100379318 (Group 10 member)

Does more money mean a higher satisfaction of individual health?

Introduction

Initially, it would make sense to agree that an increase in income would allow individuals to satisfy their preferences, as similarly stated by Dolan et al (2007). This could be due to plentiful of reasons: Adler et al. (1994) suggest that through a higher income, individuals will have access to better healthcare and improved living environment as they have fewer financial worries (Frijters et al, 2005). Currie et al (1999) find that people in good health are typically more economically productive and have a higher income. We see that there is some clear evidence that earning more will typically mean that an individual will have relatively good physical health.

Although, we must question whether an increase in income will actually lead to better health or if it is a gimmick to have a better diet with higher quality products. Howlet (2017) found that despite consumers becoming increasingly satiated with goods, it does not signify an increase in their quality of life.. This is agreed by Contoyannis et al. (2004) whereby we may observe a positive trend in health, but this could be wrongly linked to changes in income. We must also question whether an individual is satisfied with their mental health with an increase with higher satisfaction. A higher income suggests more seniority and this could bring on increased pressures to bring results, found similarly by Godin et al (2005). Hence, we must not overlook other variables which could influence an individual's satisfaction on health, by considering factors such as highest qualification, gender, where an individual lives or an individual's sleep time.

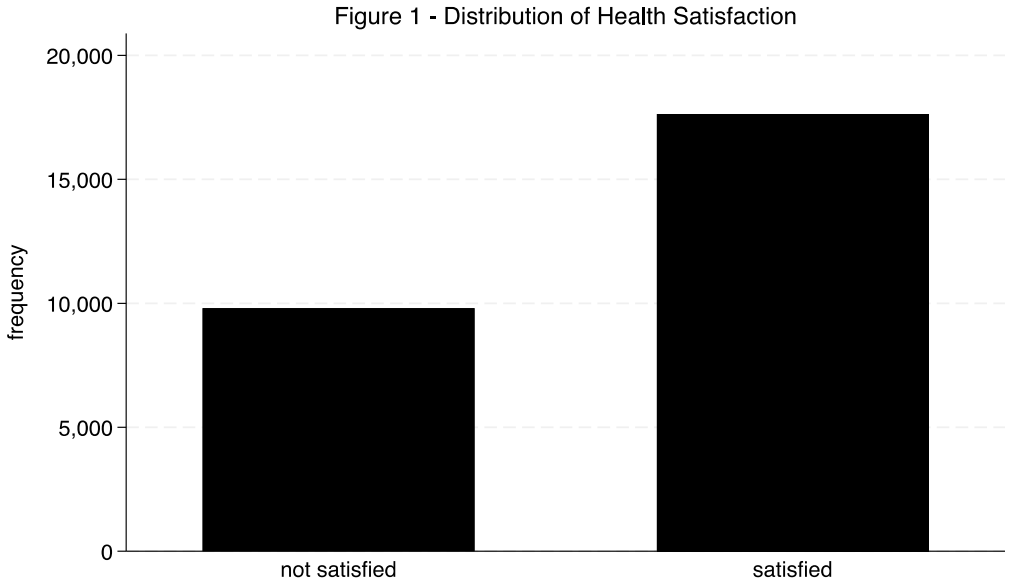
This essay primarily focuses on how income may influence an individual's judgement on their own health satisfaction and observe potential factors that stand out to differentiating whether one is satisfied or not by estimating regressions and conducting statistical tests.

During this investigation, I question "Whether a higher income means a higher level of health satisfaction" by utilising cross-sectional data from the Understanding Society Survey (UKHLS), specifically from Wave 13. Studying the relationship between health satisfaction and income can contribute to three key pillars of our society: public health, economics, and sociology. Finding the relationship between these two factors could help to improve health outcomes and mitigate health inequality.

Data and Methodology

This study uses data from the Understanding Society Survey (UKHLS). It is a nationally represented survey of respondents based in the UK. Data relating to Wave 13 is analysed and represents a cross-sectional investigation.

