Determinants of Self-Perceived Health in the UK:

An Econometric Analysis

Ryan Granger

BSc Economics & Econometrics with Professional Experience

April 2021

Final Word Count (Excluding Bibliography & Appendices): 7052

Abstract

This dissertation investigates the determinants of self-perceived health in the UK. The primary aim of this research is to establish which factors have the greatest impact on self-perceived health. My motivation for this work stems from the idea that it is important to understand whether or not people *feel* healthy, regardless of whether or not they actually *are* in good health. There are a number of benefits which arise out of feeling healthy, such as higher levels of productivity, increased motivation, reduced risk of mental illness etc. It is therefore important that those factors which determine self-perception of health are well understood.

This paper makes use of an empirical approach, using a cross-sectional dataset taken from the UK Household Longitudinal Survey (UKHLS). Ordinary least squares (OLS) estimation is initially employed, before moving to non-linear methods, built upon maximum likelihood estimation (MLE). Initially a binary response variable is used before a categorical variable is introduced, allowing for the use of a multinomial probit model.

The results show that gender does not have a statistically significant effect on self-perceived healthcare outcomes. Results also showed that existence of a long-term illness, smoking and living in an urban area were the three factors which had the largest negative impact on healthcare perception. Alcohol consumption, holding a degree, being in employment, and income were all factors which increased the probability of an individual reporting that they are in good health.

These results are explored in the context of policymaking, with a view to being able to better design policy such that the general population feel more optimistic/positive about their health. In particular, I discuss policy interventions in relation to cigarette use, traffic congestion in built-up areas, and the interplay between education and employment.

Table of Contents

1. Introduction

Healthcare outcomes and statistics of health are widely reported and discussed in the literature. When considering healthcare outcomes, many turn to macroeconomic-level aggregated variables, such as mortality rate or average life expectancy (Marmot et al. 2012). Other literature explores healthcare outcomes through the lens of microeconomic theory, behavioural considerations, and the application of micro-econometric methods (Rice, 2013; Fischer & Sousa-Poza, 2009). It is this fusion of microeconomics and econometrics upon which I base the research presented in this dissertation. Focussing on the micro-level, using a large-scale, longitudinal data set, allows for a more nuanced analysis. Of course, it would be possible to focus on specific health problems, such as high blood pressure, type 2 diabetes, hypertension etc. However, this research comes at the issue from a slightly different angle.

This paper is interested in how individuals themselves perceive their level of health. This involves looking not at a persons' medical records, or talking to a medical professional, but instead asking people:

"*On a scale from 1 – 5, how would you rate your health?"*

This is of interest, as it will allow me to explore a range of covariates in order to infer how various factors influence self-perception of health. There are, obviously, benefits and disadvantages to the use of self-perceived health variables. Among the benefits is the fact that such a broad definition of overall health will encompass all aspects of an individual's health – both physical and mental.

Another benefit is that individuals may not necessarily have been diagnosed with any particular health problem, though may still not feel in the best health. I believe that looking at medical records or the presence of a diagnosed illness is a very matter-of-fact way of looking at the issue, and that there are a greater range of factors at play in determining how healthy someone feels, beyond looking only at whether or not they have been diagnosed with an illness.

Downsides of the use of these variables include the potential for response-bias and the somewhat ambiguous definitions of categories. For example, one person may rate their health as “good”, meanwhile another participant with approximately the same standard of physical and mental health might report their health as “very good”. That is to say, there is no concrete definition for what would constitute “good” health vs “very good” health. The fact that how each individual chooses to interpret these categories may differ is something which will be important to bear in mind throughout the analysis of the results of this research.

Given the above, I formulate the following central research aim:

- **What are the most significant factors that determine self-perceived health in the UK?**

In order to answer this question, I make use of linear and non-linear binary and multinomial outcome models. I review my findings in the context of the current literature and discuss potential policy implications. In better understanding those factors which determine perceived healthcare in the UK, policy makers will be better able to shape policy such that those things which promote a positive perception of health are encouraged, while those which negatively impact health perception are discouraged. This research takes the view that while understanding the “true” state of someone’s health is important, equally as important is understanding how people perceive their own health.

Section 2 provides a review of the current literature in the area of self-perceived health, as well as the determinants of health more generally. In section 3 the methodology used in order to investigate the research question is explained in detail, and the models are formally presented, alongside the motivation for using them. Section 4 then provides a description of the data and variables used. This includes providing summary statistics. Section 5 contains the tabulated results estimated by the models, and these results are analysed in detail. Following on from section 5, section 6 is concerned with the policy implications of the results, and in particular there is a focus on policy concerning smoking as well as urban living, education, and employment. An overall conclusion of this research, including its limitations, is provided in section 7.

2. Literature Review

I begin by summarising current thinking on the topic of health determinants, as well as self-reported health, focussing primarily on literature which makes use of self-reported health variables.

Shields and Shooshtari (2001) explored the determinants of self-perceived health in Canada. They investigated a range of factors including socioeconomic status, physical activity, lifestyle choices and pyscho-social factors. They made a distinction between those variables which affected perceptions at just one end of the health-scale, and those variables which could influence positive *and* negative perceptions; the authors referred to these variables as double-risk factors. The authors found that the double-risk factors for women and men differed. Double-risk factors for men were associated only with their physical health, whereas for women there were a range of non-physical factors. These factors included a low household income, low self-esteem, and high stress levels. It was also found that women who reported that they drink on a weekly basis were more likely to report that their health was either very good or excellent. The authors also found that for men and women, the diagnosis of a chronic health condition was associated with a lower standard of self-perceived health.

Shields (2008) built on the work contained within Shields and Shooshtari (2001) and investigated how self-perceived health is affected by a sense of belonging in the community in Canada. It is found that there is a statistically significant relationship between feeling a sense of belonging in the community, and physical and mental health. The paper highlighted that when asking people to rate their general health, there are psychological factors at play which will influence perception. For example, those with depression may be in excellent physical health with no serious illness, but their mental health will weigh a considerable amount on how they perceive their health. This is particularly pertinent to the research aim of this paper, as it will be important to remember that there is a significant behavioural and psychological aspect to consider when asking people to rate their health. Barnett and Gotlib (1988) found that those with smaller social networks, inadequate support and fewer close relationships were more likely to suffer from depressive symptoms.

When thinking about the issues that may arise as a result of using self-reported variables, there is a wealth of literature which has looked at this in the context of self-reported health. Response bias is one of the main issues which is discussed. Jurges (2007) and Groot (2000) both highlighted the issue of response bias when working with a self-reported health variable. Jurges approached this through exploring how self-reported health differed when looking across a range of different countries. As well as response bias, Groot discussed the issue that adaptability can pose when trying to answer questions using self-reported measures. He noted that people are able to adapt incredibly well to illness, and that people with illnesses are often observed to rate their health much higher than we would expect for people in their position, taking for example people with chronic illnesses. The research carried out by Groot also highlighted the differential in self-perceived health between younger people and older people, and the author proposed a model whereby the cut-off points of a probit model were allowed to alter depending on age.

Marmot et. al (2012) took a similar international approach to that of Jurges (2007) and investigated the determinants of equity of healthcare outcomes across WHO countries, although it is important to note here that actual outcomes were used rather than self-perceived outcomes. The study considered a range of macroeconomic and microeconomic factors. One interesting area discussed was in relation to the working age population, and the kind of work done by an individual. Working conditions and the nature of the work done will influence healthcare outcomes. Although, more interestingly the report highlighted those who are in precarious employment or may even be unemployed. Gender also fed into the discussion, with the study concluding that women are more likely to experience stress as a result of balancing paid work, housekeeping and raising a child. Whereas the review found that men's health was more likely to be affected by work conditions. It is therefore clear from this study that gender norms can certainly play role in determining the healthcare outcomes for men and women.

