answered no to the question “have you had an alcoholic drink in the past 12 months? “ i.e. non-drinkers. This comprises just over 20 percent of the Wave 12 sample (as shown in Table 5.2 above). Of the three AUDIT-C categories above, category (iii) corresponds to heavy drinkers as used for Table 5.2 above. Thus, the representation of drinking behaviour in the employment (and earnings) regressions is more sophisticated than for either the smoking regressions in Section 6.3 above or the obesity regressions in Section 8.3 below, where simple dummy variables are used. This is because (as is shown in the results in Chapter 10 below) the relationship between drinking and labour market outcomes is more complex than for smoking or obesity. In particular, individuals in categories (i) and (ii) above (AUDIT-C scores 1 to 10) have *higher* employment and earnings than the non-drinking base category (controlling for other factors). Thus, it is necessary to separate out heavy drinkers (with AUDIT-C scores of 11 or more) to derive a negative relationship between employment or earnings, and drinking behaviour.

7.4 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 the specifications are as follows:

Main specification: Dummy variables for AUDIT-C score in Wave 12 and AUDIT-C score in Wave 7. For each wave, 3 dummies are used (1 to 5, 6 to 10 and 11 to 15) making 6 dummies in total. The specification also includes the control variables listed in Section 6.5.

Variant 1: As for the main specification but omitting the Wave 7 variables for AUDIT-C score, to focus on the coefficients on the Wave 12 AUDIT-C variables in isolation.

Variant 2: As for variant 1 but using the whole Wave 12 sample rather than just the sample with data for Waves 7 through 12 inclusive.

Variant 3: AUDIT-C score dummies for Wave 12 and Wave 7 only, with no controls. (This is to show the “raw” effects of the AUDIT-C variables).

Variant 4: additional variables for self-reported health and life satisfaction

As for the 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 12 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.

As with variant 7 of the smoking regressions, 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 5: pre-Covid estimates – Wave 9

As with Variant 8 of the smoking employment regressions, to check whether the Covid pandemic affected the estimated results from the Wave 12 analysis in the main specification, we have estimated versions of the main specification which use employment and earnings in an earlier wave. Whereas the smoking regressions used Wave 10, we have used Wave 9 because Wave 10 does not include the AUDIT-C questions.

Omission of random effects specification for drinking regressions

Note that this report does not include a random effects specification for the drinking regressions because the USoc data does not have AUDIT-C score measures for all 6 waves from 7 to 12 but only for 11, 11, 9 and 7. This means that the only two consecutive waves which can be used in a random effects spec are 11 and 12 so we do not get the sample size benefits that we had with the random effects specification for the smoking regressions.

8 Methodology: Obesity regressions

8.1 Employment regression: main specification

As with the smoking and drinking regressions, the main employment model specification for obesity uses a logistic regression. The sample is the cross-sectional data for individuals in Wave 12 of the USoc Survey (because the obesity variable in USoc is only measured at Wave 12). The binary dependent variable for employment at Wave 12 of USoc is regressed against three sets of explanatory variables:

3) **Obesity at Wave 12 of USoc (measured using the dummy variable for BMI equal to or more than 40).** Because obesity is only measured in Wave 12 of the USoc survey – and only with one dummy variable – we have no choice but to use this dummy variable as the obesity measure in the USoc obesity regressions.
4) **Control variables** – exactly the same as listed in Section 6.5 above for the smoking regressions, except that the disability dummy variable is omitted from the main specification due to collinearity with the obesity variable (this is discussed in more detail in the assessment of obesity regression results in Chapter 11).

The employment regression for obesity uses data from Wave 12 only.

8.2 Earnings regression: main specification

As with the employment regression, the earnings regression for examining the relationship between drinking behaviour and employment is an Ordinary Least Squares (OLS) specification using the subsample of individuals who were in employment (either employees or self-employed) during Wave 12 and for whom there also exists a complete set of interview data for Waves 9 to 12. The earnings regression uses data from Waves 9, 10 and 11 to construct the variable for being not in employment for at least two of the previous three waves.

The dependent variable for this regression is the log of earnings in the most recent month before interview.

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 the labour market history binary variable.

8.3 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 the specifications are as follows:

Main specification: Dummy variable for obesity in Wave 12 (measured as BMI equal to or greater than 40) plus explanatory variables.

Variant 1: Obesity dummy variable only (to show the “raw” relationship between obesity and employment or earnings).

Variant 2: As for variant 1 but including the disability dummy variable in the set of controls.

Variant 3: additional variables for self-reported health and life satisfaction
As explained in variant 7 of the smoking regressions.

Variant 4: Health and Survey England regression (for employment only)

This variant uses the HSE data instead of USoc because of the more detailed obesity variable available in the HSE. The specification includes dummy variables for:

- BMI 25-30 (overweight);
- BMI 30-40 (obese);
- BMI 40+ (morbidly obese – equivalent to the obesity dummy in Usoc).

The HSE regression also includes a full set of control variables as follows:

- Gender
- Age dummy variables
- Number of children in household (1, 2, 3 or more)
- Highest qualification
- Carer for disabled adult in household
- Ethnicity
- Currently pregnant
- Housing tenure
- Region.

Omission of pre-Covid and Random Effects specifications

It is important to note that a variant with estimates from the pre-Covid period is not possible for the obesity analysis because the Usoc dummy variable is only recorded at Wave 12. Also, the obesity regressions also do not contain a random effects specification because obesity is only measured in Wave 12 of the USoc data, and the HSE is a cross-sectional survey so cannot be used to run panel regressions.

