The impact of labor conditions on subjective health

Mark Treurniet
Erasmus School of Economics

July 13, 2012

Abstract

This paper tries to identify the causal impact of labor conditions on health in two different ways. First of all, we will try to isolate the effect of general labor conditions by using the distinction between blue collar and white collar employment. Secondly, we examine the specific effect of smoke-free workplaces by exploiting the introduction of the smoking ban in the United Kingdom as a natural experiment. We use data from the British Household Panel Survey (BHPS), which allows us to control for previous health, fixed effects and pre-determined characteristics, like gender, age and education. We find no significant impact of labor conditions in general and smoke-free workplaces in specific on health.

Keywords: health, labor conditions, smoking bans, BHPS

1. Introduction

Labor conditions may affect health in various ways. Income and fringe benefits may make a healthy lifestyle affordable and may help to finance proper health care, whereas exposure to dangerous situations and physical load may cause injuries and stress may cause mental disorders. Furthermore, firms are social institutions where personal relations may play an important role. Social norms may encourage employees, for example, to sport or smoke together or to attend social drinks after work (Fletcher and Sindelar, 2009).

Occupation and health are shown to be associated (e.g. Marmot and Smith, 1997; Marmot et al, 1997). However, selection into occupation may explain the association between labor conditions and health: if higher educated individuals choose for white collar employment and education directly improves health, then white collar workers are ceteris paribus more healthy, but this can be explained by education. Furthermore, if a relatively good ex-ante health encourage individuals to choose to work under bad labor conditions and health is serially correlated, then previous health can explain part of the association between labor conditions and future health.

If labor conditions have a causal effect on health, the government can choose to target specific policies to improve health. Furthermore, the government may want to incorporate this effect in its considerations on retirement age and pension schemes.

However, little research is done on the causal relationship between labor conditions and health: Fletcher and Sindelar (2009) use American data from the Panel Study of Income Dynamics (PSID) in order to identify the relationship between early occupational choices and later health. The authors find evidence that blue collar occupation at first labor market entry is associated with lower health at later age and suggest that there exists a causal relationship. Fletcher and Sindelar treat the relationship between first occupation and later health as “black box”, meaning that they do not examine the mechanisms underlying the causal relationship.

In this paper we try to identify whether a causal relationship between labor conditions and contemporaneous self-reported health exists. We use British data from the British Household Panel Survey (BHPS). This dataset contains variables on a wide range of topics, including health, employment, socio-demographics and education. The dataset allows us to examine the relation between occupation and health and control for lagged health, cross-section fixed effects, gender, age and education.
We estimate both a lagged dependent variables model and a fixed effects model in order to bound the effect of general labor conditions on health. Finally, we estimate a model which accounts for both fixed effects and lagged dependent variables using an instrumental variable approach.

In order to identify the causal effect of labor conditions on health, researchers have to make critical assumptions about selection into occupation in order to be able to properly control for the possible bias. We extend our research by opening the black box and examine the effect of one specific labor condition: smoking on the workplace.
Fichtenberg and Glantz (2002) review 26 studies on the effects of smoke-free workplaces on cigarette consumption. The authors argue that smoke-free workplaces both protect non-smokers from the risks of passive smoking and encourage smokers to stop smoking or reduce cigarette consumption. Fichtenberg and Glantz conclude that total cigarette consumption reduces by 29% as a result of quitting and reducing consumption.
Jones et al (2011) study the impact of the introduction of smoking bans in England and Scotland on the prevalence of smoking and the consumption of cigarettes. They do not find an effect on overall smoking prevalence, but do find some evidence of the impact on cigarette consumption for certain groups.

In this paper, we exploit the introduction of smoke-free workplaces in the United Kingdom as a natural experiment in order to identify the specific impact of smoking on the job. The Smoking Health and Social Care (Scotland) Act 2005 prohibits smoking on workplaces and other enclosed public places in Scotland as of March 26, 2006. Wales and Northern Ireland followed by introducing a smoking ban on April 2, 2007 and April 30, 2007 respectively. England required workplaces and other enclosed public places to be smoke-free as of July 1, 2007 (Jones et al, 2011).
First, we will examine the impact of the introduction of these smoking bans on the prevalence of smoking and the consumption of cigarettes. Then, we investigate its impact on health.
Since the introduction of the smoking ban in the United Kingdom is exogenous to the individuals’ relative health as compared to other individuals, we are not bothered by self-selection into the

treatment. Therefore, we estimate a fixed effects model in order to identify the impact of the smoking ban.

We find some evidence on the effect of general labor conditions in our lagged dependent variables model. However, we do not find a significant impact of general labor conditions on health in our models that account for cross-section fixed effects. The results of our model that combines lagged dependent variables and cross-section fixed effects indicate that our fixed effects model is most appropriate to identify the causal effect of general labor conditions on health.
We do not find that the introduction of smoke-free workplaces significantly impacts the prevalence of smoking, the consumption of cigarettes or health.

The remainder of this paper is devoted to discuss our *Data*, *Methodology* and *Results*. Finally, we present our *Conclusions*. Each of these sections first discusses the analysis on the effect of labor conditions in general. The second part of these sections are dedicated to our research on the specific impact of smoke-free workplaces.

2. Data

We use data from the British Household Panel Survey (BHPS), which were supplied from the UK Data Archive. We use the first 18 waves (1991-2008), which contain 582,840 observations on 32,380 individuals.
The multi-purpose study follows all individuals from a selected sample of households during a number of years. The panel started with about 5,500 households and 10,300 individuals in wave 1. In wave 9, 1,500 households from Scotland and Wales were added to the sample. In wave 11, 2,000 households from Northern-Ireland were added (BHPS, 2012a). The interviews for the BHPS start on September 1 of each year (BHPS, 2012b).

For waves 1-8 and 10-19, the variable health status is based on the question: "Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been: excellent, good, fair, poor or very poor?" This question has not been asked in wave 9. However, another question on general health status is asked in wave 9 and not in other waves. Therefore, we use this question to measure health status in wave 9: "In general would you say your health is: excellent, very good, good, fair or poor?" We stress that besides the question also the response categories were different in wave 9. We handle this in two different ways, depending on the analysis. The answers are transformed in a numerical value in such a way that better health status is associated with a lower numerical value.

We establish a dummy occupation that equals 1 for white collar employment and 0 for blue collar employment. Information on occupation is drawn from the variable socio-economic class of the current job. We consider professionals, managers, supervisors, sales and services to be white collar and production, technical, operative and agricultural to be blue collar (Fletcher and Sindelar, 2009).

Information on gender is provided by the interviewer. We define the dummy gender to be 1 for males and 0 for females.

The variable age is defined as the age at the date of the interview. This variable is calculated from the birth date of the respondent and measured in full years.

Information on the highest educational qualification is retrieved from questions on obtained qualifications in the first wave and questions on obtained qualifications in the last year in each subsequent wave. Each year, the variable is updated in order to include the most recent qualifications. We establish dummies for each category. For a detailed description of the categories, we refer to *Appendix A – Educational qualification categories*.