Dorota et. al (2006) focussed their research on physical activity and its interaction with self-perceived health. They carried out a study on a sample of 598 respondents living in Łódź. Among their findings was a positive relationship between physical activity and self-perceived health. Although, physical activity arising in the form of a person's occupation or housework was not found to have any significant effect on self-perceived health.

Similarly, Piko (2000) investigated the predictors of self-perceived health within the student population of Szeged, Hungary, with a particular focus on the importance of physical activity. The author concluded that physical activity was an important predictor of self-perceived health for those who were more physically active, but not for those individuals who were less physically active. Instead, the author concludes that psychological wellbeing plays a more important role in determining self-perceived health.

Nielsen & Krasnik (2010) investigated how the self-perceived health of migrants and ethnic minorities differed from that of the rest of the population in Europe. The authors concluded that migrant and ethnic minority groups displayed a lower level of self-perceived health than the rest of the population. In concluding, the authors advocated policies which improve access to healthcare for those migrants and ethnic minority groups such that ethnic health inequalities in Europe are reduced.

Crossley and Kennedy (2002) investigated more broadly how reliable self-assessed health status is. They used data from the Australian Bureau of Statistics 1995 National Health Survey, conducted between January 1995 and January 1996. They found that among the respondents who were asked to self-report their health a first time, followed by a second time, 28% changed their response, however only 3% of respondents changed their response by more than one category. This is an interesting finding, as it gives some empirical evidence in relation to the reliability of a self-assessed health variable such as the one being used in this dissertation. The result suggests that although over one-quarter of respondents changed their answer when asked a second time, only a very small proportion of people changed their answer drastically (by more than one category). For the statistical analysis, the authors made use of linear probability models, as well as ordinary least squares (OLS) estimation with instrumental variables (IVs).

3. Methodology

In answering my research question, I made use of a range of econometric techniques. The literature is largely in agreement when considering the standard techniques which should be used for addressing research similar to that of this dissertation. Dorota (2006), Jurges (2007) and Groot (2000), all discussed in section 2, employed the use of an ordered-probit model for their analysis. This paper makes use of this type of model, as well as some analysis using OLS estimation of a linear probability model (LPM).

Throughout the analysis, the dependent variable measures health satisfaction on a Likert-Type scale ranging from 1 to 5, as follows:

1	2	3	4	5
Poor	Fair	Good	Very Good	Excellent

Although, it is described in the following section how I adapted this variable initially such that it is a binary outcome variable.

3.1 A Linear Approach

I began the analysis with a linear probability model (LPM). A model such as this provides a means of carrying out some initial, relatively simple analysis of the data. For this method, it was necessary to transform the ordered-scale dependent variable to a binary variable. The original variable was therefore transformed to a new variable, ***GoodHealth***, as follows:

HealthScale	Poor	Fair	Good	Very Good	Excellent
GoodHealth	0			1	

I then took a model of the following form:

$$Y_i^* = \boldsymbol{X}_i'\beta + \varepsilon_i$$

Here, Y_i^* is a latent (unobservable) variable, $\boldsymbol{X}$ is a column vector of covariates, β are coefficients associated with each regressor $\boldsymbol{X_i}$, ε_i is the error term.

Y_i^* can also be interpreted as the probability of the event:

$$Prob(Y_i = 1|\boldsymbol{X}_i) = F(\boldsymbol{X}, \beta)$$

F is the cumulative distribution function (CDF), and in the case of the LPM using OLS a uniform distribution over interval [0,1] is used. It is therefore assumed that the probability increases linearly between 0 and 1.

The value of Y_i^* depends on the regressors. $\boldsymbol{X_i'}$ is the vector of regressors in the model. β measures the change in the probability of success for a one-unit change in the explanatory variable:

$$\Delta P(Y_i = 1|X_i) = \beta \Delta X_i$$

Note here that Y_i^* is unobservable, and only Y_i is observed. In particular:

$$Y_i = 1 \; if \; Y_i^* \geq 0$$

In this case, good health is observed when the propensity to be in good health is above a particular threshold. If not above this threshold it is observed that $Y_i = 0$.

This model is reasonably simplistic to implement and interpret, though it is not without issue. One of the main issues with the LPM is that probabilities predicted by the model can be outside of the range [0,1]. Obviously, there is no meaningful interpretation when this is the case. There is also the assumption that the probability increases linearly over a uniform distribution. This is likely to not be the case, and so I therefore adapted the model and added some further levels of sophistication in order to draw more robust conclusions.

In addition to this, in considering the Gauss-Markov assumptions necessary for OLS estimators to be a Best Linear Unbiased Estimator (BLUE), there are assumptions in place on variables captured by the error term, ε_i:

1) That these variables are uncorrelated

2) That they have zero mean. i.e., $\mathbb{E}[\varepsilon_i] = 0$

3) That errors are homoscedastic

It can be shown that the LPM necessarily violates assumption 3, and therefore it can be concluded that an LPM is neither efficient, nor BLUE.

The LPM also limits the model to the use of a binary dependent variable, where it is preferable to be able to use a ranked scale in order to make more nuanced analysis.

3.2 A Non-Linear Approach

Still using the binary outcome variable, I next adapted the model used in 3.1 and moved to the use of non-linear models. I used the logit and probit models for non-linear estimation. One of the benefits of using logit/probit is that these functions are bounded between 0 and 1. It could be argued that this issue is not too bothersome to me, seeing as I am more interested in causal inference as opposed to prediction. Nonetheless, these non-linear methods are widely used in the literature, and I felt it would be interesting and beneficial to compare the results gained by using the different estimation techniques.

For the probit model, the CDF for the standard normal cumulative distribution is used:

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \phi(v)\, dv$$

Where ϕ is the probability density function (PDF):

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

And for the logit model, $F(x)$ is the logistic function:

$$F(x) = \frac{e^x}{1 + e^x} = \Lambda(x)$$

Using an LPM, it is possible and also straightforward to interpret the coefficients directly from running the regression. This is possible as in a linear model:

$$Y_i = \boldsymbol{X_i'}\beta + u_i$$

And the marginal effect can therefore be found by:

$$\frac{\partial E[Y_i|\boldsymbol{X}_i]}{\partial \boldsymbol{X}_i} = \frac{\partial \boldsymbol{X_i'}\beta}{\partial \boldsymbol{X}_i} = \beta$$

In other words, the marginal effect is nothing more than the coefficient outputted when running the regression.

However, this is not the case for a non-linear model. Without marginal effects, only the sign of each coefficient can be interpreted – not the magnitude – so it is therefore necessary to calculate the marginal effects for each of the independent variables.

There are, of course, many forms of marginal effect available. These include the average marginal effect (AME), marginal effect at the mean (MEM) and marginal effects at representative values (MER).

Results analysed in this paper report the average marginal effect (AME), which can be expressed as follows:

$$\frac{1}{n}\sum_{i=1}^{n} \beta_{x_1} f(\beta x_i)$$

It can be seen that the AME gives the average effect of a given variable, x_1, on the probability of observing each of the dependent outcomes. For the binary outcome model, it gives the average effect of x_1 on $P(Y = 1)$.

3.3 An Ordinal Model

I continued by forming an ordered probit model, whereby I am not limited to a binary dependent variable. Instead, I used a ranking of health satisfaction on a Likert-type scale, as described at the beginning of this section of the paper. Again, I took a model of the form:

$$Y_i^* = \boldsymbol{X}_i'\beta + \varepsilon_i$$

For a model of this type, it is necessary to estimate β as usual. Additionally, it is now necessary to estimate the cut-off points, μ_i.