9 Regression Results - Smoking

9.1 Employment regressions

Coefficients on smoking variables from main specification and variants

Table 10.1 shows the coefficients on the smoking variable in the employment regressions for the main specification as well as the variants explained in Section 6.4 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 11 (or in Wave 2, or ever having smoked, or Wave 11, depending on the precise specification being estimated).

In Table 10.1 (and in subsequent results tables in this report), 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 11 of USoc are just under 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 just under per cent less likely to be in employment in Wave 8 than never-smokers. The coefficient for smoking in Wave 11 is statistically significant at the 5% level but the coefficient for smoking in Wave 2 is not. The coefficient for smoking before Wave 2 is also 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 Wave 11 of the USoc survey and the probability of being in employment in Wave 12.

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 11 is markedly larger here, at -0.1013, or around minus 10 per cent.

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 -0.0563 – suggesting that smokers in Wave 7 are just under 6 per cent less likely to be in employment than non-smokers at Wave 7. Variant 3 shows that if

the ‘ever smoked’ variable is omitted, the coefficients on the Wave 11 smoking variable is slightly larger (at -0.563) than for the main specification, but the difference is not particularly great.

Variant 4 shows that when contemporaneous smoking status in Wave 12 is used in the regression (instead of lagged smoking status in Wave 11), the coefficient on smoking status shows a slightly smaller negative impact of smoking: minus 4.5 per cent compared to minus 5.6 per cent in 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 10,252 to 22,074 individuals. The coefficient on smoking in Wave 11 shows slightly smaller effects with the larger sample (-0.0406 compared to -0.0450) 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 6-12 and regressors from waves 5-11. This increases the sample size to 166,920 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 -0.0329, implying that smokers in each Wave of USoc are just over 3 per cent less likely to be employed than non-smokers in the next wave of the survey. This is a smaller result than for Variant 2, which is the closest analogous specification to the Wave 12 logistic regression model.

In Variant 7, which includes additional explanatory variables for self-reported health status and life satisfaction, the employment penalty from smoking in Wave 11 is slightly smaller than in the main specification (a coefficient of -0.0432, meaning that Wave 11 smokers are just under 4 per cent less likely to be in employment, controlling for other factors), but not by much.

Finally, in Variant 8, which uses employment data from wave 10 and smoking data from Wave 9, the coefficient on smoking in Wave 9 is relatively small (at -2.2 per cent) and is not statistically significant. However, unlike in other specifications the variables for smoking at Wave 2 and “ever smoked” *are* significant (and negative). Adding the coefficients for Wave 2 and ever smoked, individuals who smoked at wave 2 and before Wave 2 are almost 7 percentage points less likely to be in employment than people who did not smoke in those periods, controlling for other factors.

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 9.1 Smoking and employment regressions: Main regression and variants

	Main specification		Variant 1		Variant 2		Variant 3		Variant 4		Variant 5		Variant 6		Variant 7		Variant 8	
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]	m.e.	[z]
Wave 12*	-		-		-		-		-.0450	3.08	-.0406	3.52	-					
Wave 11**	-.0473	3.09	-.1013	4.80	-.0563	3.94	-.0555	3.03	-		-		-.0329	8.94	-.0432	2.90	-.0222	1.48
Wave 2	-.0186	1.14	-.0298	1.62	-		-.0128	0.83	-		-		-		-.0182	1.13	-.0469	3.39
ever	.0119	1.10	-.0154	1.27	-		-		-		-		.0026	0.71	.0128	1.18	-.0217	2.33
Number of obs	10,252		10,252		10,252		10,252		10,252		22,074		166,920		10,252		13,114	
Pseudo R^2	0.2001		0.0066		0.1772		0.2000		0.1725		0.1632		N/A		0.2037		0.2287	
Chi-sq test on combined smoking vars with P-value	Chi-2 (3) = 20.90 (P = 0.001)		Chi-2 (3) = 66.61 (P = 0.000)		n/a		Chi-2 (2) = 19.64 (P = 0.000)		n/a		n/a		Chi-2 (2) = 81.65 (P = 0.0000)		Chi-2 (2) = 18.72 (P = 0.0003)		Chi-2 (2) = 29.38 (P = 0.0000)	
Wald test for significance of regression with P-value	Chi-2 (49) = 1538.13 (P = 0.0000)		Chi-2 (3) = 66.61 (P = 0.0000		Chi-2 (48) = 1612.81 (P = 0.0000)		Chi-2 (48) = 1539.92 (P = 0.0000)		Chi-2 (48) = 1596.81 (P = 0.0000)		Chi-2 (48) = 2176.16 (P = 0.0000)		Chi-2 (49) = 13765.10 (P = 0.0000)		Chi-2 (51) = 1571.67 (P = 0.0000)		Chi-2 (51) = 2034.63 (P = 0.0000)	

*Except Variant 8 (Wave 9)

**Except variant 6 (Wave t-1) and variant 8 (Wave 9)

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

m.e. – marginal effect

9.2 Earnings regressions

Coefficients on smoking variables from main specification and variants

Table 9.2 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 -0.0905 on log weekly earnings at Wave 12 for smoking at Wave 11. This implies an earnings penalty for smokers of just over 9 per cent. The coefficient on smoking status at Wave 2 is also negative but is a lot smaller in absolute terms (-0.0186) and is not statistically significant. The coefficient on the "ever smoked" variable is *positive* but 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 11 smoking and earnings (an earnings penalty of almost 24 per cent). The Wave 2 smoking and "ever smoked" variables are statistically insignificant.