The dummies employed, retired and student are based on the question: "Please look at this card and tell me which best describes your current situation: self-employed, in paid employment (full or part-time), unemployed, retired from paid work altogether, on maternity leave, looking after family or home, full-time student / at school, long term sick or disabled, on a government training schema or something else?" The dummy employed equals 1 if the respondent is in paid employment and 0 otherwise. The dummy retired equals 1 if the respondent is retired from paid work and 0 otherwise. The dummy student equals 1 if the respondent is a full-time student and 0 otherwise. Since these dummies will be used to restrict the sample, we do not need to create dummies for the other categories.

For wave 10-19, the variable smoking status is based on the question: "Do you smoke cigarettes?" Since this question has not been asked in wave 9 and since we only use this variable for the impact evaluation of the smoking ban, we do not measure the smoking status for earlier waves. The answers are transformed into a dummy, where 1 indicates that the individual is a smoker and 0 means that the individual is a non-smoker.

We measure the daily cigarette consumption by the answer on the question: "Approximately how many cigarettes a day do you usually smoke, including those you roll yourself?" This question has only been asked to smokers. For self-reported non-smokers, we set the daily cigarette consumption to zero.

We establish a dummy which indicates whether the individual lives in Scotland or other parts of the United Kingdom. Information for this dummy is collected from a variable, which is derived from confidential data of the BHPS.

2.1. Occupation

2.1.1. Sample

Since we are interested in the impact of general labor conditions and since we want to exclude potential distorting effects of intervening activities like entrepreneurship, unemployment, maternity, family care or disabilities, we restrict our sample to all individuals that are employed in all available waves after schooling and before retirement.

As mentioned before, for wave 9, the variable health status is based on a question with different response categories. In order to make the answers comparable to the observations in the other waves, we collapse the five category scale to a four category scale with categories excellent, good or very good, fair and poor or very poor (Jones et al, 2011). We prefer to transform the scale over dropping all observations for wave 9, since the latter solution will cause a large loss of data in our lagged dependent variables model and especially in our model that accounts for both fixed effects and lagged dependent variables. Furthermore, for waves 1-8 and 10-18, the two least frequently chosen response categories are combined.

2.1.1. Descriptive statistics

Descriptive statistics of the relevant variables in our research on the impact of general labor conditions on health are reported in table 1. 71.4% of the individuals in this sample is employed in white collar jobs versus 28.6% in blue collar jobs. Furthermore, about half of our respondents is male. The average age equals 39.3 years.

Table 1 - Descriptive statistics occupation

Variable	*Ind*	*Obs*	*Mean*	*St Dev*	*Min*	*Max*
Health status (1=excellent, 5=poor or very poor)	12490	73877	1.918	0.760	1	4
- Excellent health	12490	73877	0.298	0.458	0	1
- Good or very good health	12490	73877	0.520	0.500	0	1
- Fair health	12490	73877	0.147	0.354	0	1
- Poor or very poor health	12490	73877	0.035	0.183	0	1
Occupation	12255	73148	0.714	0.452	0	1
Gender	12495	74027	0.507	0.500	0	1
Age	12511	74048	39.349	12.458	15	95
Highest educational qualification	11889	69238	-	-	-	-
- Higher degree	11889	69238	0.036	0.187	0	1
- First degree	11889	69238	0.144	0.351	0	1
- Teaching qualification	11889	69238	0.025	0.156	0	1
- Other higher qualification	11889	69238	0.263	0.440	0	1
- Nursing qualification	11889	69238	0.015	0.123	0	1
- GCE A levels	11889	69238	0.130	0.336	0	1
- GCE O levels	11889	69238	0.201	0.401	0	1
- Commercial qualification	11889	69238	0.020	0.140	0	1
- GCSE grades	11889	69238	0.033	0.178	0	1
- Apprenticeship	11889	69238	0.010	0.099	0	1
- Other qualification	11889	69238	0.005	0.073	0	1
- No qualification	11889	69238	0.116	0.320	0	1
- Still at school	11889	69238	0.002	0.040	0	1

2.2. Smoking ban

2.2.1. Sample

We use a different sample for our impact evaluation of the smoking ban.

Observations from individuals that are not included in the panel both before and after the introduction of the smoking ban cannot help estimating the effect of the smoking ban. However, these observations may only distort the estimation process by affecting the common trend. Therefore, we restrict our sample to individuals from Scotland who have observations in wave 15 and 16 and individuals from other parts of the United Kingdom who have observations in wave 16 and 17[1].

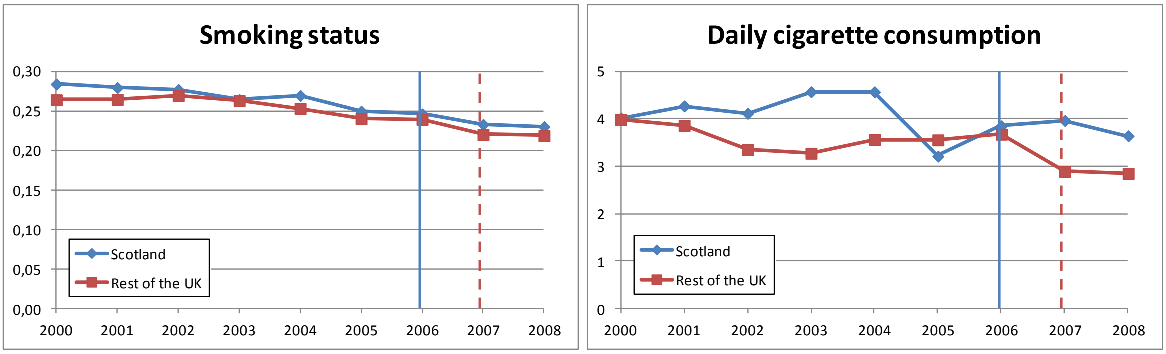
We are interesting in the effect of working conditions on health. Therefore, we further restrict our sample to all individuals that are employed in the period before the introduction of the ban and not yet retired in the period after the introduction of the ban[2].

As mentioned before, for wave 9, the variables smoking status and health status could not be based on an identical question. Since individuals may react in a different way on a deviating question, we exclude all observations up until wave 9 from our sample.

2.2.2. Descriptive statistics

The trends of the prevalence of smoking and the consumption of cigarettes for both Scotland and the rest of the United Kingdom are shown in graph 1. The first wave after the introduction of the bans are indicated by the solid vertical lines for Scotland and the dashed lines for the other parts of the United Kingdom. Both smoking prevalence and cigarette consumption seem to have a downward sloping trend. Especially the trends of daily cigarette consumption seems to differ between Scotland and the remainder of the United Kingdom. In the next section, we will explain how we will test for the similarity of the trends of the different countries after controlling for differences between individuals.

Graph 1 – Smoking status and daily cigarette consumption



Descriptive statistics of the relevant variables in our analysis of the impact of the smoking ban are reported in table 2. 18.4% of the individuals live in Scotland versus 81.6% in the other parts of the United Kingdom. Due to a different sample selection, the descriptive statistics are different from the descriptive statistics in the previous subsection. However, the distribution over

[1] More precisely, individuals in wave 16 that live in Scotland should have an observation in wave 15 and wave 16 in order to be included in the sample, whereas individuals in wave 16 that live in other parts of the United Kingdom should have an observation in wave 16 and wave 17.