Using an ordered scale, the following is observed:

$$HealthScale = \begin{cases} 1, & Y_i^* \geq 0 \\ 2, & 0 < Y_i^* \leq \mu_1 \\ & \quad \vdots \\ 5, & Y_i^* \geq \mu_4 \end{cases}$$

This model estimates:

$$Prob(Y_i = j|\boldsymbol{X}_i) = F(\boldsymbol{X}, \beta)$$

That is, it estimates the probability that $Y_i = j$, where j is one of the possible outcomes on the ordered scale of health outcomes.

4. Data

In this section, I describe the data used for the econometric work, and provide a description of the variables included in the analysis. Data used in this research is from wave 9 (2017-18) of the UK Household Longitudinal Survey (University of Essex et al., 2020).

Table 1, below, provides descriptive statistics of the continuous variables used throughout the analysis.

Table 1

Descriptive Statistics of Continuous Variables

Variable	Mean	S. Dev.	Skew	Kurtosis
Net Income (£100s)	13.45	8.25	0.32	-0.46
Age	49.21	18.87	0.03	-0.93
n = 30768				

It can be seen from table 1 that the average net income in the sample is £1345, with a standard deviation of £825. Table 1 also shows that the age of the average respondent was just over 49 years, with a standard deviation of just under 19 years. For graphical representations of these distributions, see Appendix A.

Looking now at the dependent variable, Table 2, below, shows the proportion of responses which were reported into each of the possible outcomes on the Likert-type response scale.

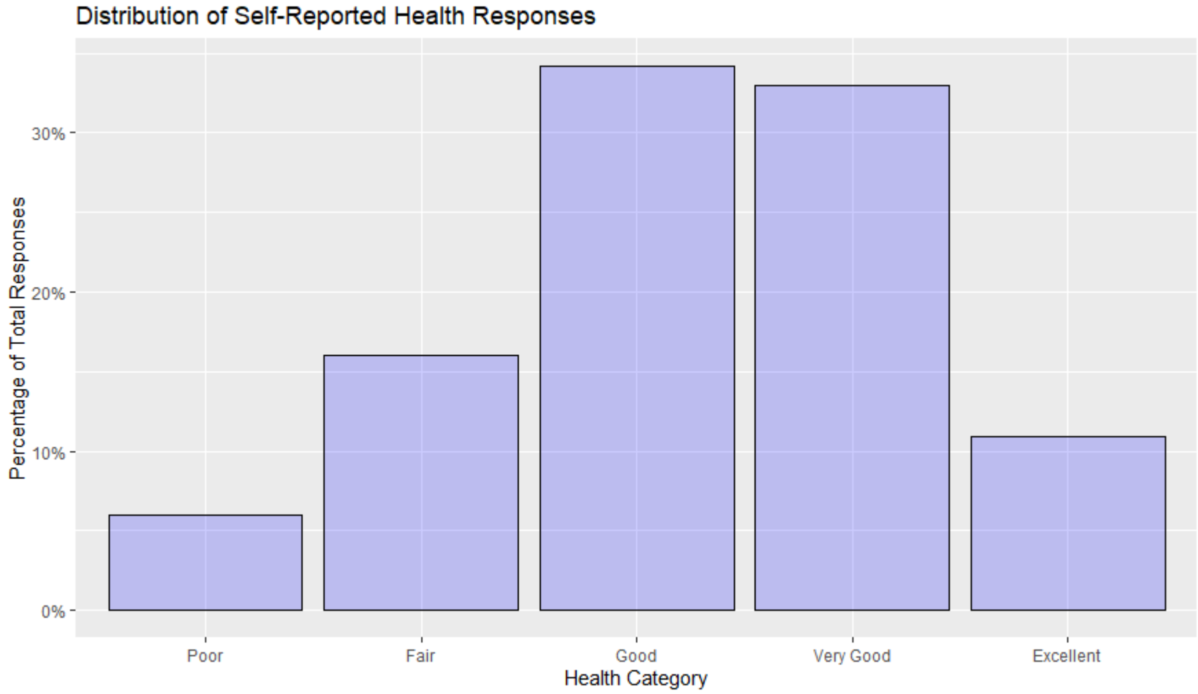
Table 2

Self-Reported Health: Category Proportions

Variable	Poor	Fair	Good	Very Good	Excellent
Count	1830	4923	10520	10136	3359
% of Total	6%	16%	34%	33%	11%
n = 30768					

Figure 1, below, graphically illustrates the distribution of responses.

Figure 1



I found that the majority of responses (34%) were reported into the “Good” category, closely followed by the “Very Good” category, which contained 33% of all responses. The “Poor” category accounted for just 6% of responses, while 16% of all respondents reported themselves being in “Fair” health. This means that in total, 22% of all respondents reported their level of health being below a “Good” standard.

Looking next at the binary covariates used in the analysis, the covariates cover a broad range of socio-economic factors, and lifestyle choices. This was to allow me to draw some interesting conclusions from the model, as well as provide a range of points of discussion when it came to addressing the aim of this research.

Table 3, below, contains the mean value of these variables

Table 3

Binary Variable Averages

Variable	Mean
Degree	0.38
Female	0.57
Employed	0.56
Smokes	0.14
Long-Term Illness	0.37
Urban	0.75
Married	0.52
Drinks Alcohol	0.77
BAME	0.16
n = 30768	

These averages are helpful for understanding the proportion of observations in the data for which the value of each variable value equals 1 versus 0. For example, it can be seen from Table 3 that 38% of respondents reported as having a degree level qualification. It also shows that 57% of respondents were female, and 75% of all respondents reported living in an urban area.

5. Results & Analysis

The results gained from estimating the LPM, probit and logit models are presented in Table 4, below.

Table 4

LPM & Logit/Probit Regression Results

GoodHealth	**LPM**	**Probit AMEs**	**Logit AMEs**
Net Income (£100s)	0.000’ (0.000)	0.001** (0.000)	0.001** (0.000)
Age	-0.002*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Degree	0.043*** (0.004)	0.044*** (0.005)	0.046*** (0.005)
Female	0.002 (0.004)	0.004 (0.004)	0.004 (0.004)
Employed	0.055*** (0.005)	0.046*** (0.005)	0.049*** (0.005)
Smokes	-0.107*** (0.007)	-0.093*** (0.005)	-0.092*** (0.005)
Long-Term Illness	-0.357*** (0.005)	-0.283*** (0.003)	-0.283*** (0.003)
Urban	-0.026*** (0.005)	-0.026*** (0.005)	-0.025*** (0.005)
Married	0.023*** (0.005)	0.019*** (0.004)	0.019*** (0.004)
Drinks Alcohol	0.083*** (0.006)	0.066*** (0.005)	0.066*** (0.005)
BAME	-0.004 (0.007)	-0.016* (0.006)	-0.017* (0.006)

$* \, p < 0.05$, $** \, p < 0.01$, $*** \, p < 0.001$

I found that the estimates produced by the LPM, and those produced using the logit/probit models are largely similar. Interestingly, none of the models were able to estimate a statistically significant gender differential when considering self-perceived health.

The results indicated that existing long-term illness was the most influential factor in determining self-perceived health. the LPM model predicted that the existence of a long-term health condition is associated with a 36% reduction in the probability that a person will report being in good health. Both the probit and logit models estimated a 28% reduction in the probability of being in good health, where an individual had a long-term illness. This result is in-line with what was found in the literature. For example, Shields and Shooshtari (2001), previously discussed in section 2, also found that a chronic/long-term illness had a negative impact on self-reported health.

One of the more interesting results from the models, was the effect of living in an urban area. Results from all models concluded that living in an urban area reduced the probability of being in good health by approximately 2.5%. This is perhaps not unexpected. Those living in more built-up, urban areas are likely to be in those areas which have poorer air quality, and higher overall levels of air pollution**.** Bai et al. (2012) explored how health and wellbeing are impacted by living in an urban environment. They highlighted air pollution, as well as a range of other urban hazards. These hazards included crowded living conditions, poor sanitation, inadequate waste disposal, and traffic congestion.