Variant 2 shows that when the Wave 11 smoking variable is included, but no other smoking variables, the earnings penalty for smokers is around 9 per cent, while Variant 3 shows a slightly larger earnings penalty for Wave 11 smokers of around 9.8 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 12 smoking measure is used instead of the Wave 11 smoking measure, the earnings penalty to smoking is slightly larger than Variant 2, at 9.8 per cent. Variant 5 shows that when the full Wave 12 sample is included rather than the balanced 12-wave sample (expanding the number of individual observations from 6,124 to 13,456) the coefficient on smoking is slightly larger (-0.1076 rather than -0.0977). 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 107,489. The estimate for the earnings penalty from smoking at Wave t-1 is just over 5 per cent which is smaller than the Wave 11 estimate from the main specification.

Variant 7 shows that including additional variables for self-reported health status and life satisfaction in the earnings regression actually makes the coefficient on smoking status in Wave 11 slightly larger, implying an earnings penalty of 9.55 per cent for smokers. Finally, Variant 8, which uses earnings information in Wave 10 and smoking information in Wave 9, results in a coefficient on Wave 9 smoking which is similar to the random effects specification in Variant 6 but not statistically significant.

The F-test of the significance of the whole regression is significant in all eight regression specifications in Table 9.2.

Table 9.2 Smoking and earnings regressions: Main regression and variants

	Main specification		Variant 1		Variant 2		Variant 3		Variant 4		Variant 5		Variant 6		Variant 7		Variant 8	
Smoking in wave:	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]	ceoff	[t]
Wave 12*	-		-		-		-		-.0977	3.02	-.1076	3.67	-					
Wave 11**	-.0905	2.33	-.2390	5.04	-.0902	2.94	-.0979	2.27	-		-		-.0517	3.61	-.0955	2.24	-.0590	1.61
Wave 2	-.1006	0.49	-.0442	1.07	-		-.0067	0.20	-		-		-		-.0224	0.61	-.0508	1.47
ever	.0231	0.91	.0033	0.12	-		-		-		-		-.0324	1.00	..0269	0.61	0.0013	0.06
Number of obs	6,124		6,124		6,124		6,124		6,124		13,456		107,489		6,124		8.268	
R^2	0.2227		0.0394		0.1772		0.2225		0.1725		0.1919		N/A		0.2248		0.2137	
F-test on combined smoking vars with P-value	F (3, 6073) = 3.12 (P = 0.0249)		F (4, 6073) = 32.72 (P = 0.0000)		n/a		F (2, 6074) = 4.32 (P = 0.0133)		n/a		n/a		Chi-2 (2) =59.10 (P = 0.0014)		F (3, 6071) = 3.03 (P = 0.0282)		F (3, 8217) = 4.93 (P = 0.0020)	
F-test*** for significance of regression with P-value	F (50, 6073) = 25.92 (P = 0.0000)		F (3, 6073) = 18.11 (P = 0.0000)		F (49, 6075) = 30.32 (P = 0.0000)		F (49,6074) = 26.42 (P = 0.0000)		F (49, 6075) = 31.02 (P = 0.0000)		F (49, 13406) = 41.21 (P = 0.0000)		Chi-2 (49) = 8423.34 (P = 0.0000)		F (52, 6017) = 25.54 (P = 0.0000)		F (50, 8217) = 37.69 (P = 0.0000)	

*Except Variant 8 (Wave 9)
**Except variant 6 (Wave t-1) and variant 8 (Wave 9)
***Except variant 6 (Chi-squared test)
Note: grey shading indicates statistically significant result at the 5% level.
m.e. – marginal effect

10 Regression Results – drinking

10.1 Employment regressions

Table 10.1 shows the results from the alcohol consumption regressions for employment, using the regression specifications as set out in Section 7.1 (for the main specification) and Section 7.4 (for the variants). The top panel of Table 10.1 shows the results from the main specification and variants 1 and 2, while the lower panel shows the results from variants 3 through 5.

The results from the main specification show that the marginal effects on the AUDIT-C scores follow a "downward-sloping" relationship. An AUDIT-C score of 1 to 5 is associated with the probability of employment being just over 4 percentage points higher than the base category of people who had not had anything to drink at all in the 12 months prior to the Wave 12 interview (controlling for other factors). For an AUDIT-C score of 6 to 10, employment probability is just over 2 percentage points higher. An AUDIT-C score of 11 to 15 (which we define as "heavy drinking") is associated with an employment probability almost 4 percentage points *lower*. In other words, drinkers with AUDIT-C scores between 1 and 10 are more likely to be in employment (conditional on other factors) than non-drinkers, but heavy drinkers (i.e. those with AUDIT-C scores of 11 or above) are less likely to be employment than non-drinkers (or drinkers with AUDIT-C scores of 10 or below).

The lagged AUDIT-C score variables from Wave 7are all positive with marginal effects of around 6 per cent (for scores of 1 to 5 and 11 to 15) or 8 per cent (for scores of 6 to 10). [any more discussion here?]

The chi-squared test of the combined AUDIT-C score variables in the main specification (and variant 1, where the Wave 7 lagged variables are omitted) shows that they are jointly significant. However, the AUDIT-C score dummy for heavy drinkers in Wave 12 is not individually significant in either specification. This shows the importance of including the full set of AUDIT-C variables for Wave 12 (and Wave 7 where appropriate) in the regressions rather than just a single "heavy drinking" dummy.