[2] More precisely, individuals in wave 16 that live in Scotland should be employed in wave 15 and not yet be retired in wave 16 in order to be included in the sample, whereas individuals in wave 16 that live in other parts of the United Kingdom should be employed in wave 16 and not yet be retired in wave 17.

different health status categories does not show large differences as compared to the previous subsection.

Table 2 - Descriptive statistics smoking ban

Variable	*Ind*	*Obs*	*Mean*	*St Dev*	*Min*	*Max*
Health status (1=excellent, 5=very poor)	6758	56077	1.974	0.809	1	5
- Excellent health	6758	56077	0.285	0.451	0	1
- Good health	6758	56077	0.506	0.500	0	1
- Fair health	6758	56077	0.165	0.371	0	1
- Poor health	6758	56077	0.038	0.191	0	1
- Very poor health	6758	56077	0.006	0.077	0	1
Smoking status	6716	54897	0.248	0.432	0	1
Daily cigarette consumption	6716	54763	3.528	7.413	0	80
Scotland	6758	55556	0.184	0.388	0	1

3. Methodology

Ferrer-i-Carbonnel and Frijters (2004) compare the implications of assumptions regarding the interpretation of self-reported satisfaction and the existence and effects of unobserved factors on the estimates of the determinants of self-reported satisfaction. The research focuses on the determinants of happiness, but the authors argue that their reasoning also applies to other satisfactions, like satisfaction with health.

Firstly, Ferrer-i-Carbonnel and Frijters compare cardinal and ordinal comparability assumptions on the interpretation of the health status.
The cardinal comparability assumption means that the difference in health between a health status of 1 and 2 is the same as between a health status of 3 and 4 and is equal for each individual[3]. In this case, estimation is often done using OLS.
When the health status is ordinally comparable, one assumes that the difference in health between different health status is unknown, but equal for each individual[4]. This assumption can be modeled by latent variable models, like ordered probit and logit.

Ferrer-i-Carbonnel and Frijters conclude that assuming cardinal or ordinal comparability is relatively unimportant to the results. OLS and ordered probit do not yield qualitatively different results. In our analysis we use OLS and check the robustness of our results by estimating an ordered probit model.
However, properly controlling for time-invariant unobserved factors proves to be essential. Ferrer-i-Carbonnel and Frijters show that models with cross-section fixed-effects and models that do not account for fixed unobserved factors yield qualitative different results.

[3] More formally, the difference in the underlying concept health is a linear function of the difference in health status: $health_i - health_j = \omega(health\_status_i - health\_status_j)$.
[4] More formally, if $health\_status_i > health\_status_j$, then for the underlying concept health holds that $health_i > health_j$.

We estimate both a lagged dependent variables model and a fixed effects model. We interpret the resulting estimates as bounds of the true effect (Angrist and Pischke, 2008).

Note that since the number of individuals N in our dataset is much larger than the number of periods T, stationarity is no issue in our analysis.

3.1. Occupation

3.1.1. Lagged dependent variables

We estimate a lagged dependent variables model in order to investigate the impact of general labor conditions (Angrist and Pischke, 2008, chapter 5). This model allows us to control for the reverse causality between occupation and health: ex-ante health may influence the occupational choice. The model is based on the identifying assumption that after controlling for the lagged health status and the included covariates, there is no selection bias. This means that no other factor affects both occupation and health:

$$\begin{aligned} &E[health\_status_{0it}|health\_status_{it-1}, \lambda_t, X_{it}, occupation_{it},] \\ &\quad = E[health\_status_{0it}|health\_status_{it-1}, \lambda_t, X_{it}] \end{aligned} \tag{1}$$

where $health\_status_{0it}$ is the health status for individual i in wave t provided that the individual would be employed in a blue collar job, $health\_status_{it-1}$ is the lagged health status, the year effects λ_t capture the differences in self-reported health over time, the observed covariates X_{it} include exogenous characteristics that influence both the selection into the treatment and self-reported health and the treatment dummy $occupation_{it}$ indicates whether the individual is employed in a blue collar or white collar job.

We estimate the following linear equation:

$$health\_status_{it} = \alpha + \theta health\_status_{it-1} + \lambda_t + \rho occupation_{it} + X_{it}\delta + \varepsilon_{it} \tag{2}$$

where $health\_status_{it}$ is the observed self-reported health status for individual i in wave t.

Besides the lagged health status, we control for year effects, gender, age and highest educational qualification. Since we want to allow for a more rapid health deterioration at later age, we also control for squared age.

Occupation may affect the income, health care insurance and behavior of employees. These variables may act as mechanisms by which occupation influences health: white collar jobs may, for example, offer higher salaries. If the higher income allows individuals to afford healthier products, this may improve health. We do not control for these time-variant individual-specific variables in the regression, since we attribute their effect on health as an indirect effect of general labor conditions on health and we are interested in the full impact of occupation on health.

3.1.2. Fixed effects

We further investigate on the relation between occupation and health by estimating a fixed effects model (Angrist and Pischke, 2008, chapter 5). This model allows us to control for all unobserved variables that affect both the occupational choice and health. Fixed effect estimation is based on the assumption that both the unobserved characteristics A_i and their influence on the dependent variable do not vary over time. Therefore, we can capture these unobserved characteristics by a single parameter α_i. The model is based on the identifying assumption that given the cross-section fixed effect and the covariates, individuals in blue collar employment are similar to individuals in white collar employment:

$$E[health\_status_{0it}|\alpha_i, \lambda_t, X_{it}, occupation_{it}] = E[health\_status_{0it}|\alpha_i, \lambda_t, X_{it}] \quad (3)$$

where the parameter α_i is the cross-section fixed effect.

We identify the effect of occupation health by the following linear equation:

$$health\_status_{it} = \alpha_i + \lambda_t + \rho occupation_{it} + X_{it}\delta + \varepsilon_{it} \quad (4)$$

From (4) follows that:

$$\overline{health\_status}_i = \alpha_i + \bar{\lambda} + \rho\overline{occupation}_i + \bar{X}_i\delta + \bar{\varepsilon}_i \quad (5)$$

where $\overline{health\_status}_i$, $\bar{\lambda}$, $\overline{occupation}_i$, $\bar{X}_i$ and $\bar{\varepsilon}_i$ are the individual's averages of respectively $health\_status_{it}$, λ_t, $occupation_{it}$, X_{it} and ε_{it} over time[5].

Since the number of cross-section fixed effects equals the number of individuals in the sample, the cross-section fixed effects are not consistently estimated in a panel where the number of periods T is fixed while the sample size $N \rightarrow \infty$. However, the other parameters are consistently estimated.
Since we are not interested in the cross-section fixed effects themselves, but only in the other parameters, we kill these unobserved fixed effects by estimating in deviations from means:

$$\begin{aligned} health\_status_{it} - \overline{health\_status}_i \\ = (\lambda_t - \bar{\lambda}) + \rho(occupation_{it} - \overline{occupation}_i) + (X_{it} - \bar{X}_i)\delta + (\varepsilon_{it} - \bar{\varepsilon}_i) \end{aligned} \quad (6)$$

However, we should correct the standard errors for the fact that degrees of freedom are lost in estimating the means.