One result of particular relevance in current times, was how the probability of being in good health changes purely as a result of being from a BAME background. There has been much discussion recently regarding how Covid-19 has disproportionately affected BAME individuals, and what that implies about a wider societal problem, in terms of socio-economic factors, for example. Bentley (2020) discussed the ethnic and structural inequalities in relation to the susceptibility of COVID-19. My LPM did not find that coming from a BAME background had a statistically significant effect on self-perceived health. However, the probit and logit models did find a statistically significant relationship, predicting that the probability of reporting to be in good health reduced by 1.6% and 1.7%, respectively, for those from a BAME background.

Other factors which were associated with worse predicted outcomes were smoking and increasing age. These results were not unexpected. Adverse health risks which arise as a result of smoking are well known and documented in the literature (Bartecchi et al,. 1994). There is also a wide array of literature documenting health problems which present themselves with older age (Paul et al., 2006).

Looking at those factors which lead to more positive outcomes, the models found that higher income, having a degree, being a woman, having a job, and being married are all factors which increased the probability of a person reporting as being in good health. In addition to this, drinking alcohol was shown to increase the probability of being in good health. This is unexpected, and a result which is explored in greater detail at a later stage.

All models found that having a degree increased the probability of being in good health by approximately 4% - 5%, which concurs with much of the literature linking higher levels of education to increased levels of health. Being in employment was predicted by the LPM to increase the probability of being in good health by approximately 6%. The logit and probit models both predicted a 5% increase. I suspect that mental health plays a role here, as those who are unemployed may feel as though they do not have any purpose, and this feeling of no sense of purpose may manifest itself as a range of mental illnesses, such as anxiety and depression.

Ross and Wu (1995) investigated the relationship between education and health. In explaining the education-health relationship, they considered economic standing, working conditions, social-psychological factors, and health lifestyle. They concluded that those who are more highly educated are more likely to be employed. In addition, they summarised by discussing how a greater level of education allows for an individual to boost the control they have over their life. Those who are well educated are also more likely to form healthy relationships with others and have a healthier lifestyle. For example, those better educated may better understand the risks of smoking and drinking and may therefore act in a way such that they lead healthier lives.

The binary variable models have already been helpful in giving an indication as to how the covariates impact self-perceived health. The results from estimating the ordered probit model are now presented in Table 5, below, and allow me to undertake a more nuanced analysis of the factors which alter how people perceive their health.

Table 5

Ordered Probit Marginal Effects

HealthScale	**Poor**	**Fair**	**Good**	**Very Good**	**Excellent**
Net Income (£100s)	0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	0.001*** (0.000)	0.000*** (0.000)
Age	0.000*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)
Degree	-0.010*** (0.001)	-0.035*** (0.002)	-0.034*** (0.003)	0.056*** (0.004)	0.023*** (0.002)
Female	0.000 (0.001)	-0.002 (0.002)	-0.001 (0.002)	0.002 (0.004)	0.001 (0.001)
Employed	-0.003*** (0.001)	-0.011*** (0.003)	-0.010*** (0.003)	0.017*** (0.005)	0.007*** (0.002)
Smokes	0.030*** (0.002)	0.090*** (0.004)	0.040*** (0.001)	-0.121*** (0.005)	-0.039*** (0.001)
Long-Term Illness	0.093*** (0.002)	0.241*** (0.004)	0.090*** (0.003)	-0.304*** (0.004)	-0.119*** (0.002)
Urban	0.006*** (0.001)	0.020*** (0.003)	0.019*** (0.003)	-0.032*** (0.004)	-0.013*** (0.002)
Married	-0.002** (0.001)	-0.008** (0.002)	-0.007** (0.002)	0.012** (0.004)	0.005** (0.002)
Drinks Alcohol	-0.013*** (0.001)	-0.041*** (0.004)	-0.030*** (0.002)	0.061*** (0.005)	0.022*** (0.002)
BAME	0.005*** (0.001)	0.017*** (0.004)	0.014*** (0.003)	-0.026*** (0.006)	-0.010*** (0.002)

***** $\boldsymbol{p < 0.05}$, ****** $\boldsymbol{p < 0.01}$, ******* $\boldsymbol{p < 0.001}$

The results indicated that having a long-term illness has a substantial effect on the probability of reporting into many of the categories. My results showed that those with a long-term illness were 30% less likely to report that they were in very good health, and 9% more likely to report being in poor health.

Again, the results provided no evidence of any gender differential which exists for self-perceived health status, as the estimates produced by the model are statistically insignificant.

I found, unsurprisingly, that those who smoke were more likely to report a poorer level of health than those who do not. The results showed that smoking reduces the probability of reporting very good health by 12%. I also found that those who smoke were about 4% less likely to report that they have excellent health. My results indicated that smoking increases the probability of reporting into the remaining categories (poor, fair and good).

Interestingly, the results showed that an additional £100 of income has a negligible effect on the probability of a person into any one of the health categories. Exploring further the effect of drinking alcohol, drinking reduced the probability that an individual perceived their health as being poor, fair, or good, and increased the probability that they will perceive their health as being either very good or excellent. This is interesting, as it is well known that drinking alcohol (particularly in excess) can give rise to a wide range of health problems. These health risks include liver problems, increased risk of cardiovascular disease, the presentation of mental health issues, and gastrointestinal health problems (Thakker, 1998).

Given that many people do drink responsibly, or only in social situations, it may be that those who drink socially and have active and fulfilling social lives feel happier than those who do not. They may be less likely to develop mental health problems, and perhaps actually feel healthier overall, as a result of their social lives. Thakker (1998) explores the benefits of alcohol consumption and notes that alcohol consumption has been observed to have psychological benefits through the reduction of feelings of tension, stress, depression, and other mental illnesses.

Another explanation for the result could be that those who drink excessively, to the point where their level of alcohol consumption has an adverse effect on their level of health may be more likely to be in denial about health problems arising from alcohol consumption or may not even realise that their level of alcohol consumption has had any impact on their health. It is also important to keep in consideration that there will inevitably be a differential which exists between the perceived effect of alcohol on a person's health, and the actual impact it is having on their health.

These results also showed that coming from a BAME background increased the probability of reporting poor, fair, or good health. Meanwhile reducing the probability that an individual would report as being in very good or excellent health. Specifically, I found that those from a BAME background were 1% less likely to report being in excellent health, and 2.6% less likely to report being in very good health. I found that BAME individuals were 0.5% more likely to report being in poor health, 1.7% more likely to report a fair level of health, and 1.4% more likely to report being in good health.

From the results and analysis, I was able to rank the impact of each factor as shown in table 6, below:

Table 6

Ranking of determinants of perceived health

Positive Impact	***Negative Impact***
1) Alcohol	1) Long-Term Illness
2) Employed	2) Smokes
3) Degree	3) Urban
4) Married	4) BAME
5) Income	5) Age

In terms of those factors which were not included in this research, the lack of any variable measuring physical activity is the most striking omission. This was as a result of being unable to find a suitable variable.

6. Policy Implications

I will now discuss what my results suggest in terms of policy implications. It was observed in the results that having a long-term illness was the factor which had the largest negative impact on self-reported health. This is a difficult factor to target effectively with policy, as the long-term illness category encompasses such a wide range of healthcare problems. Many of these problems may arise as a result of lifestyle choices which have an adverse effect on a person's health, however many (such as cancer) are genetic health issues for which there is little that policy intervention is likely to be able to do. It makes sense, therefore, to focus on those factors which can feasibly be tackled with targeted policy intervention.