In Variant 1 (where lagged AUDIT-C scores are omitted), the marginal effect on the heavy drinker dummy (AUDIT-C score of 11 to 15) is smaller, at just under 2 per cent. The marginal effect of heavy drinking is also around 2 per cent in Variant 2 (where the full Wave 12 sample is used rather than just the sample with data for Wave 7 and Wave 12. In Variant 3 (where no control variables are included other than the AUDIT-C scores in Waves 7 and 12), the heavy drinking dummy for Wave

12 also has a marginal effect of around minus 2 per cent. In Variant 4 (where life satisfaction variables are included),

Table 10.1 Drinking and employment regressions: Main regression and variants

	Main specification		Variant 1		Variant 2	
AUDIT scores:	m.e.	[z]	m.e.	[z]	m.e.	[z]
Wave 12:						
Score 1-5	0.0410	3.38	0.0517	4.39	0.0535	4.81
Score 6-10	0.0218	2.18	0.0428	4.49	0.0517	5.82
Score 11-15	-0.0393	1.06	-0.0178	0.54	-0.0211	0.71
Wave 7:						
Score 1-5	0.0618	4.46				
Score 6-10	0.0818	6.39				
Score 11-15	0.0609	2.14				
Number of obs		17,935		17,935		22,118
Pseudo R^2		0.1902		0.1871		0.1647
Chi-sq test on combined AUDIT vars with P-value		Chi-sq (6) = 66.91 P=0.0000		Chi-sq(3) = 31.10 P=0.0000		Chi-sq(3) = 45.87 P=0.0000
Wald test for significance of regression with P-value		Chi-sq (53) = 2003.48 P=0.0000		Chi-sq (50) = 1992.05 P=0.0000		Chi-sq (50) = 2148.17 P=0.0000

	Variant 3		Variant 4		Variant 5	
AUDIT scores:	m.e.	[z]	m.e.	[z]	m.e.	[z]
Wave 12:						
Score 1-5	0.0565	4.09	0.0409	3.40	0.0552	4.81
Score 6-10	0.0493	4.47	0.0214	2.15	0.0837	5.82
Score 11-15	-0.0170	0.48	-0.0341	0.98	-0.0141	0.71
Wave 7:						
Score 1-5	0.0954	4.46	0.0619	4.48	0.0451	3.72
Score 6-10	0.1508	6.39	0.0812	6.34	0.0433	3.48
Score 11-15	0.0911	2.14	0.0644	2.26	0.0041	0.17
Number of obs		17,935		17,935		23,362
Pseudo R^2		0.0166		0.1936		0.2222
Chi-sq test on combined AUDIT vars with P-value		Chi-sq (6) = 225.59 P=0.0000		Chi-sq(6) = 66.02 P=0.0000		Chi-sq(6) = 160.45 P=0.0000
Wald test for significance of regression with P-value		Chi-sq (6) = 225.59 P=0.0000		Chi-sq (55) = 2032.83 P=0.0000		Chi-sq (53) = 3005.60 P=0.0000

*Except Variant 5 (Wave 10)
Note: grey shading indicates statistically significant result at the 5% level.
m.e. – marginal effect

10.2 Earnings regressions

Table 10.2 shows the results from the alcohol consumption regressions for earnings, using the regression specifications as set out in Section 7.2 (for the main specification) and Section 7.4 (for the variants). As with Table 10.1, the top panel of Table 10.2 shows the results from the main specification and variants 1 and 2, while the lower panel shows the results from variants 3 through 5.

The results from the main specification show that, as with the employment regression, there is a "downward-sloping" relationship between AUDIT-C score and coefficients. Drinkers with an AUDIT-C score of between 1 and 5 have *higher* earnings than non-drinkers in Wave 12, controlling for other factors (the earnings premium for this group is almost 6 per cent). There is a smaller earnings premium for drinkers with an AUDIT-C score of between 6 and 10 (just over 1.5 per cent). Heavy drinkers (those with AUDIT-C scores of 11 or above) experience an earnings *penalty* of just over 10 per cent compared to non-drinkers.

The lagged AUDIT-C scores for Wave 7 in the main specification are all positive, with earnings premiums of over 15 per cent for those with scores of 6 to 10 and 11 to 15 in Wave 7. [why?] The AUDIT-C score variables in the main specification are jointly significant at the 5% level (as shown by the F-test statistics). This is also the case in Variant 1 where the Wave 7 AUDIT-C dummies are omitted. However, the heavy drinker variable (AUDIT-C score of 11 to 15 in Wave 12) is not individually significant in either specification. This is a similar finding to the employment regressions in Section 10.1 above.

In variant 2 (where the sample is expanded to include anyone with full cross-sectional data in Wave 12 plus labour market history in Waves 9-11, rather than just those respondents who also have data for Wave 7), the coefficient on the heavy drinker variable is somewhat lower, at around minus 4 per cent. The AUDIT-C variables for Wave 12 remain jointly significant in this specification.