Furthermore, from (4) follows that:

$$\overline{\overline{health\_status}} = \bar{\alpha} + \bar{\lambda} + \rho\overline{\overline{occupation}} + \bar{\bar{X}}\delta + \bar{\bar{\varepsilon}} \quad (7)$$

where $\overline{\overline{health\_status}}$, $\bar{\alpha}$, $\overline{\overline{occupation}}$, $\bar{\bar{X}}$ and $\bar{\bar{\varepsilon}}$ are the grand averages of respectively $health\_status_{it}$, α_i, $occupation_{it}$, X_{it} and ε_{it}[6].

---

[5] More formally, $\overline{health\_status}_i \equiv \frac{\sum_{t=1}^{T} health\_status_{it}}{T}$, $\bar{\lambda} \equiv \frac{\sum_{t=1}^{T}\lambda_t}{T}$, $\overline{occupation}_i \equiv \frac{\sum_{t=1}^{T} occupation_{it}}{T}$, $\bar{X}_i \equiv \frac{\sum_{t=1}^{T} X_{it}}{T}$ and $\bar{\varepsilon}_i \equiv \frac{\sum_{t=1}^{T}\varepsilon_{it}}{T}$.

[6] More formally, $\overline{\overline{health\_status}} \equiv \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} health\_status_{it}}{N \cdot T}$, $\bar{\alpha} \equiv \frac{\sum_{i=1}^{N}\alpha_i}{N}$, $\overline{\overline{occupation}} \equiv \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} occupation_{it}}{N \cdot T}$, $\bar{\bar{X}} \equiv \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}}{N \cdot T}$ and $\bar{\bar{\varepsilon}} \equiv \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}}{N \cdot T}$.

We sum (6) and (7) in order to be able to identify the average cross-section fixed effect $\bar{\alpha}$ in addition to the parameters ρ and δ (Gould, 2011):

$$\begin{aligned} health\_status_{it} - \overline{health\_status_i} + \overline{\overline{health\_status}} \\ = \bar{\alpha} + \lambda_t + \rho\left(occupation_{it} - \overline{occupation_i} + \overline{\overline{occupation}}\right) \\ + \left(X_{it} - \bar{X}_i + \bar{\bar{X}}\right)\delta + (\varepsilon_{it} - \bar{\varepsilon}_i + \bar{\bar{\varepsilon}}) \end{aligned} \quad (8)$$

Since the effects of time-constant characteristics are captured by the cross-section fixed effect, we do not include variables that differ between individuals, but not over time, like for example gender, race, birth weight and parental education in the regression. Furthermore, since all period effects are captured by the year dummies, we do not include variables that differ over time, but not among individuals, like the state of the economy and changes in health policy. Finally, the effect of age can be separated in a time-invariant individual-specific part and a common time-variant part: $age_{it}\delta_{age} = age_{i1}\delta_{age} + (t-1)\delta_{age}$. Since the age at the beginning is captured by the cross-section fixed effect and since the increase in age is captured by the year effects, we do not need to include age in the regression (Ferrer-i-Carbonnel and Frijters, 2004).
However, since health deterioration is likely to accelerate at higher ages, we want to control for a non-linear effect of age. Since the marginal effect of squared age depends on the age of the individual, this effect cannot be captured by the year effects. Therefore, we do include squared age besides the cross-section fixed effects and year effects as control variable in our fixed effects model.
Again, we do not control for variables that may act as mechanisms.

3.1.3. Fixed effects and lagged dependent variables

Finally, we try to estimate a model that includes both unobserved cross-section fixed effects and lagged dependent variables, since the identifying assumption of this model is weaker (Angrist and Pischke, 2008, chapter 5). We assume that that after controlling for the cross-section fixed effects, the lagged health status and the included covariates, no selection bias is left:

$$\begin{aligned} E\left[health\_status_{0it} \middle| \alpha_i, health_{status_{it-1}}, \lambda_t, X_{it}, occupation_{it}\right] \\ = E\left[health\_status_{0it} \middle| \alpha_i, health_{status_{it-1}}, \lambda_t, X_{it}\right] \end{aligned} \quad (9)$$

We identify the effect of occupation health by the following linear equation:

$$health\_status_{it} = \alpha_i + \theta health\_status_{it-1} + \lambda_t + \rho occupation_{it} + X_{it}\delta + \varepsilon_{it} \quad (10)$$

In order to kill the fixed effects, we take first differences[7]:

$$\Delta health\_status_{it} = \theta \Delta health\_status_{it-1} + \Delta\lambda_t + \rho\Delta occupation_{it} + \Delta X_{it}\delta + \Delta\varepsilon_{it} \quad (11)$$

where $\Delta health\_status_{it}$, $\Delta health\_status_{it-1}$, $\Delta\lambda_t$, $\Delta occupation_{it}$, ΔX_{it} and $\Delta\varepsilon_{it}$ are the first differences of respectively $health\_status_{it}$, $health\_status_{it-1}$, λ_t, $occupation_{it}$, X_{it} and ε_{it}[8].

[7] Note that using deviations from means would cause that the error term is a function of all lags of ε_{it}. Hence, this error term is necessarily correlated with all lags of health status. Consequently, no lag of health status can be used as an instrument to solve the problem of exogeneity.

Although this model does properly account for the economic selection bias, this model introduces an econometric bias: since $\Delta\varepsilon_{it}$ is necessarily correlated with the lagged dependent variable, OLS estimates are inconsistent. We try to solve this exogeneity problem by using lags of $health\_status_{it-1}$ as instruments for $\Delta health\_status_{it-1}$. In order to be a good instrument, these lags should (i) be exogenous to $\Delta\varepsilon_{it}$ and (ii) be sufficiently correlated with $\Delta health\_status_{it-1}$. However, if there is autocorrelation in the residuals, there may be no valid instrument and hence, no consistent estimator.

We use the Sargan test in order to test whether correlation exists between the error terms and the instruments. The Sargan test procedure involves regressing the residuals of the instrumental variables estimation on the instruments. The null hypothesis of exogeneity will be rejected if the instruments have significant explanatory power for the residuals (Heij et al, 2004, chapter 5).

For similar reasons as presented in the previous subsections, we only control for squared age besides the lagged health status, the cross-section fixed effect and the year effects

3.2. Smoking ban

We estimate additional fixed effects models to isolate the impact of smoke-free workplaces on the prevalence of smoking, the consumption of cigarettes and health. We estimate (8), using successively the smoking status, the consumption of cigarettes and the health status as dependent variables and replacing the occupation dummy by a dummy that indicates whether the smoking ban has been implemented.

Since the smoking ban has been implemented in England, Wales and Northern Ireland a year later than in Scotland, we can both estimate the effect of the smoking ban on health and control for differences in self-reported health over time if we assume that both before and after the implementation of the smoking ban health evolves similarly in Scotland and the rest of the United Kingdom. In case of parallel trends, we can use the trend in the other country as reference in order to estimate what the health would have been if the ban would not have been introduced. However, if the trends are not similar, we cannot reliable estimate how health would have changed if the ban would not have been introduced.