The results showed that smoking was one of these lifestyle factors included in the analysis which had an adverse effect on self-assessed health. I believe smoking to be a particularly important issue to tackle, as it affects not only those who actively choose to smoke, but also those around them. It is well known that goods of this nature are referred to as demerit goods. We can assess the impact of smoking through the lens of externalities.

Figure 2

Negative Externality in Consumption

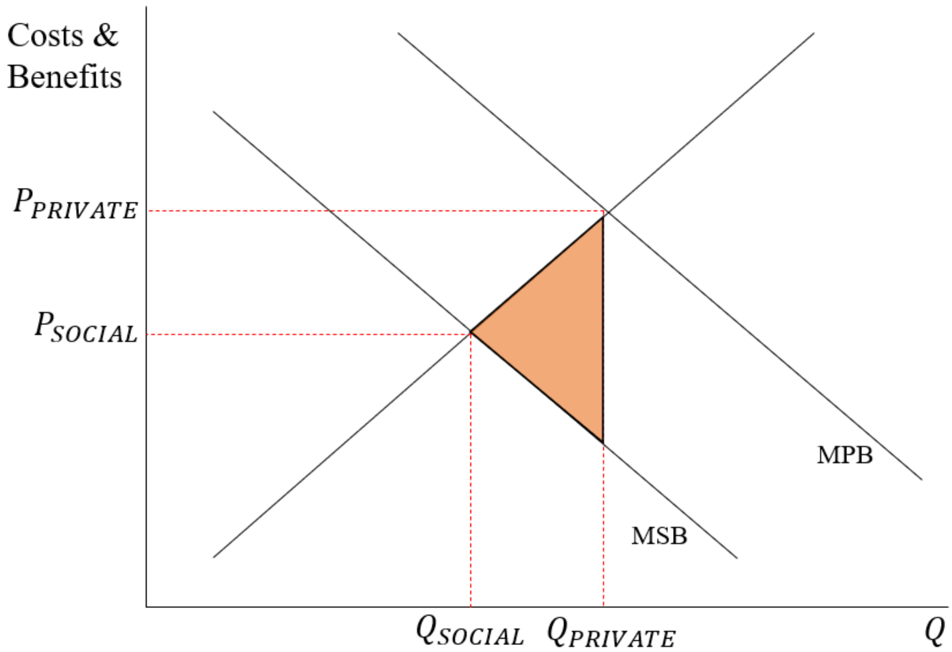


Figure 2 illustrates why consumption of these demerit goods, such as cigarettes, is such an issue and why it should be discouraged as far as possible, ideally with the help of policy intervention. It can be seen from Figure 2, that the marginal private benefit of consuming cigarettes is greater than the benefit to society. This is represented by the quantity:

$$Q_{PRIVATE} - Q_{SOCIAL}$$

The effect of this overconsumption can also be seen by the shaded area in figure 1, which represents the deadweight welfare loss to society. This negative externality from smoking arises as the consumption of cigarettes does not only harm the health of those who have made the conscious decision to smoke, but also those around them, as a result of passive smoking.

There has already been significant policy intervention in an attempt reduce the social cost of smoking. In 2007, the UK government introduced the smoking ban (Health Act 2006, c 28) which made it illegal to smoke in a number of places, including the workplace, pubs, restaurants, and nightclubs.

Given this significant policy, which is already in place, I consider that a more effective future intervention would be to ensure there is easy access to healthcare services (such as those provided by the NHS) for those who want to quit smoking. This is, of course, only helpful for those who have already chosen to start smoking. An even more effective policy could be to improve education in schools such that young people are properly informed about the risk of smoking. It is already mandatory in UK schools that for children aged 11 and older, sex education must be provided, and schools are required to have a written policy on sex education. A similar initiative targeting smoking could be hugely beneficial in terms of deterring young people from developing a smoking addiction in the first place. This would improve their own individual outcomes as well as those outcomes for society as a whole, due to the elimination of negative externalities which arise from smoking.

The results also showed that urban living had an adverse effect on self-assessed health. Continuing population growth and a lack of housing has meant that an increasing number of residential properties are being built in cities. It is already well established in the literature, that there are a number of health risks associated with urban living (Bai et al., 2012). The question for policy makers revolves around which of those hazards can be effectively targeted by policy, and precisely how a policy can be designed such that the health risks of living in urban areas can be mitigated. Among those hazards which exist as a result of urban living are high levels of air pollution, crowded living conditions, poor sanitation, inadequate waste disposal and traffic congestion.

Arguably, poor sanitation and crowded living conditions go hand-in-hand. The main risk which arises as a result of crowded living conditions is that disease and illness is able to spread more easily. This has been a particular issue in recent times, in the context of the Covid-19 pandemic. Ensuring high levels of sanitation would reduce the risk of disease and illness in the first instance, and as a result would lower the additional health risk which arises out of crowded living conditions.

Traffic congestion is known to lower the quality of air substantially, and in many areas, emissions arising as a result of traffic congestion have become the primary source of pollution. Pollutants released include carbon monoxide, carbon dioxide, nitrogen oxide and particulates. In the UK, we have already seen some cities establishing so-called “clean air zones”, in addition to London’s congestion charge and ultra-low emission zone (ULEZ). The idea of these policies is that people are charged for driving within certain areas of urban cities, which acts as a disincentive to drive, and hence lowering the levels of vehicle pollution. Leape (2006) shows that following the introduction of the congestion charge in London, there was a 34% drop in the number of kilometres driven by vehicles within the zone, as well as showing the largest reduction in time travelled in the zone, was for those vehicles traveling at a speed less than 10km/h. Therefore, it seems sensible that the UK government puts disincentives of this nature in place, in order to maintain an acceptable quality of air in cities.

I have so far focussed on those factors which were shown to have a negative impact on self-perceived health, and so I will not provide some discussion of these factors which were shown to have a positive impact. These are the factors for which policy should be designed such that they are encouraged. Focussing on employment and education, these things are obviously linked as those who hold higher level academic qualifications will find it easier to gain employment. Not least as a result of having the level of qualification such that they have the training required for a greater number of vacancies which exist in the job market at any given time. So therefore, I suggest that policy in this area should do two things:

1. Encourage higher levels of education

2. Provide opportunities for lower-skilled adults

Firstly, focussing on encouraging higher levels of education, I do not mean only encouraging the progression from GCSEs onto A-levels and then university. Although this is certainly a very popular and credible education progression pathway, there are other alternatives. Apprenticeships, for example, are actually a much better option for certain individuals. However, I do not believe that this is presented as a realistic progression option for many school leavers in the UK, and that in many educational settings the more traditional route to university is pushed heavily whereas alternative such as apprenticeships are largely forgotten.

Providing retraining opportunities for lower-skilled adults is another policy initiative which I believe would be hugely beneficial for boosting levels of employment and self-perceived health. Many adults many be in professions which are becoming outdated; for example, those in industries which are being increasingly taken over by artificial intelligence. The policy challenge here is to provide retraining opportunities at an affordable price. For both of the potential policy ideas which have been outlined above, it is important to appreciate how education and employment are both linked, and how they both work together in influencing self-perceived health. Encouraging a high level of education in the population, as well as providing opportunities for individuals to retrain, will allow for higher levels of employment in the economy. Our results suggest that, in turn, this should lead to a more positive perception of health.

7. Conclusion

In this research, I investigated self-perceived health through the use of an econometric model. In concluding, I shall give a summary of the main findings as well as provide some discussion on the limitations of this research, and aspects of this topic which could be explored further in future work.

Considering those factors which had a negative impact on perceived health, this research found that a long-term existing illness was the factor most likely to impact negatively on self-reported health. Smoking, living in an urban area, being from a BAME background and increased age were all other factors which had an adverse effect on an individual’s perception of their health.