In variant 3 (which features the Wave 12 and Wave 7 AUDIT-C scores and no other controls), the coefficient on heavy drinking in Wave 12 is larger (an earnings penalty of around 11.5 per cent compared to non-drinkers). In Variant 4 (which includes life satisfaction indicators) the earnings penalty for heavy drinking in Wave 12 is larger still (around 12 per cent). Finally, in Variant 5 (which uses pre-COVID earnings data from wave 9) *all* the AUDIT-C indicators for Wave 9 are positive and significant. [This suggests that the negative association between heavy drinking and earnings *appears* after COVID]

Table 10.1 Drinking and earnings regressions: Main regression and variants

	Main specification		Variant 1		Variant 2	
AUDIT scores:	Coeff	[t]	Coeff	[t]	Coeff	[t]
Wave 12:						
Score 1-5	0.0597	2.24	0.0564	2.15	0.0497	2.08
Score 6-10	0.0163	0.70	0.0568	2.53	0.0507	2.50
Score 11-15	-0.1044	1.28	-0.0797	0.71	-0.0413	0.42
Wave 7:						
Score 1-5	0.0566	1.52				
Score 6-10	0.1573	4.43				
Score 11-15	0.1624	2.08				
Number of obs	11,010		11,010		13,463	
R^2	0.1695		0.1655		0.1647	
F test on combined AUDIT vars with P-value	F (6, 10956) = 6.41 P=0.0000		F (3, 10959) = 3.10 P=0.0257		F (3, 13412) = 2.81 P=0.0000	
F test for significance of regression with P-value	F(53, 10956) = 30.61 P=0.0000		F(50, 10959) = 31.25 P=0.0000		F(50, 13412) = 38.78 P=0.0000	

	Variant 3		Variant 4		Variant 5	
AUDIT scores:	Coeff	[t]	Coeff	[t]	Coeff	[t]
Wave 12*:						
Score 1-5	0.0737	2.58	0.0521	1.99	0.0666	2.02
Score 6-10	0.0466	1.88	0.0171	0.74	0.1258	3.64
Score 11-15	-0.1154	1.09	-0.1199	1.19	0.1371	2.39
Wave 7:						
Score 1-5	0.0476	1.32	0.0508	1.38	0.0257	0.81
Score 6-10	0.2106	6.31	0.1477	4.20	0.1175	3.78
Score 11-15	0.1783	2.24	0.1516	1.95	0.0251	0.37
Number of obs	11,010		11,010		15,253	
R^2	0.0397		0.1887		0.2184	
Chi-sq test on combined AUDIT vars with P-value	F (6, 11002) = 14.49 P=0.0000		F (6, 10953) = 5.980 P=0.0000		F (3, 13412) = 15.18 P=0.0000	
Wald test for significance of regression with P-value	F(7,11002) = 28.13 P=0.0000		F(56, 10953) = 31.98 P=0.0000		F(54, 15198) = 65.31 P=0.0000	

*Except Variant 5 (Wave 10)

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

m.e. – marginal effect

11 Regression Results – obesity

11.1 Employment regressions

Table 11.1 shows the results from the regressions for the relationship between obesity and employment regressions for employment, using the regression specifications as set out in Section 8.1 (for the main specification) and Section 8.3 (for the variants). The top panel of Table 10.1 shows the results from the main specification and variants 1 and 2, while the lower panel shows the results from variants 3 and 4.

The results from the main specification show a negative association between the USoc “very overweight” variable (BMI of 40 or above) and employment. The marginal effect is -0.0473, meaning that people with a BMI of 40 or above are just under 5 percentage points less likely to be in employment than the rest of the sample, controlling for other factors. The obesity variable is statistically significant at the 5% level. In Variant 1 (which includes just the obesity variable with no other controls) the marginal effect is just over minus 10 per cent. This suggests that including other control variables reduces the estimated negative association between obesity and employment. This is confirmed by the results from Variant 2, which shows that when a disability dummy variable is included in the set of controls, the estimated marginal effect of obesity falls to less than 1 per cent and it is no longer statistically significant. Given the substantial overlap between the subsample in USoc who are disabled and the subsample of people with BMI of 40 or above, disentangling the effects of obesity on employment from the wider effect of disability on employment is a challenging task.

Variant 3 adds life satisfaction indicators to the set of controls (but omits disability). The results show a slightly lower association between obesity and employment compared to the main specification (a marginal effect of around minus 3.5 per cent) and the obesity variable is no longer significant at the 5% level.

Finally, variant 3 uses HSE data instead of USoc. The sample size of HSE is smaller (just over 6,000 observations compared with 11,000 for USoc) but we are able to include dummy variables for overweight (BMI between 25 and 30) and obese (BMI between 30 and 40) as well as the variable for BMI of 40 or over (morbidly obese). The results show a *positive* association between BMI 25-30 and employment and BMI 30-40 and employment. While the coefficient on the variable for BMI of 40 or over is negative, the result is not significant at the 5% level.

Table 11.1 Obesity and employment regressions: Main regression and variants

	Main specification		Variant 1		Variant 2	
Dataset	Usoc		Usoc		Usoc	
Obesity variables	m.e.	[z]	m.e.	[z]	m.e.	[z]
Wave 12:						
BMI 40 or above	-0.0481	2.27	-0.1003	4.03	-0.0097	0.44
Number of obs	18,466		18,466		18,466	
Pseudo R^2	0.1480		0.0011		0.1615	
Wald test for significance of regression with P-value	Chi-sq (47) = 1697.03 P=0.0000		Chi-sq (1) = 16.13 P=0.0001		Chi-sq (48) = 1786.32 P=0.0000	

	Variant 3		Variant 4 (HSE)	
Dataset	USoc		HSE	
Obesity variables	m.e.	[z]	m.e.	[z]
BMI 25-30	n/a		0.0317	2.53
BM 30-40	n/a		0.0439	3.13
BMI 40 or above	-0.0353	1.63	-0.0151	-0.51
Number of obs	18,466		6,867	
Pseudo R^2	0.1480		0.2122	
Wald test for obesity variables	n/a		Chi-sq (3) = 13.29 P=0.0000	
Wald test for significance of regression with P-value	Chi-sq (47) = 1701.29 P=0.0000		Chi-sq (38) = 1143.22 P=0.0000	

11.2 Earnings regressions

A full set of earnings regressions was estimated for the Understanding Society Wave 12 data using the specifications set out in Sections 8.2 and 8.3 above. Unfortunately, the coefficient on the variable for BMI of 40 or above was not statistically significant in *any* of the regression specifications. This means that we are unable to establish any reliable relationship between obesity and earnings, whether controlling for other factors or not. Accordingly, we have not included any estimated impact of obesity on earnings in the aggregate productivity calculations in Chapter 12 below.