The focal point of this investigation is annual income and how it may impact on an individual's satisfaction on their health. In the UKHLS, one of the questions asked was what their total personal monthly labour income was. I will be using this explanatory variable as my focal point for this investigation; to determine whether there is statistical evidence to show that a higher wage increases an individual's health satisfaction. To follow, the dependent variable is health satisfaction: where the respondent was asked whether they were satisfied with their health or not. I decided to recode the responses to: 'not satisfied' (grouping categories 1-4) or 'satisfied' (grouping categories 5-7) with personal health (Table A1), following Daroudi et al (2016), visually seen on Figure 1.



I would like to impose a Linear Probability Model (LPM), which will be performed by using the Ordinary-Least Squares (OLS) Estimator:

$\Pr(healthsat = 1) = \beta_0 + \beta_1 \ln(wage) + \gamma\alpha_i$ (1)

'Pr(*healthsat*)' represents the probability that an individual says that they are satisfied with their health and *wage* is our estimated annual labour income. Controls included within my analysis are similar to the determinants used by Jones et al (2008) and Jones et al (2009), with some tailored

additions from the variables provided in the UKHLS survey. Variables included are: age, gender, highest level of qualification, ethnicity, postal code, number of drinks per week, hours worked, hours of overtime work, income, sleep time, if the individual smokes, how many vegetables they have a week, number of days they walk a week, if they are a carer, job satisfaction, marital status, type of job & moderate activity they have. A full list of how the variables are coded with descriptive statistics are shown in the appendix, resp. (Table A2/A3).

The LPM is good for provide us with a base model to work with. However, it follows a linear distribution which can hold back the reliability of binary-based dependent variable regressions as it may estimate coefficients to be outside the 0 to 1 probability range, violating the law of probability. A comparison between the LPM, Logit and Probit (resp.) models can be seen on Figure 2.