This research also found that being in employment, having a university degree, being married, and higher levels of income were all, unsurprisingly, factors which increased the probability of having a positive perception of one's own health. Although, perhaps the most striking result of this research is that alcohol consumption was found to be the factor which had the greatest positive impact on self-perceived health. It is important to note that the measure of alcohol consumption used in this study was not interested in the amount of alcohol individuals had consumed, but simply whether or not they had consumed any alcohol at all. In that sense, I think it is fair to conclude that this is not a particularly helpful finding.

For future work, it would be interesting to further explore the effect of alcohol consumption on self-reported health, where the amount consumed is considered. It would also be interesting to see how self-perceived health differs for different age groups. I think there is also scope here to further develop on the econometric models used in this analysis. This could be done through conducting research using panel data rather than cross sectional data. This would allow for the analysis of changes in self-perceived health over time. It would be interesting too, to explore further how the ranking of determinants differs for different age groups.

As well as providing a thorough analysis of results, this paper has discussed the relevant policy which exists, or could be introduced, to promote positive health perception. In particular, policy in relation to smoking and air pollution was discussed in some detail. I suggested that education around smoking should be prioritised such that it is seen as important an issue as sex education. In relation to mitigating the health issues which arise as a result of urban living, I believe that there is strong evidence to suggest that implementation of clean air zones in cities is an effective way of reducing the risks posed by air pollution. Ensuring high levels of sanitation in those areas which are most densely populated is also a suggested policy idea, as this would have the knock-on effect of reducing the risk posed by crowded living conditions. The important interplay between education and employment was also highlighted in the context of designing policy which should ultimately seek to promote positive perceptions of health by individuals.

References

Bai, X., Nath, I., Capon, A., Hasan, N., Jaron, D. (2012). Health and wellbeing in the changing urban environment: complex challenges, scientific responses, and the way forward. *Current Opinion in Environmental Stability, 4*, 465-472. http://dx.doi.org/10.1016/j.cosust.2012.09.009

Barnett, P. A., & Gotlib, I. H. (1988). Psychosocial functioning and depression: distinguishing among antecedents, concomitants, and consequences. *Psychological bulletin*, *104*(1), 97–126. https://doi.org/10.1037/0033-2909.104.1.97

Bentley, G. R. (2020). Don’t blame the BAME: Ethnic and structural inequalities in susceptibilities to COVID-19. *American Journal of Human Biology, 5*(32). https://doi.org/10.1002/ajhb.23478

Bartecci, C. E., MacKenzie, T. D., Schrier, R. W. (1994). The human costs of tobacco use. *The New England Journal of Medicine, 330*(13), 907-912. https://doi.org/10.1056/nejm199403313301307

Crossley, T. F., & Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*. https://doi.org/10.1016/S0167-6296(02)00007-3

Dorota, K., Makowiec-Dąbrowska, T., Dziankowska-Zaborszczyk, E., Jegier, A. (2006). Physical Activity and Self Perceived Health Status. *International Journal of Occupational Medicine and Environmental Health, 19*(1), 61-69. https://doi.org/10.2478/v10001-006-0005-x

Fischer, J. A. V., Souza-Pouza, A. Does job satisfaction improve the health of workers? New evidence using panel data and objective measures of health. *Health Economics, 18,* 71-89. https://doi.org/10.1002/hec.1341

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*. https://doi.org/10.1016/S0167-6296(99)00037-5

Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics*. https://doi.org/10.1002/hec.1134

Leape, J. (2006). The London Congestion Charge. *Journal of Economic Perspectives, 20*(4), 157-176. https://doi.org/10.1257/jep.20.4.157

Mammal, K. J., & Ramm, E. B. (2008). Alcohol Consumption: Risks and Benefits. *Current Atherosclerosis Report, 10,* 536-543

Marmot, M., Allen, J., Bell, R., Bloomer, E., & Goldblatt, P. (2012). WHO European review of social determinants of health and the health divide. *The Lancet*. https://doi.org/10.1016/S0140-6736(12)61228-8

Nielsen, S. S., & Krasnik, A. (2010). Poorer self-perceived health among migrants and ethnic minorities versus the majority population in Europe: a systematic review. *International Journal of Public Health, 55,* 357-371. https://doi.org/10.1007/s00038-010-0145-4

Paul, C., Ayis, S., Ebrahim, S. (2006). Psychological distress, loneliness and disability in old age. *Psychology, Health & Medicine, 11*(2), 221-232. https://doi.org/10.1080/13548500500262945

Piko, B. (2000). Health-related predictors of self-perceived health in a student population: the importance of physical activity. *Journal of Community Health, 25*(2).

Rice, T. 2013. The behavioural economics of health and health care. *Annual Review of Public Health, 34*, 431-437. https://doi.org/10.1146/annurev-publhealth-031912-114353

Ross, C. E., & Wu, C.-l. (1995). The links between education and health. *American Sociological Review, 60*(5), 719–745. https://doi.org/10.2307/2096319

Shields, M., & Shooshtari, S. (2001). Determinants of self-perceived health. *Health Reports, 13*(1), 35-52.

Shields, M. (2008). Community belonging and self-perceived health. *Health Reports, 19*(2), 1-10.

Thakker, K. D. (1998). An Overview of Health Risks and Benefits of Alcohol Consumption. *Alcoholism: Clinical and Experimental Research*, *22*(7), 285-298. https://doi.org/10.1111/j.1530-0277.1998.tb04381.x

UK Health Act 2006, c. 28. Available at: https://www.legislation.gov.uk/ukpga/2006/28/contents/ (Accessed: 11 April 2021).

University of Essex, Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2020). *Understanding Society: Waves 1-10, 2009-2019 and Harmonised BHPS: Waves 1-18, 1991-2009*. [Data set]. 13th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-14.