As explained in Chapter 4, it was not possible to estimate earnings regressions using the HSE data because HSE does not contain any information on individual earnings.

12 Estimating the overall productivity losses from smoking, drinking and obesity

This section of the paper uses the results from the regression specifications estimated in Chapters 9, 10 and 11, 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, drinking and obesity. The calculations to derive the productivity losses are shown in Table 12.1 (for smoking) Table 12.2 (for heavy drinking) and Table 12.3 (for obesity).

The overall estimated productivity impacts for smoking, heavy drinking and obesity sum to £31.2 billion in total – just under 1.4% of UK Gross Domestic Product in 2022.

12.1 Methodology

Smoking and heavy drinking

The calculations for the productivity impacts of smoking and heavy drinking in this report use two sets of results from the relevant regressions in Chapters 9 (for smoking) and 10 (for drinking). One set is the estimated impact of smoking on employment (from Table 9.1) and the estimated impact of heavy drinking on employment (from Table 10.1). The others are the estimated impact of smoking on earnings (from Table 9.2) and the impact of heavy drinking on earnings (from Table 10.2). The smoking calculations here use the coefficient on smoking in the previous wave of USoc (Wave 11) from the main specification of the employment and earnings regressions. The heavy drinking calculations use the coefficient on the dummy for AUDIT-C score of 11 to 15 in Wave 12 from the main specification of the employment and earnings regressions.

The calculation for smoking proceeds in four stages as follows. First, the marginal effect of smoking on employment (calculated from the employment regression as -0.0473) 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 (around £26,500 per year). Third, the coefficient from the earnings regression in the main specification of Table 9.2 (a 9.05 percent earnings penalty for smokers) 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 under £2,400 per year)[6]. Finally, the overall productivity loss to the UK economy from smoking is calculated as the sum of two components:

[6] 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.

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.

The calculation for heavy drinking proceeds similarly, using the marginal effect of heavy drinking on employment (-0.0393), average annual earnings for current heavy drinkers in the USoc data (around £32,500) and the coefficient from the earnings regression in the main specification of table 10.2 (a 10.44% earnings penalty for heavy drinkers). This is used to estimate what the average earnings of heavy drinkers in the UK economy would be if they were not heavy drinkers (an increase of just under £3,400 per year). Finally the overall productivity loss to the UK economy from heavy drinking is calculated as the sum of two components:

a) The increase in overall earnings for current heavy drinkers already in employment if they were not heavy drinkers;
b) The additional earnings for heavy drinkers who are not currently in employment but would be if they had were not heavy drinkers.

Obesity

The calculation of productivity impacts for obesity (defined as a BMI of 40 or above) is more straightforward than for smoking or heavy drinking because the analysis in this report was not able to find any relationship between obesity and earnings (either “raw” or controlling for other explanatory variables). Therefore, the productivity loss calculation for obesity only includes the equivalent of item (b) in the lists for drinking and smoking above (i.e. the additional earnings for people with BMI of 40 or above who are not currently in employment but would be if they had BMI of below 40. This calculation uses the marginal effect of a BMI of 40 or above on employment from the USoc regression main specification (-0.0481) and average earnings across the UK economy (around £33,700 per year).

12.2 Results and discussion

Smoking

Table 12.1 shows the detailed calculations involved in the smoking estimates. Component (a) is estimated at £8.3bn and component (b) at £9.8bn, leading to an overall estimate for the productivity losses arising from smoking in the UK economy of £18.1bn.

Table 12.1. 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 20-69 in UK	42.56	ONS (2022a)
2: Total smoking prevalence in the age group 20-69	14.37%	Usoc Wave 12
3: marginal effect of smoking on employment	-0.0478	Marginal effect calculation based on smoking (t-1) variable in Table 9.1 main specification
4: Amount of additional employment for smokers aged 20-69 if they did not smoke	289,000	(1) x (2) x (3)
Wage levels for current smokers if UK smoking prevalence were zero		
5: Average annual earnings across UK economy	£33,696	ONS (2022b)
6: Average earnings for smokers as percentage of average earnings across whole economy	78.50%	Usoc Wave 12
7: Average annual earnings for smokers in UK economy	£26,451	(5) x (6)
8: earnings premium for non-smokers compared to smokers controlling for other factors	9.05%	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	£2,394	(7) x (8)
10: total earnings for current smokers if they had never smoked	£28,845	(7) + (9)
11: total smoking prevalence among people in work	12.37%	USoc Wave 12
12: total UK employment	33.09	ONS (2022c)
13: total number of smokers in employment	4.09m	(11) x (12)
Increase in UK productivity arising from:		
14: Increase in earnings for current smokers already in work	£9.8bn	(9) x (13)
15: Earnings of people who are not currently in work but would be if smoking prevalence were zero	£8.3bn	(4) x (10)
Total loss in UK productivity from smoking	**£18.1bn**	**(14) + (15)**

Heavy drinking

Table 12.2 shows the detailed calculations involved in the heavy drinking estimates. Component (a) is estimated at £3.6bn and component (b) at £7.1bn, leading to an overall estimate for the productivity losses arising from heavy drinking in the UK economy of £10.6bn. This is smaller than the estimate of losses due to smoking. Mainly the smaller result for heavy drinking arises because the incidence of heavy drinking in the USoc sample (at just under 6% for the overall sample aged 20-69 and just over 6% for the sample in work) is smaller than the incidence of smoking in the USoc sample (which is just over 14% across all 20-69 year olds and just over 12% for those in work).