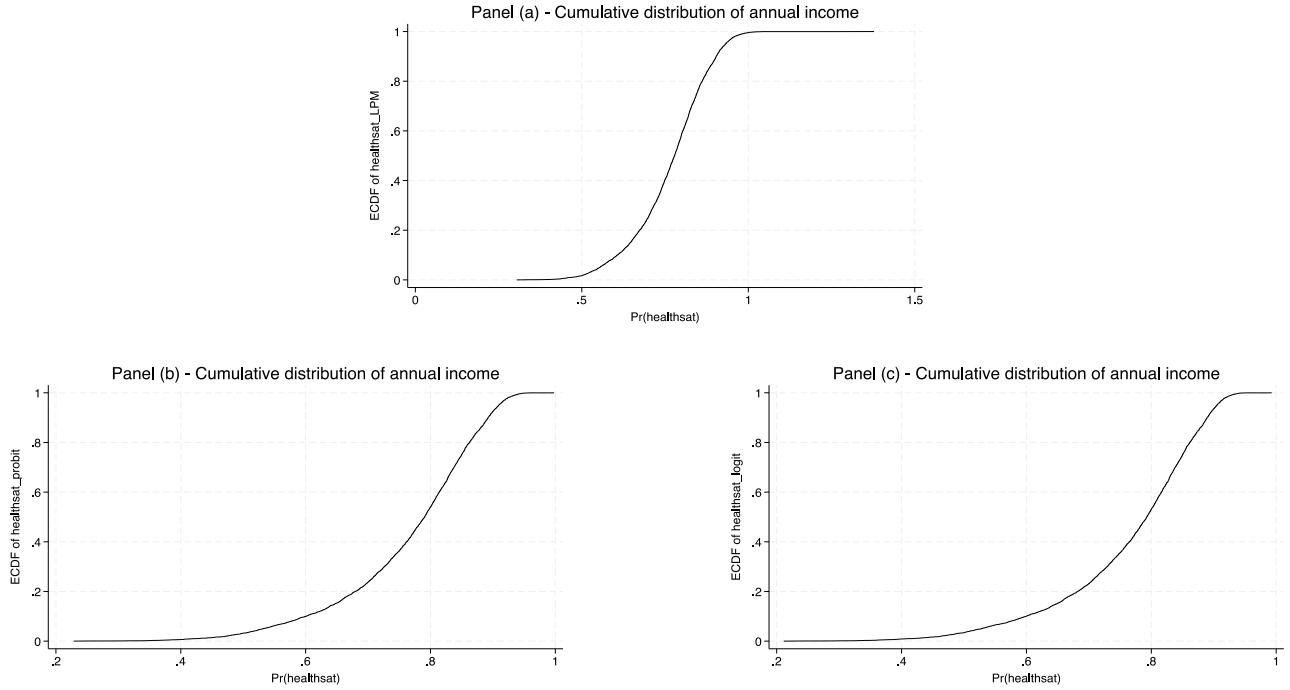


Figure 2 – Cumulative Distribution Functions

Therefore, for more accurate results, I will use a sigmoid-distribution regression:

$$\Pr(healthsat = 1) = G(X_i\beta) = \Lambda(X_i\beta) \quad (2)$$

$\Lambda(.)$ represents the cumulative logistic function.

I have decided to move forward with equation (2) as it expresses similar attributes to the binary satisfaction function seen in Graham et al (2009).

Results

Firstly, despite its importance, I will not be using income satisfaction as an explanatory as it affects the p-value of log wage (Table 1x), full results for base regression are on Table (B1).

Table 1x – Comparison of lnwage with/without incsat

	With Income Satisfaction	Without Income Satisfaction
lnwage	0.0049	0.0589***
	(0.0558)	(0.0533)
incsat	1.916***	------
N	9067	9072

Notes: Coefficients represent marginal effects. Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Nevertheless, I will be exploring the relationship between health satisfaction, income, and highest level of education. Table 1a shows results for secondary graduates. Similar to previously, the LPM generates heteroskedastic errors (p-value < 0.01) and predicted values lie outside the 0-1 probability range. Thus, sigmoid-distributed models are preferred in this scenario.

We see the probit and logit models produce similar results by having relative values for both pseudo R^2 and count R^2. It is common for these values to be small in binary-based dependent variables, as found by Veall et al (1994). Count R^2 is calculated to assess the accuracy of the binary choice models and measures the proportion of observations correctly predicted by the model. The values 70.70 and 70.73 (respectively) suggests there is relatively high correlation between the dependent and explanatory variables.

To interpret the effects, margin effects have been calculated. For example, coefficients suggest that (hvc) secondary graduates with an additional unit of annual income are approximately 6-7% points more likely to be satisfied with their health. We can also interpret the results using odds ratio, exclusive to logit models. Results here predicts that higher levels of income will 1.4 times more likely be satisfied with their health. FitzRoy et al (2018), Frijters et al (2005) come to similar conclusions.

Table 1a – Relationship between Health Satisfaction, Income, and Education (Secondary)

	LPM	Probit	Logit
lnwage	0.0668*** (0.0165)	0.0721*** (0.0523)	0.0761*** (0.0240) [1.403][x]
Diagnostics			
R^2	0.107		
Pseudo R^2		0.087	0.087
Count R^2 (baseline)		70.70 (70.49)	70.73 (67.85)
Average VIF value	4.97		
Heteroskedasticity (BP) Test	75.41 (0.000)		
Predicted Values Range	Min: 0.062 Max: 1.484		
N		3789	

Notes: Coefficients represent marginal effects. Robust standard errors are in parentheses. * p < 0.1, ** p < 0.05, p < 001. Breush-Pagan test used all explanatory variables. [x] Odds ratio. Other variables included in the model: see appendix.
Full regression results can be found in Appendix (B2).

Table 1b – Relationship between Health Satisfaction, Income, and Education (UG)

	LPM	Probit	Logit
lnwage	0.0525*** (0.0165)	0.0562*** (0.0177)	0.0567*** (0.0185) [1.3129][x]
Diagnostics			
R^2	0.089		
Pseudo R^2		0.075	0.075
Average VIF value	5.43		
Count R^2 (baseline)		72.99 (71.62)	72.93 (71.62)
Heteroskedasticity (BP) Test	86.55 (0.000)		
Predicted Values Range	Min: 0.119 Max: 1.534		
N		3214	

Notes: " ", Appendix (B3)

Table 1b suggests the same upwards-trend relationship, for undergraduates, as previously seen. As seen before, the LPM generates heteroskedastic errors and predicted values that are > 1. The R^2, pseudo-R^2 and count R^2 values are all relatively 'weaker' compared to our prior model. Nevertheless,

marginal effects are still highly significant and from them, we can suggest that (hvc), undergraduates with an additional unit of annual income are approximately 5-6% points more likely to be satisfied with their health. Again, the odds ratio can be used for the logit model, with results suggesting that an individual is 1.3 times more likely to be satisfied with their health with higher income.