We can test the parallel trends assumption by including interactions between the wave dummies and a dummy indicating whether the individual lives in Scotland in the regression[9]. If the parameters of these interactions differ from each other, then health does not evolve similarly in both countries after controlling for differences between individuals. Therefore, we perform a

[8] More formally, $\Delta health\_status_{it} \equiv health\_status_{it} - health\_status_{it-1}$, $\Delta health\_status_{it-1} \equiv health\_status_{it-1} - health\_status_{it-2}$, $\Delta\lambda_t \equiv \lambda_t - \lambda_{t-1}$, $\Delta occupation_{it} \equiv occupation_{it} - occupation_{it-1}$, $\Delta X_{it} \equiv X_{it} - X_{it-1}$ and $\Delta\varepsilon_{it} \equiv \varepsilon_{it} - \varepsilon_{it-1}$.

[9] We do not include the interaction between wave 16 and the dummy indicating whether the individual lives in Scotland, since this interaction is used to identify the effect of the smoking ban:

$$ban_{it} = \begin{cases} 0 & \text{if } t < 16 \\ Scotland_{it} & \text{if } t = 16 \\ 1 & \text{if } t > 16 \end{cases}$$

where ban_{it} indicates whether the smoking ban has been implemented and $Scotland_{it}$ indicates whether the individual lives in Scotland.

Wald test by restricting the parameters of the interaction terms to be equal in order to test whether the parameters differ significantly from each other.

Note that this identification strategy assumes that the introduction of the smoking ban has no preliminary or delayed effects on health. Both would influence the year effects and hence, the parameter of the smoking ban.

Again, we include squared age besides the cross-section fixed effects and year effects as control variable in these fixed effects models.

4. Results

4.1. Occupation

4.1.1. Lagged dependent variables

We regress the smoking status on the occupation dummy, while controlling for lagged health status, year effects, gender, age, squared age and highest educational qualification. We initially treat the explanatory lagged health status as cardinal variable, meaning that we include one variable with the health status category as value. The parameter estimates are shown in table 3. We find that white collar employment significantly improves health as compared to blue collar employment.

We test for joint significance of the highest educational qualification dummies. We find that these dummies are jointly significant ($F = 6.217$, $p = 0.000$). The year effects are jointly significant as well ($F = 4.548, p = 0.000$).

In order to test for heteroskedasticity, we regress the squared residuals on the year dummies. These year effect are jointly significant ($F = 12.539$, $p = 0.000$), which indicates that there is significant heteroskedasticity. Moreover, the Durbin-Watson statistic suggests that the residuals are weakly autocorrelated ($DW = 2.205$). Therefore, we calculate White period standard errors, which account for between-period correlation of the standard errors. The Jarque-Bera test indicates that the residuals are not normally distributed ($S = 0.547$, $K = 3.828$, $JB = 4520.129$, $p = 0.000$). However, the skewness and the kurtosis do not signal extreme values.

Both age and squared age are insignificant in the previous model. Since the p-value that belongs to squared age ($p = 0.791$) exceeds the p-value associated with age ($p = 0.484$), we re-estimate the previous model without squared age as explanatory variable. We report the parameter estimates in table 3. Again, we find that white collar employment significantly improves health.

Both the highest educational qualification dummies ($F = 6.345$, $p = 0.000$) and the year effects ($F = 4.554, p = 0.000$) are jointly significant.

Our heteroskedasticity test shows that there is significant heteroskedasticity ($F = 12.580$, $p = 0.000$) and the Durbin-Watson statistic suggests that the residuals are weakly autocorrelated ($DW = 2.205$). We account for this by calculating White period standard errors. The Jarque-Bera test indicates that the residuals are not normally distributed ($S = 0.547$, $K = 3.828$, $JB = 4519.715$, $p = 0.000$), but the skewness and the kurtosis indicate that the distribution is not very different from the normal distribution.

Table 3 - Regressions occupation: lagged dependent variables

Dependent variable	Health status	Health status	Health status	Health status
Sample	Full	Full	Full	Full
Specification	OLS	OLS	OLS	Ordered probit
Cross-section fixed-effects	No	No	No	No
Year effects	Yes	Yes	Yes	Yes
Occupation	-0.031***	-0.031***	-0.031***	-0.060***
	(0.008)	(0.008)	(0.008)	(0.012)
Lagged health status	0.499***	0.499***		0.837***
	(0.006)	(0.006)		(0.007)
- Excellent health[1]			-1.014***	
			(0.012)	
- Good or very good health[1]			-0.511***	
			(0.011)	
- Poor or very poor health[1]			0.439	
			(0.026)	
Gender	-0.043***	-0.043***	-0.043***	-0.077***
	(0.008)	(0.008)	(0.008)	(0.010)
Age	0.001	0.002***	0.002***	0.003***
	(0.002)	(0.000)	(0.000)	(0.000)
Age-squared	0.000			
	(0.000)			
Highest educational qualification				
- Higher degree[2]	-0.070	-0.073	-0.071	-0.139
	(0.070)	(0.070)	(0.070)	(0.119)
- First degree[2]	-0.099	--0.100	-0.099	-0.186
	(0.068)	(0.068)	(0.068)	(0.117)
- Teaching qualification[2]	-0.094	-0.086	-0.085	-0.153
	(0.071)	(0.071)	(0.071)	(0.120)
- Other higher qualification[2]	-0.030	-0.032	-0.030	-0.066
	(0.068)	(0.068)	(0.067)	(0.116)
- Nursing qualification[2]	-0.093	-0.094	-0.093	-0.166
	(0.072)	(0.072)	(0.072)	(0.123)
- GCE A levels[2]	-0.036	-0.038	-0.036	-0.069
	(0.068)	(0.068)	(0.068)	(0.117)
- GCE O levels[2]	-0.025	-0.027	-0.026	-0.054
	(0.068)	(0.068)	(0.068)	(0.116)
- Commercial qualification[2]	-0.026	-0.027	-0.027	-0.049
	(0.072)	(0.072)	(0.072)	(0.121)
- GCSE grades[2]	-0.014	-0.016	-0.014	-0.028
	(0.070)	(0.070)	(0.069)	(0.119)
- Apprenticeship[2]	-0.029	-0.030	-0.028	-0.053
	(0.076)	(0.076)	(0.075)	(0.126)
- Other qualification[2]	-0.023	-0.025	-0.023	-0.065
	(0.083)	(0.084)	(0.083)	(0.135)
- No qualification[2]	0.013	0.011	0.013	0.006
	(0.069)	(0.069)	(0.068)	(0.117)
Constant	1.005***	0.998***	2.508***	
	(0.073)	(0.069)	(0.068)	
(Pseudo) R-squared	0.256	0.256	0.256	0.131
Individuals	9178	9178	9178	9178
Observations	57643	57643	57643	57643

White period standard error between brackets. *** p<0.01, ** p<0.05, * p<0.1

[1] As compared to fair health. [2] As compared to still at school.