Table 12.2. Calculations of overall productivity gains to the UK economy from immediate cessation of heavy drinking in the population

Additional employment if UK prevalence of heavy drinking (AUDIT-C score 11-15) were zero		
Statistic	**Value**	**Source**
1: Total population aged 20-69 in UK	42.56	ONS (2022a)
2: Total heavy drinking prevalence in the age group 20-69	5.93%	Usoc Wave 12
3: marginal effect of heavy drinking on employment	-0.0393	Marginal effect calculation based on AUDIT-C score 11-15 variable in Table 10.1 main specification
4: Amount of additional employment for heavy drinkers aged 20-69 if they were not heavy drinkers	99,000	(1) x (2) x (3)
Wage levels for current smokers if UK heavy drinking prevalence were zero		
5: Average annual earnings across UK economy	£33,696	ONS (2022b)
6: Average earnings for heavy drinkers as percentage of average earnings across whole economy	96.39%	Usoc Wave 12
7: Average annual earnings for heavy drinkers in UK economy	£32,479	(5) x (6)
8: earnings premium for people who are not heavy drinkers compared to heavy drinkers controlling for other factors	10.44%	Inverse of coefficient on smoking (t-1) variable in Table 4.3 main specification
9: increase in earnings for current heavy drinkers if they were not heavy drinkers	£3,391	(7) x (8)
10: total earnings for current heavy drinkers if they were not heavy drinkers	£35,870	(7) + (9)
11: total heavy drinking prevalence among people in work	6.29%	USoc Wave 12
12: total UK employment	33.09	ONS (2022c)
13: total number of heavy drinkers in employment	2.08	(11) x (12)
Increase in UK productivity arising from:		
14: Increase in earnings for current heavy drinkers already in work	£7.1bn	(9) x (13)
15: Earnings of people who are not currently in work but would be if heavy drinking prevalence were zero	£3.6bn	(4) x (10)
Total loss in UK productivity	**£10.6bn**	**(14) + (15)**

Obesity

Finally, Table 12.3 shows the calculations involved in the estimates of the productivity costs of obesity. There is only one component here – the costs due to the negative employment impact of obesity (defined as BMI equal to or greater than 40 or "morbid" obesity), which is estimated at £2.4 billion. Thus, the productivity costs of obesity are smaller than for smoking or heavy drinking. Partly this is because the proportion of people with BMI of 40 or above in the USoc sample aged 20 to 69 is relatively low (at 3.5%). The other reason is that there is no statistically reliable estimate of the impact of obesity on earnings for those in work and so the calculation for obesity is based on the negative employment impact only.

Table 12.3. Calculations of overall productivity gains to the UK economy if the incidence of morbid obesity (BMI of 40 or more) were zero

Additional employment if UK prevalence of BMI scores of 40 or above were zero		
Statistic	**Value**	**Source**
1: Total population aged 20-69 in UK	42.56	ONS (2022a)
2: Total prevalence of BMI >=40 in the age group 20-69	3.50%	Usoc Wave 12
3: marginal effect of BMI>=40 on employment	-0.0481	Marginal effect calculation based on (BMI >=40) variable in Table 11.1 main specification
4: Amount of additional employment for people with BMI>=40 aged 20-69 if they had BMI below 40	72,000	(1) x (2) x (3)
5: Average annual earnings across UK economy	£33,696	ONS (2022b)
Increase in UK productivity arising from:		
6: Earnings of people who are not currently in work but would be prevalence of BMI>=40 were zero	£2.4bn	(4) x (5)
Total loss in UK productivity	**£2.4bn**	**(6)**

6 Conclusions

The results from the analysis in this research report suggest that smoking and heavy drinking (defined as having an AUDIT-C score of 11 or above in Understanding Society) are both negatively associated with the probability of being employed in the UK, and that there is also penalties associated with both smoking and heavy drinking. For obesity (defined as having a BMI score of 40 or above in the USoc data) there is a negative association with employment, but not earnings.

The results from the preferred specifications of the employment regressions in this report indicate that for adults aged between 20 and 69 in Wave 12 of the Understanding Society survey, smoking in Wave 11 was associated with being just under 5 per cent less likely to be in employment. Heavy drinking in Wave 12 was associated with being just under 4 per cent less likely to be in employment, and a BMI of 40 or above was associated with being just under 5 per cent less likely to be in employment. All these regression results control for other factors.

The results from the preferred specifications of the earnings regressions suggest that for adults in USoc, smoking in Wave 11 was associated with an earnings penalty of just over 9 per cent, while heavy drinking in Wave 12 was associated with an earnings penalty of just over 10 per cent. Both of these analyses control for other factors.

The estimated associations from the employment and earnings regressions imply sizeable productivity losses from smoking and drinking in particualr. 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 £18 billion. A similar calculation for heavy drinkers suggests that the total cost of heavy drinking to the UK economy is around £10.6 billion. There is also an estimated negative productivity impact of obesity but it is smaller, at around £2.4 billion. In total, the estimated productivity impact of smoking, heavy drinking and morbid obesity summed together is estimated at £31.2 billion – around 1.4 per cent of UK GDP in 2022.