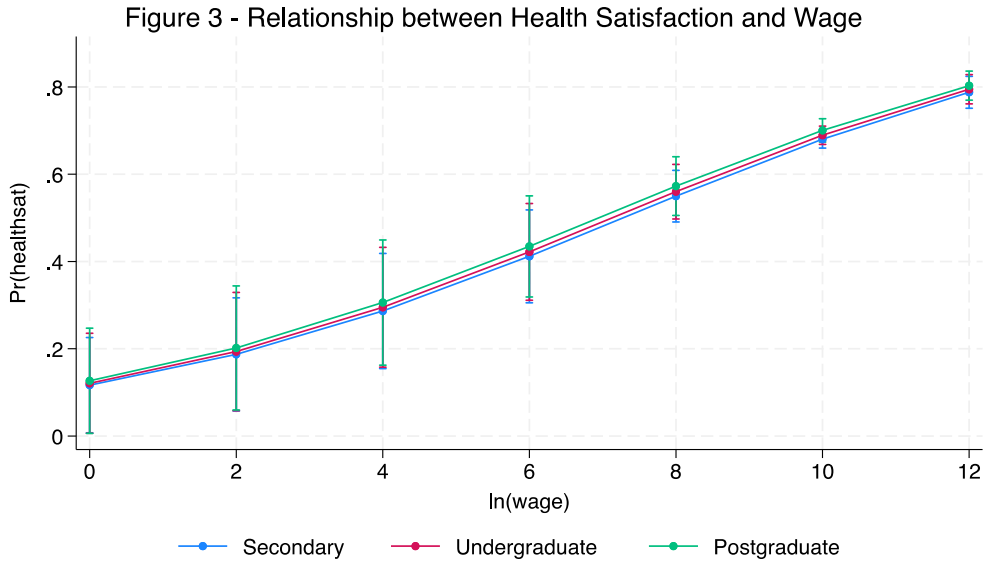
We can also analyse the effects of income and health satisfaction with postgraduates. Table 1c again shows the positive relationship between income and health satisfaction. The marginal effects suggest that (hvc) for every increase in income, the postgraduate is ~5-7% points more likely to be satisfied with their health. Again, this can be interpreted through the odds ratio, which suggests that an increase in income will make an individual 1.3 times more likely to say they are satisfied with their health.

Table 1c – Relationship between Health Satisfaction, Income and Education (PG)

	LPM	***Probit***	Logit
lnwage	0.0667*** (0.0129)	0.0473*** (0.0291)	0.0455*** (0.02) [1.320] [x]
Diagnostics			
R^2	0.083		
Pseudo R^2		0.0072	0.072
Count R^2 (baseline)		75.32 (74.30)	75.38 (74.25)
Average VIF value	5.55		
Heteroskedasticity (BP) Test	59.43 (0.000)		
Predicted Values Range	Min: 0.199 Max: 1.304		
N		1856	

Notes: " ", Appendix (B4)

Overall, we see that all three cases suggest that a higher income leads to an increase in likelihood that an individual is satisfied with their health, with postgraduates generally having the higher probability of being satisfied with their health (shown on Figure 3).



However, the marginal effects gradually diminish when the individual's highest qualification increases. This finding is quite interesting and similar to what Ross et al (1995) found, where higher educated people frequently engage in positive behaviours which may protect their health as they don't excessively drink and smoke less. With this in mind, our findings correlate with these outcomes: highly educated people already engage with healthier lifestyles and so an increase in their annual income won't have as significant of an impact compared to an individual who graduated from high school. Helmert et al (1989), Millar et al (1986) have similar conclusions.