We check the robustness of our results by treating the explanatory lagged health status as ordinal value, implying that we include a dummy for each lagged health status except fair health, which is our reference category. Again, the results are shown in table 3. We still find that white collar employment significantly improves health.
We test the validity of the cardinality assumption for the explanatory lagged health status with a Wald test on the restriction $\frac{1}{2}\delta_{health\_status_{-1}=1} = \delta_{health\_status_{-1}=2} = -\delta_{health\_status_{-1}=4}$. For a mathematical derivation of this restriction, we refer to *Appendix B – Mathematical derivation cardinality restriction*. We do not reject the null hypothesis of cardinality on 5% significance level ($F = 2.990, p = 0.050$).
Again, the highest educational qualification dummies ($F = 6.379, p = 0.000$) and the year effects ($F = 4.489, p = 0.000$) are jointly significant.
Our test indicate that the residuals are heteroskedastic ($F = 12.580$, $p = 0.000$), weakly autocorrelated ($DW = 2.205$) and not normally distributed ($S = 0.550$, $K = 3.824$, $JB = 4539.462, p = 0.000$). However, the skewness and the kurtosis do not show abnormal values.

Furthermore, we check the robustness of the results by estimating an ordered probit model. The parameter are shown in table 3. We find no qualitative different results as compared to the OLS estimates. This is in line with the findings of Ferrer-i-Carbonnel and Frijters (2004).

4.1.2. Fixed effects

Subsequently, we try to identify the impact of general labor conditions by estimating a fixed effects model. We regress health status on occupation, while we control for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 4. We find no significant evidence that occupation influences health.
We test for joint significance of the cross-section fixed effects and the year effects and find that both the cross-section fixed-effects ($F = 5.157$, $p = 0.000$) and the year effects ($F = 7.209$, $p = 0.000$) are jointly significant.

Table 4 - Regressions occupation: fixed effects

Dependent variable	Health status
Sample	Full
Specification	OLS
Cross-section fixed-effects	Yes
Year effects	Yes
Occupation	-0.011
	(0.012)
Age-squared	0.000***
	(0.000)
Constant	1.478***
	(0.074)
R-squared	0.513
Individuals	12235
Observations	72969

White period standard error between brackets. *** p<0.01, ** p<0.05, * p<0.1

Our heteroskedasticity test indicates that the residuals are heteroskedastic ($F = 7.521$, $p = 0.000$). Furthermore, the Durbin-Watson statistic suggests that the residuals are weakly autocorrelated ($DW = 2.024$). Therefore, we calculate White period standard errors. The Jarque-Bera test indicates that the residuals are not normally distributed ($S = 0.458$, $K = 4.330$, $JB = 7921.668$, $p = 0.000$). However, based on the skewness and kurtosis, we conclude that the residuals are roughly normally distributed.

4.1.3. Fixed effects and lagged dependent variables

Finally, we try to estimate a model in which we control for both fixed effects and the lagged health status. We also control for squared age. As explained in the *Methodology*, we need to find a valid instrument in order to solve the exogeneity problem.

We start by using the second and all available preceding lags as instrumental variables. We perform the Sargan test on the validity of instruments. We reject the null hypothesis of exogeneity ($LM = 249.385$, $p = 0.000$). Hence, this set of instruments is invalid. We repeat this procedure by using the third and all available preceding lags as instrumental variables. Again, the set of instruments is inappropriate ($LM = 187.932$, $p = 0.000$). The fourth and all available preceding lags are not valid either ($LM = 135.382$, $p = 0.015$).

However, we do not find significant evidence at 5% level that the set of the fifth and all available preceding lags is not exogenous to the error term ($LM = 107.821$, $p = 0.080$). In order to further test the suitability of this set of instruments, we regress the health status on the fourth lag of health status, where we exclude the last wave from the sample[10]. The fourth lag of health status is significantly correlated with health status ($t = 78.234$, $p = 0.000$). Hence, the fifth and all available preceding lags seem to be a good set of instruments.

The instrumental variables estimation results with this set of instruments are reported in table 5. The impact of occupation is insignificant.

The lagged health status does not have a significant impact on health. We test for the significance of the year effects and find that the year effects are jointly significant ($F = 4.505$, $p = 0.000$).

We use our heteroskedasticity test in order to conclude that the residuals are heteroskedastic ($F = 2.732$, $p = 0.000$). Hence, we calculate White period standard errors. Furthermore, the Jarque-Bera test shows that the residuals are not normally distributed ($S = 0.080$, $K = 4.317$, $JB = 3822.733$, $p = 0.000$). However, the skewness and the kurtosis are not extremely high.

Since the squared age parameter is not significant, we repeat the above procedure without using squared age as control variable. Again, the set of the fifth and all available preceding lags seems to be a good set of instruments ($LM = 106.866$, $p = 0.095$).

The estimated parameters are shown in table 5. The parameter of occupation is not significant.

In this model, the lagged dependent health status has no significant effect on health. However, the year effects are jointly significant ($F = 8.042$, $p = 0.000$).

We find that the residuals are heteroskedastic ($F = 2.831$, $p = 0.000$) and not normally distributed ($S = 0.060$, $K = 4.297$, $JB = 3687.581$, $p = 0.000$), but the skewness and the kurtosis do not show extreme values.

[10] Note that this is similar to regressing the first lag of health status on the fifth lag of health status.

Since the effect of the lagged health status is not significant, the fixed effects model seems most appropriate in order to identify the causal effect of occupation on health.

Table 5 - Regressions occupation: fixed effects and lagged dependent variables

Dependent variable	Health status	Health status
Sample	Full	Full
Specification	OLS	OLS
Cross-section fixed-effects	Yes	Yes
Year effects	Yes	Yes
Lagged health status	0.026	0.055
	(0.090)	(0.089)
Occupation	-0.252	-0.338
	(0.284)	(0.274)
Age-squared	0.000	
	0.000	
Individuals	8072	8072
Observations	52131	52135

White period standard error between brackets. *** p<0.01, ** p<0.05, * p<0.1

4.2. Smoking ban

First of all, we examine whether the introduction of the smoking ban affected the prevalence of smoking. We regress the smoking status on a dummy indicating the smoking ban, while accounting for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 6. We find no significant evidence that the introduction of the smoking ban reduced the prevalence of smoking.
We test for joint significance of both the cross-section fixed effects and the year effects. Both the cross-section fixed effects ($F = 32.618$, $p = 0.000$) and the year effects ($F = 6.713$, $p = 0.000$) are jointly significant.
We test for similar trends as described in the *Methodology*. We do not find significant evidence that the prevalence of smoking does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 0.469$, p =0.857). Hence, the parallel trends assumption seems satisfied.
In order to test for heteroskedasticity, we regress the squared residuals on the year effects. The year effect are jointly significant ($F = 16.123$, $p = 0.000$), which indicates that there is significant heteroskedasticity. Furthermore, the Durbin-Watson statistic suggests that the residuals are autocorrelated ($DW = 1.512$). Therefore, we calculate White period standard errors, which account for between-period correlation of the standard errors. The Jarque-Bera test indicates that the residuals are not normally distributed ($S = -0.342$, $K = 12.283$, $JB = 196296.1$, $p = 0.000$). The high kurtosis indicates that we are dealing with fat tails in the distribution of the error terms. Since we underestimate the standard errors by assuming a normal distribution, we should be careful to conclude significant relationships.