Overall, it has been harder to identify reliable relationships between alcohol consumption and labour market outcomes, and obesity and labour market outcomes, using the USoc data (or the HSE data in the case of obesity) than it was for smoking. This is partly because the estimated relationship between drinking and employment and earnings, and between obesity and employment and earnings, is more complex than for smoking. A simple binary variable identifying smokers in the data and differentiating them from non-smokers is sufficient to identify sizeable and statistically significant negative associations between smoking and labour market outcomes. By contrast, low to moderate amounts of alcohol consumption are associated with *positive* labour market outcomes, and it is only for heavy drinkers – a relatively small proportion of the overall population of drinkers – that negative effects can be identified. Similarly, negative associations between obesity and employment are only identifiable for relatively extreme levels of BMI (40 and above) – partly due to the data limitations of Understanding Society, but also in the HSE, which has data on people who are overweight or obese but with BMI below 40. Further research,

particularly (if possible) with panel datasets which feature a longer time series of AUDIT score data for drinking and BMI data for obesity, would be most useful in this regard.

References

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.

Cabinet Office (2003), *Alcohol misuse: How much does it cost?* Cabinet Office Strategy Unit report. https://www.ias.org.uk/uploads/pdf/Economic%20impacts%20docs/costi%20uk.pdf

De Sio S, Tittarelli R, Di Martino G, Buomprisco G, Perri R, Bruno G, Pantano F, Mannocchi G and Marinelli E (2020), "Alcohol consumption and employment: a cross-sectional study of office workers and unemployed people". *PeerJ Hubs*. https://peerj.com/articles/8774/

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

Department of Health, Social Services and Public Safety [Northern Ireland] (2010), *Social Costs of Alcohol Misuse in Northern Ireland for 2008/09*. https://www.health-ni.gov.uk/sites/default/files/publications/dhssps/scam-2008-09.pdf

Frontier Economics (2022), *Estimating the Full Costs of Obesity: A report for Novo Nordisk*. https://www.frontier-economics.com/media/hgwd4e4a/the-full-cost-of-obesity-in-the-uk.pdf

Heien, D (1996), "The relationship between alcohol consumption and earnings". *Journal of Studies on Alcohol and Drugs*, 57(5), pp 536-542. https://doi.org/10.15288/jsa.1996.57.536

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

Institute for Social and Economic Research (2022), "Understanding Society: Survey response rates". November 2022. https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/responsetables.pdf

Institute for Social and Economic Research (2022b), "Understanding Society main study changes due to the Covid-19 pandemic (Wave 12 release)". November 2022. https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/understanding_society_changes_during_covid.pdf

Institute of Alcohol Studies (2020), “The costs of alcohol to society”. Alcohol Knowledge Centre briefing. https://www.ias.org.uk/wp-content/uploads/2020/12/The-costs-of-alcohol-to-society.pdf

Jawad A and Reed E (2023), *Holding Us Back: Tobacco, alcohol and unhealthy food and drink.* Obesity Health Alliance, Alcohol Health Alliance and ASH.

Kim J T and von dem Knesebeck O (2018), “Income and obesity: what is the direction of the relationship? A systematic review and meta-analysis”. BMJ Open 2018;8: e019862. doi:10.1136/ bmjopen-2017-019862. https://bmjopen.bmj.com/content/bmjopen/8/1/e019862.full.pdf

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.

Mangot-Sala L, Smidt N and Liefbroer A (2022), “Disentangling the association between alcohol consumption and employment status: causation, selection or confounding?” *European Journal of Public Health*, 32(6), pp 926-932. https://doi.org/10.1093/eurpub/ckac141

Monsivais P, Martin A, Suhrcke M, Forouhi N and Wareham N (2015), “Job-loss and weight gain in British adults: Evidence from two longitudinal studies”, *Social Science and Medicine*, Vol 143, pp 223-231. https://doi.org/10.1016/j.socscimed.2015.08.052

National Institute on Drug Abuse Clinical Trials Network [NIDA CTN] (2023), “Instrument: AUDIT-C Questionnaire”. https://cde.nida.nih.gov/instrument/f229c68a-67ce-9a58-e040-bb89ad432be4

Office for National Statistics [ONS] (2022a), “Population estimates for the UK, Wales, Scotland and Northern Ireland mid 2021”: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021d

Office for National Statistics [ONS] (2022b), “Average weekly earnings in Great Britain: June 2023”: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/averageweeklyearningsingreatbritain/june2023

Office for National Statistics [ONS] (2022b), “Employment in the UK: June 2023”: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/employmentintheuk/june2023

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

Public Health England (2017), “Health matters: Obesity and the food environment”, 31 March 2017. https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2

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

Reed, H (2021), “The impact of smoking history on employment prospects, earnings and productivity: an analysis using UK panel data”. London, ASH. https://ash.org.uk/resources/view/smoking-employability-and-earnings

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

Thørissen M, Bonsaksen T, Hashemi N, Kjeken I, van Mechelen W and Aas R (2019), “Association between alcohol consumption and impaired work performance (presenteeism): a systematic review. *BMJ Open,* Volume 9 Issue 7. https://bmjopen.bmj.com/content/9/7/e029184

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.

World Health Organization (2001), “AUDIT: the Alcohol Use Disorders Identification Test: guidelines for use in primary health care”. https://www.who.int/publications/i/item/WHO-MSD-MSB-01.6a

York Health Economics Consortium [University of York] (2010), *The societal cost of alcohol misuse in Scotland for 2007.* Edinburgh: Scottish Government Social Research. https://drugslibrary.wordpress.stir.ac.uk/files/2017/03/SGalocohol_cost_to_society.pdf