We can also explore the difference in health satisfaction with an increase in income by splitting the sample by gender. Doing this allows us to observe whether there is a difference in a change of health satisfaction between the genders. Results found are consistent to before; higher income will make an individual more likely to be satisfied with their health. Ross et al (1994) suggests that women health is worse than men, although the gap closes with age, thus suggesting that an increase in wages should have a larger effect on women. Figures 4a and b agree with this statement – we see that initially women start off with lower satisfaction and gradually converge with men's health satisfaction. We can see a clearer effect after running an interaction term between age and gender: where female satisfaction converges towards men's HS, at around 51, to then increasing being after reaching the local minimum. This could be due to plentiful reasons, such as an earlier retirement compared to men, suggested by Fooken (1982).

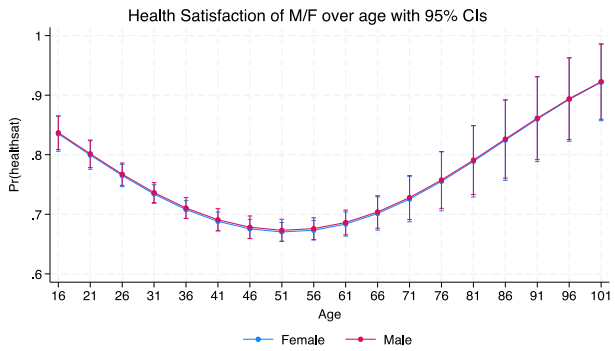


Figure 4a – Prediction plots (By gender).

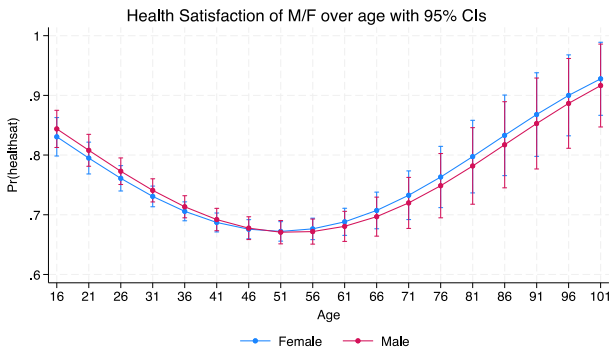


Figure 4b – Prediction plots: interaction term between age and gender.

On the other hand, when observing statistically, results suggest that men are 1.5 times more likely to be satisfied with their health with an increase in income as opposed to females with only a 1.2 times more likely (refer to Table 2). There could be various reasons for this, such as more drastic lifestyle changes in men: with an increase in wages, they may want to adapt a healthier lifestyle; Verbrugge et al (1989) have found that men are more likely to be overweight than women. Another reason could be that women workers tend to dominate in the work hierarchy (earn more), meaning the changes in their income won't impact as much as men (Rytina, 1982).

Table 2 – Relationship between Health Satisfaction and Income, by Gender

Variable	Male	Female
lnwage	1.516*** (0.116)	1.213*** (0.082)
Max wage	183,996	199,992
N	4051	5021

Notes: Coefficients represent the odds-ratio. RSEs in parentheses. * p < 0.1, ** p < 0.05, p < 001.

Conclusion

All in all, this study has attempted to investigate the relationship between health satisfaction and income, whilst looking at the impacts of education and gender. Through statistical tests and performing sub-group regressions, our results suggest that health satisfaction is majorly influenced by income (Currie et al, 1999): a higher level of education determining the magnitude of change in health satisfaction that comes with an increase in income, which is consistent with previous findings.

Although we have found some holistic reports, there are some limitations. Firstly, the number of observations in our analysis reduced drastically due to invalidation and so our model may not provide a holistic opinion on all individuals who filled out the survey. Therefore, having a larger

sample to work with will improve the accuracy of the findings. Additionally, the dataset utilised was cross-sectional limits our interpretations on how an individual's health satisfaction changes across a timeframe. This area could be improved upon by utilising time-series based data.

Moreover, further investigation into this topic could be to analyse the difference in health satisfaction between individuals with and without consistent wages: Ross et al (1994) alludes to this point as a significant determinant in one's opinion on their health satisfaction.