Since squared age does not have a significant effect on health status in the previous model, we remove squared age as a control variable from the regression. The resulting estimates are shown

in table 6. We still find no significant evidence that the introduction of the smoking ban reduced the smoking prevalence.
Both the cross-section fixed effects ($F = 32.811$, $p = 0.000$) and the year effects ($F = 27.883$, $p = 0.000$) are jointly significant.
We do not find significant evidence that the cigarette consumption does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 0.466, p = 0.860$).
Our heteroskedasticity test indicates that the residuals are heteroskedastic ($F = 16.121$, $p = 0.000$) and the Durbin-Watson statistic suggests that the residuals are serially correlated ($DW = 1.512$). The Jarque-Bera test indicates that the residuals are not normally distributed ($S = -0.342$, $K = 12.284$, $JB = 196366.6$, $p = 0.000$). Again, the kurtosis is high, standard errors are underestimated and we should be careful to conclude significant relationships.

Subsequently, we investigate whether the introduction of the smoking ban affected the consumption of cigarettes. We regress the daily cigarette consumption on a dummy indicating the smoking ban, while accounting for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 6. We find no significant evidence that the introduction of the smoking ban reduced the consumption of cigarettes.
Both the cross-section fixed effects ($F = 37.456$, $p = 0.000$) and the year effects ($F = 2.586$, $p = 0.008$) are jointly significant.
We do not find significant evidence that the cigarette consumption does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 0.643, p = 0.721$).
The tests indicate that the residuals are heteroskedastic ($F = 8.469, p = 0.000$), autocorrelated ($DW = 1.491$) and not normally distributed ($S = 0.045, K = 20.846, JB = 719847.8, p = 0.000$). Again, the high kurtosis indicates that we should be careful to conclude significant effects.

Table 6 - Regressions smoking ban: smoking status and daily cigarette consumption

Dependent variable	Smoking status	Smoking status	Daily cigarette consumption
Sample	Full	Full	Full
Specification	OLS	OLS	OLS
Cross-section fixed-effects	Yes	Yes	Yes
Year effects	Yes	Yes	Yes
Ban	0.000	0.000	-0.104
	(0.006)	(0.006)	(0.105)
Age-squared	0.000		-0.001***
	(0.000)		(0.000)
Constant	0.262***	0.249***	5.649***
	(0.040)	(0.002)	(0.629)
R-squared	0.822	0.822	0.841
Individuals	6716	6716	6716
Observations	54376	54381	54244

White period standard error between brackets. *** p<0.01, ** p<0.05, * p<0.1

Although we do not find a significant effect of the introduction of the smoking ban on the prevalence of smoking or the consumption of cigarettes, the ban did induce people to smoke elsewhere. Hence, health may be affected due to decreased passive smoking on the job.
We investigate the effect of the smoking ban on health by regressing the health status on a dummy indicating the smoking ban, while we account for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 7. We find no significant evidence that the introduction of the smoking ban affected health.
Both the cross-section fixed effects ($F = 7.444$, $p = 0.000$) and the year effects ($F = 8.365$, $p = 0.000$) are jointly significant.
We do not find significant evidence that health does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 1.052$, $p = 0.392$). Thus, the parallel trends assumption seems satisfied.
The residuals are heteroskedastic ($F = 7.877$, $p = 0.000$) and weakly serial correlated ($DW = 1.939$). Therefore, we calculate White period standard errors. The residuals are not normally distributed ($S = 0.573$, $K = 4.794$, $JB = 10487.31$, $p = 0.000$), but the skewness and the kurtosis do not signal extreme values.

We found no significant evidence that the introduction of the smoking ban affected health. However, this result might exist due to opposing effects for smokers and non-smokers: smokers may experience more stress, whereas non-smokers are redeemed from passive smoking.
In order to examine the specific effect on smokers, we further restrict the sample to all individuals that smoke in the period before the introduction of the ban[11]. Again, we regress the health status on a dummy indicating the smoking ban, while we account for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 7. We find no significant evidence that the introduction of the smoking ban affected health for smokers.
Both the cross-section fixed effects ($F = 6.752$, $p = 0.000$) and the year effects ($F = 3.082$, $p = 0.002$) are jointly significant.
We do not find significant evidence that health for smokers does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 1.187$, $p = 0.306$).
The residuals are heteroskedastic ($F = 2.353$, $p = 0.016$), weakly autocorrelated ($DW = 1.951$) and not normally distributed ($S = 0.524$, $K = 4.630$, $JB = 2045.825$, $p = 0.000$), but the skewness and kurtosis do not show alarming values.

Finally, we consider the impact of the introduction of the smoking ban on the health of non-smokers. We restrict the sample to all individuals that do not smoke in the period before the introduction of the smoking ban[12]. Again, we regress the health status on a dummy indicating the smoking ban, while we account for cross-section fixed effects, year effects and squared age. The parameter estimates are shown in table 7. We find no significant evidence that the introduction of the smoking ban affected health for non-smokers.

[11] More precisely, individuals in wave 16 that live in Scotland should smoke in wave 15 in order to be included in the sample, whereas individuals in wave 16 that live in other parts of the United Kingdom should smoke in wave 16.
[12] More precisely, individuals in wave 16 that live in Scotland should not smoke in wave 15 in order to be included in the sample, whereas individuals in wave 16 that live in other parts of the United Kingdom should not smoke in wave 16.

Both the cross-section fixed effects ($F = 7.486$, $p = 0.000$) and the year effects ($F = 6.065$, $p = 0.000$) are jointly significant.
We do not find significant evidence that health for non-smokers does not evolve similarly in Scotland and other parts of the United Kingdom ($F = 1.293, p = 0.249$).
The residuals are heteroskedastic ($F = 6.150, p = 0.000$), weakly autocorrelated ($DW = 1.932$). and not normally distributed ($S = 0.598$, $K = 4.841$, $JB = 8388.166$, $p = 0.000$). However, the skewness and the kurtosis do not strongly deviate from normal values.

Table 7 - Regressions smoking ban: health status

Dependent variable	Health status	Health status	Health status
Sample	Full	Smoker	Non-smoker
Specification	OLS	OLS	OLS
Cross-section fixed-effects	Yes	Yes	Yes
Year effects	Yes	Yes	Yes
Ban	-0.000	0.003	-0.002
	(0.022)	(0.048)	(0.025)
Age-squared	0.000***	0.000	0.000***
	(0.000)	(0.000)	(0.000)
Constant	1.632***	1.840***	1.555***
	(0.090)	(0.179)	(0.106)
R-squared	0.509	0.488	0.509
Individuals	6758	1608	5055
Observations	55543	13071	41795

White period standard error between brackets. *** p<0.01, ** p<0.05, * p<0.1

5. Conclusions

5.1. Occupation

We used data from the British Household Panel Survey (BHPS) and estimated the impact of general labor conditions on health with two different models. The results of our lagged dependent variables model suggests that white collar employment improves health as compared to blue collar employment. However, we do not find a significant relationship in our fixed effects model. Angrist and Pischke (2008) suggest to interpret the estimates of the two models as bounds of the true effects. Hence, we can conclude that we did not find unambiguous evidence that general labor conditions affect health.

Furthermore, we did estimate a model that controls for both fixed effects and lagged dependent variables by using an instrumental variables approach. We found no significant effect of occupation on health. Since the lagged dependent variable had no significant effect in this model, the fixed effects model seems most appropriate to identify the causal effect of occupation on health. Thus, we can conclude that we did not identify a significant impact of labor conditions in general on health.