Appendix A

Distributions of Age and Net Monthly Income

Figure A1

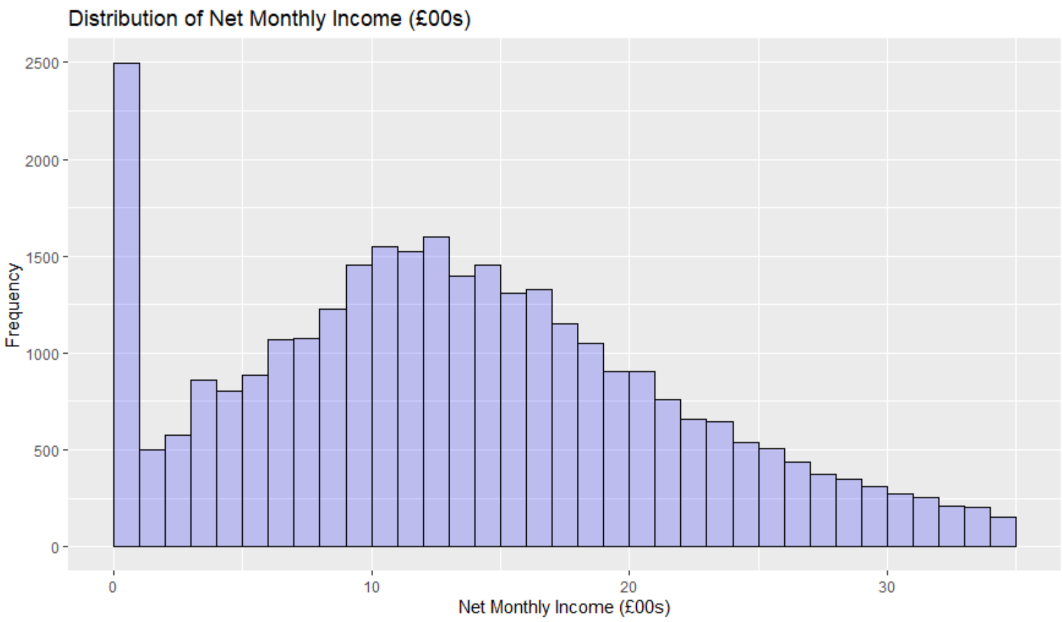


Figure A2

Figure A3

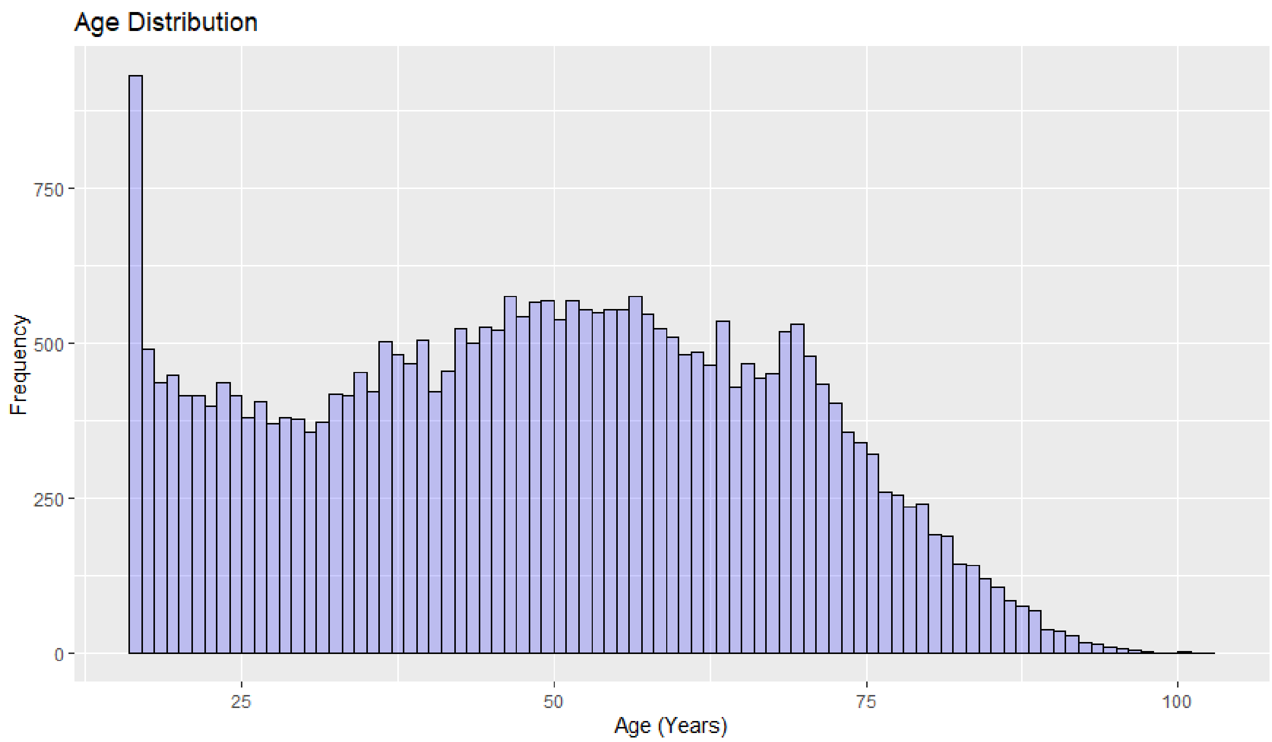
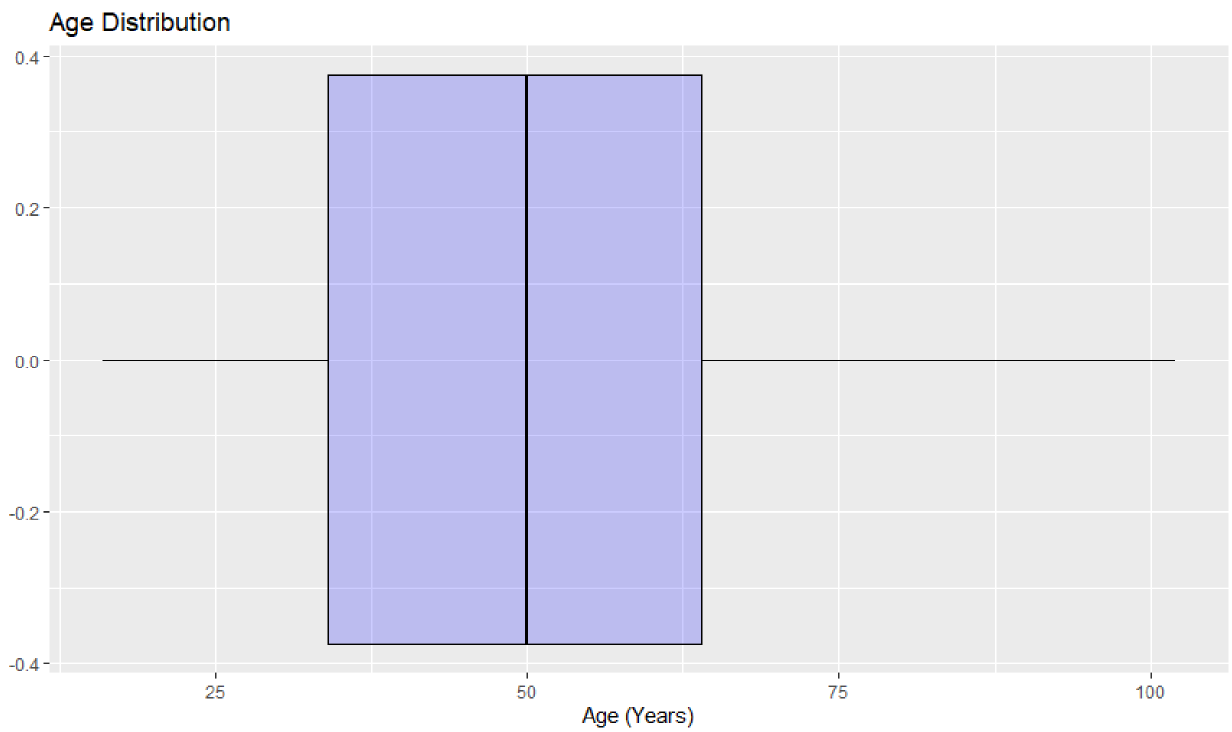


Figure A4



Appendix B

Definition of Variables

Net Income: Net monthly income (£100s).

Age: Age of respondent.

Degree: Binary variable indicating whether the respondent holds a university degree.

Female: Binary Variable indicating whether the respondent is female.

Employed: Binary variable indicating whether the respondent is in employment.

Smokes: Binary variable indicating whether the respondent smokes cigarettes.

Long-Term Illness: Binary variable indicating whether the respondent has a long-term illness or health problem which has been ongoing for 12 months or more.

Urban: Binary variable indicating whether the respondent lives in an urban area.

Married: Binary variable indicating whether or not the respondent is married.

Drinks Alcohol: Binary variable indicating whether or not the respondent has drunk alcohol in the past 12 months.

BAME: Binary variable indicating whether the respondent is from a BAME background.