In order to be careful in concluding significant relations, we did choose to rule out potential distorting effects of intervening activities like entrepreneurship, unemployment, maternity, family care or disabilities, by restricting our sample to all individuals that are employed in all available waves after schooling and before retirement. Therefore, individuals that became disabled because of labor conditions are excluded from the sample. Hence, we are likely to underestimate the causal effect of general labor conditions on health.

5.2. Smoking ban

The introduction of the smoking ban did not lead to significant disproportional direct changes in smoking behavior: both the smoking prevalence and the consumption of cigarettes did not decrease directly after the introduction of the ban. We cannot conclude that the introduction of the smoking ban did not affect smoking behavior at all: individuals might have anticipated the introduction of the ban and hence, decreased smoking already before the introduction of the ban. Furthermore, the introduction of the smoking ban might have a delayed effect on smoking behavior. Note that that our identification strategy assumes that the introduction of the smoking ban does not have a preliminary or delayed effect. Both would affect the year effects and thus, the parameter of the smoking ban. Therefore, we cannot definitively establish that the smoking ban has no direct effect on smoking behavior at all.

Although we did not find a reduction in the prevalence of smoking and the consumption of cigarettes, health might have improved due to decreased passive smoking. However, for both smokers and non-smokers, we do not find a significant disproportional direct change in health.

We can conclude that smoke-free workplaces are not a mechanism trough which labor conditions in general affect health in the short run.

5.3. Limitations and further research

We used a unbalanced panel. The reason for missing data might be correlated with the error terms (Wooldridge, 2009). Non-healthy people may be more likely to die, for example. We did not account for this correlation in our analysis. Further research is needed to investigate on the missing data and its effects for the parameter estimates.

Our research has been devoted to examine the direct impact of labor conditions on health. Further research need to be done to identify the long-term effects of both general labor conditions and smoke-free workplaces on health.

6. References

Angrist, Joshua David and Jörn-Steffen Pischke (2008), *Mostly harmless econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

BHPS (2012a), *BHPS*. Retrieved from: https://www.iser.essex.ac.uk/bhps

BHPS (2012b), *Design*. Retrieved from: https://www.iser.essex.ac.uk/bhps/faqs/design

Ferrer-i-Carbonnel, Ada and Paul Frijters (2004), How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497): 641-659.

Fichtenberg, Caroline M. and Stanton A. Glantz (2002), Effect of smoke-free workplaces on smoking behavior: systematic review. *British Medical Journal*, 325(7357): 188-194.

Fletcher, Jason M. and Jody L. Sindelar (2009), *Estimating causal effects of early occupational choice on later health: evidence using the PSID*. NBER Working Paper Series, 15256. National Bureau of Economic Research, Cambridge, MA.

Gould, William (2011), *Interpreting the intercept in the fixed-effects model*. Retrieved from: http://www.stata.com/support/faqs/stat/xtreg2.html.

Heij, Christiaan, Paul de Boer, Philip Hans Franses, Teun Kloek and Herman K. van Dijk (2004), *Econometric methods with applications in business and economics*. Oxford University Press, Oxford.

Jones, Andrew M., Audrey Laporte, Nigel Rice and Eugenio Zucchelli (2011), *A model of the impact of smoking bans on smoking with evidence from bans in England and Scotland*. HEDG Working Papers, 11/05. Health, Econometrics and Data Group, Department of Economics, University of York.

Marmot, Michael G. and George Davey Smith (1997), Socio-economic differences in health. *Journal of Health Psychology*, 2(3): 283-296.

Marmot, Michael G., H. Bosma, H. Hemingway, E. Brunner and S. Stansfeld (1997), Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet*, 350: 235-239.

Taylor, Marcia Freed, John Brice, Nick Buck and Elaine Prentice-Lane (2010), *British Household Panel Survey User manual Volume A: Introduction, technical report and appendices*. Institute for Social and Economic Research, University of Essex.

Wooldridge, Jeffrey M. (2009). *Introductory econometrics: A modern approach*. South-Western Cengage Learning, Mason, OH.

7. Appendix A – Educational qualification categories

The British Household Panel Survey (BHPS) uses the same definition of educational qualification categories as the General Household Survey (Taylor et al, 2010):

1. University or CNAA Higher Degree
2. University or CNAA First Degree
3. Teaching Qualifications
4. City & Guilds Certificate (Full Technological/Part III), HNC, HND, BEC/TEC/BTEC Higher Certificate/Diploma, University Diploma, Any other technical, professional or higher qualifications

5. Nursing Qualifications
6. A Levels, Scottish Higher Grades, Scottish School Leaving Certificate Higher Grade, Scottish Certificate of Sixth Year Studies, Higher School Certificate, Ordinary National Certificate/Diploma, BEC/TEC/BTEC National/General Certificate or Diploma or City & Guilds Certificate (Advanced/Final/Part II)
7. O Levels (pre 1975), O Level grades A-C (1975 or later), GCSE grades A-C, CSE grade 1, Scottish O Grades (pass or bands A-C or 1-3), Scottish School Leaving Certificate Lower Grade, School Certificate or Matric, Scottish Standard Grade Level 1-3 or City & Guilds Certificate (Craft/Intermediate/Ordinary/Part I)
8. Clerical or Commercial Qualifications
9. CSE Grades 2-5, O Level grades D-E, GCSE grades D-G, Scottish SCE Ordinary Grade bands D-E or 4-5 or Scottish Standard Grade levels 4-7
10. Recognised trade apprenticeship
11. Youth Training Certificate, Any other qualifications
12. No qualifications and not at school
13. No qualifications and still at school

8. Appendix B – Mathematical derivation cardinality restriction

We rewrite the restricted model with the explanatory lagged health status as cardinal variable in the unrestricted model with the explanatory lagged health status as ordinal variable:

$$\alpha + \delta_{health\_status_{-1}} health\_status_{it-1} = \alpha + \sum_{k=1}^{4} \delta_{health\_status_{-1}} k I(health\_status_{it-1} = k) \quad (12)$$

$$= \beta + \sum_{k=1}^{4} \delta_{health\_status_{-1}=k} I(health\_status_{it-1} = k)$$

where $I(health\_status_{it-1} = k)$ equals 1 if the condition $health\_status_{it-1} = k$ is satisfied and 0 otherwise.

This yields the next set of restrictions:

$$\alpha + k\delta_{health\_status_{-1}} = \beta + \delta_{health\_status_{-1}=k}, \qquad \text{for } k = 1,2,3,4 \quad (13)$$

Subtracting the restriction for $k = m$ from the restriction for $k = m - 1$ yields:

$$\delta_{health\_status_{-1}} = \delta_{health\_status_{-1}=m} - \delta_{health\_status_{-1}=m-1}, \qquad \text{for } m = 2,3,4 \quad (14)$$

We use fair health as reference category. Hence, we restrict $\delta_{lagged_{health_{status}}=3}$ to be zero and we can rewrite the restrictions on $\delta_{health\_status_{-1}=k}$:

$$\frac{1}{2}\delta_{health\_status_{-1}=1} = \delta_{health\_status_{-1}=2} = -\delta_{health\_status_{-1}=4} \quad (15)$$