Appendix C

R Code

```
library(sandwich)

library(clubSandwich)

library(rgr)

library(readxl)

library(writexl)

library(ggplot2)

library(mfx)

library(erer)

library(margins)

library(psych)

library(dplyr)

library(scales)


#Load data


uk_data = read_excel("C://Users/ryanj/Google Drive/University/Year
      3/EC3DIS/Data/Diss_Current.xls")


#recode values <0 as NA


uk_data[uk_data < 0] <- NA


#Recode and set factor variables as appropriate#


#Age


uk_data$age <- uk_data$i_age_dv


#Health Outcome (Full Scale)


uk_data$FShealth = NULL
```

```
uk_data$FShealth[uk_data$i_scsf1 == 5] <- 1

uk_data$FShealth[uk_data$i_scsf1 == 4] <- 2

uk_data$FShealth[uk_data$i_scsf1 == 3] <- 3

uk_data$FShealth[uk_data$i_scsf1 == 2] <- 4

uk_data$FShealth[uk_data$i_scsf1 == 1] <- 5


uk_data$FShealth <- factor(uk_data$FShealth,
                         levels = c(1,2,3,4,5),
                         labels = c("Poor", "Fair", "Good", "Very Good",
      "Excellent"))


#Health Outcome (Binary)


uk_data$GoodHealth = NULL


uk_data$GoodHealth[uk_data$FShealth == "Poor" | uk_data$FShealth == "Fair"
      ] <- 0

uk_data$GoodHealth[uk_data$FShealth == "Good" | uk_data$FShealth == "Very
      Good" | uk_data$FShealth == "Excellent"] <- 1


#qualification


uk_data$qualification = NULL


uk_data$qualification[uk_data$i_hiqual_dv == 1 | uk_data$i_hiqual_dv == 2]
      <- 1

uk_data$qualification[uk_data$i_hiqual_dv == 3] <- 2

uk_data$qualification[uk_data$i_hiqual_dv == 4] <- 3

uk_data$qualification[uk_data$i_hiqual_dv == 5] <- 4

uk_data$qualification[uk_data$i_hiqual_dv > 5] <- NA


uk_data$qualification <- factor(uk_data$qualification,
                             levels = c(1,2,3,4),
                             labels = c("degree", "A-level", "GCSE",
      "None"))


#degree
```

```
uk_data$degree = NULL

uk_data$degree[uk_data$i_hiqual_dv == 1 | uk_data$i_hiqual_dv == 2] <- 1
uk_data$degree[uk_data$i_hiqual_dv > 2] <- 0

#Sex

uk_data$female = NULL

uk_data$female[uk_data$i_sex_dv == 2] <- 1
uk_data$female[uk_data$i_sex_dv == 1] <- 0
uk_data$female[uk_data$i_sex_dv < 1] <- NA

#BAME

uk_data$bame = NULL

uk_data$bame[uk_data$i_ethn_dv > 4] <- 1
uk_data$bame[uk_data$i_ethn_dv < 5] <- 0

#Smoker

uk_data$smokes = NULL

uk_data$smokes[uk_data$i_smoker == 2] <- 0
uk_data$smokes[uk_data$i_smoker == 1] <- 1

#employed

uk_data$employed = NULL

uk_data$employed[uk_data$i_employ == 2] <- 0
uk_data$employed[uk_data$i_employ == 1] <- 1

#Satisfied with Job
```

```
uk_data$jbsat = NULL

uk_data$jbsat[uk_data$i_jbsat > 4] <- 1
uk_data$jbsat[uk_data$i_jbsat < 5] <- 0

#long-term illness

uk_data$illnessLT = NULL

uk_data$illnessLT[uk_data$i_health == 2] <- 0
uk_data$illnessLT[uk_data$i_health == 1] <- 1

#Drink in last 12 months

uk_data$drinks = NULL

uk_data$drinks[uk_data$i_auditc1 == 2] <- 0
uk_data$drinks[uk_data$i_auditc1 == 1] <- 1

#Urban Area

uk_data$urban = NULL

uk_data$urban[uk_data$i_urban_dv == 2] <- 0
uk_data$urban[uk_data$i_urban_dv == 1] <- 1

#married

uk_data$married = NULL

uk_data$married[uk_data$i_marstat != 2] <- 0
uk_data$married[uk_data$i_marstat == 2] <- 1

#create log(wage) variable

uk_data$wage100 <- uk_data$i_fimnnet_dv/100
```

```
uk_data <- subset(uk_data, wage100 < 35)

#Drop all variables not needed for LPM

keepsLPM <- c("wage100","age" , "GoodHealth", "degree", "female", "smokes",
      "employed", "illnessLT", "urban", "married", "drinks", "bame")
lpm_data = uk_data[keepsLPM]

lpm_data <- na.omit(lpm_data)

#Descriptive Statistics

describe(lpm_data)

#Data Visualizations of Continuous Variables

qplot(lpm_data$wage100,
      geom = "histogram",
      breaks = seq(0, 35, by=1),
      main = "Distribution of Net Monthly Income (£00s)",
      xlab = "Net Monthly Income (£00s)",
      ylab = "Frequency",
      fill = I("blue"),
      col = I("black"),
      alpha = I(.2),
      xlim = c(0, 35))

qplot(lpm_data$wage100,
      geom = "boxplot",
      main = "Distribution of Net Monthly Income (£00s)",
      fill = I("blue"),
      col = I("black"),
      alpha = I(.2))

qplot(lpm_data$age,
```

```
      geom = "histogram",
      breaks = seq(16, 103, by=1),
      main = "Age Distribution",
      xlab = "Age (Years)",
      ylab = "Frequency",
      fill = I("blue"),
      col = I("black"),
      alpha = I(.2),
      xlim = c(16, 103))

qplot(lpm_data$age,
      geom = "boxplot",
      main = "Age Distribution",
      xlab = "Age (Years)",
      fill = I("blue"),
      col = I("black"),
      alpha = I(.2))


#Linear Probability Model#

Xi = c("wage100", "age", "degree", "female", "employed", "smokes",
      "illnessLT", "urban", "married", "drinks", "bame")
lpm_outcome <- "GoodHealth"
eqn <- as.formula(paste(lpm_outcome, paste(Xi,  collapse = "+"), sep =
      "~"))

lpm <- lm(eqn, data = lpm_data)
summary(lpm)
coeftest(lpm, vcov = vcovHC(lpm, type="HC1"))

lpm_data$predictLPM <- predict(lpm, newdata = lpm_data)

#Probit Model# (same data and vectors used as for LPM)

probit <- glm(eqn, data = lpm_data, family = binomial(link = "probit"))
summary(probit)
```

```
coeftest(probit, vcov = vcovHC(probit, type="HC1"))
probitMargins <- margins(probit)
summary(probitMargins)

lpm_data$predictProbit <- predict.glm(probit, newdata = lpm_data, type =
      "response")

#Logit Model# (same data and vectors used as for LPM/probit)

logit <- glm(eqn, data = lpm_data, family = binomial(link = "logit"))
summary(logit)
coeftest(logit, vcov = vcovHC(logit, type="HC1"))
logitMargins <- margins(logit)
summary(logitMargins)

lpm_data$predictLogit <- predict.glm(logit, newdata = lpm_data, type =
      "response")

#Data for Ordered Probit Model#

keepsOPM <- c("age" , "FShealth", "degree", "female", "smokes", "employed",
      "illnessLT", "wage100", "urban", "married", "drinks", "bame")
opm_data = uk_data[keepsOPM]
opm_data <- na.omit(opm_data)

#Descriptive Statistics (Full Scale Variable)

ggplot(opm_data, aes(x = FShealth)) + geom_bar(aes(y =
      (..count..)/sum(..count..)), fill="blue", color = "black", alpha =
      0.2) +
  scale_y_continuous(labels=percent) + xlab("Health Category") +
  ggtitle("Distribution of Self-Reported Health Responses") +
      ylab("Percentage of Total Responses")

FStable <- table(opm_data$FShealth)
prop.table(FStable)
round(prop.table(FStable), digits = 2)
```

```
#Ordered Probit

opm_outcome <- "FShealth"
eqn2 <- as.formula(paste(opm_outcome, paste(Xi,  collapse = "+"), sep =
      "~"))

oprobit <- polr(eqn2, data = opm_data, Hess = TRUE)
summary(oprobit)

#compute marginal effects#

ME <- ocME(w = oprobit); oprobit
ME$out
```