(2200 Words)

Bibliography

Adler, N.E., Boyce, T., Chesney, M.A., Cohen, S., Folkman, S., Kahn, R.L. and Syme, S.L. (1994). Socioeconomic status and health. The challenge of the gradient. *The American Psychologist*, [online] 49(1), pp.15–24. doi:https://doi.org/10.1037//0003-066x.49.1.15.

Contoyannis, P., Jones, A.M. and Rice, N. (2004). The Dynamics of Health in the British Household Panel Survey. *Journal of Applied Econometrics*, [online] 19(4), pp.473–503. Available at: https://www.jstor.org/stable/pdf/25146298.pdf?refreqid=fastly-default%3Ae9fe68b68b9ee36f159c746103ddd1c4&ab_segments=&origin=&initiator=&acceptTC=1

Currie, J., Ucla, N., Brigitte, C. and Madrian (1999). *HEALTH, HEALTH INSURANCE AND THE LABOR MARKET*. [online] Available at: http://www.econ.ucla.edu/people/papers/currie/hole.pdf.

Daroudi, R., Rashidian, A., Zeraati, H., Oliyaeemanesh, A. and Akbari Sari, A. (2016). Life and health satisfaction in the adult population of Iran. *Epidemiology and Health*, [online] 38, p.e2016047. doi:https://doi.org/10.4178/epih.e2016047.

Dolan, P. and White, M.P. (2007). How Can Measures of Subjective Well-Being Be Used to Inform Public Policy? *Perspectives on psychological science : a journal of the Association for Psychological Science*, [online] 2(1), pp.71–85. doi:https://doi.org/10.1111/j.1745-6916.2007.00030.x.

FitzRoy, F.R. and Nolan, M.A. (2018). Education, income and happiness: panel evidence for the UK. *Empirical Economics*, 58. doi:https://doi.org/10.1007/s00181-018-1586-5.

Fooken, I. (1982). Patterns of Health Behavior, Life Satisfaction, and Future Time Perspective in a Group of Old Aged Women: Data of 'Survivors' from a Longitudinal Study on Aging. *International Journal of Behavioral Development*, 5(3), pp.367–390. doi:https://doi.org/10.1177/016502548200500306.

Godin, I., Kittel, F., Coppieters, Y. and Siegrist, J. (2005). A prospective study of cumulative job stress in relation to mental health. *BMC Public Health*, 5(1). doi:https://doi.org/10.1186/1471-2458-5-67.

Graham, C. and Lora, E. (2009). *Paradox and perception : measuring quality of life in Latin America*. Washington, D.C.: Inter-American Development Bank : Brookings Institution Press.

Helmert, U., Herman, B., Joeckel, K.-H., Greiser, E. and Madans, J.H. (1989). Social class and risk factors for coronary heart disease in the Federal Republic of Germany. Results of the baseline survey of the German Cardiovascular Prevention Study (GCP). *Journal of Epidemiology and Community Health*, 43(1), pp.37–42. doi:https://doi.org/10.1136/jech.43.1.37.

Jones, A.M. and Schurer, S. (2009). How does heterogeneity shape the socioeconomic gradient in health satisfaction? *Journal of Applied Econometrics*, 26(4), pp.549–579. doi:https://doi.org/10.1002/jae.1134.

Jones, A.M. and Wildman, J. (2008). Health, income and relative deprivation: Evidence from the BHPS. *Journal of Health Economics*, 27(2), pp.308–324. doi:https://doi.org/10.1016/j.jhealeco.2007.05.007.

Millar, W.J. and Wigle, D.T. (1986). Socioeconomic disparities in risk factors for cardiovascular disease. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale*

canadienne, [online] 134(2), pp.127–32. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1490638/.

Mohanty, M. (2019). *Effects of job satisfaction on the worker's wage and weekly hours: A simultaneous equations approach*. [online] Available at: https://doi.org/10.1016/j.socec.2019.01.004.

Ross, C.E. and Bird, C.E. (1994). Sex Stratification and Health Lifestyle: Consequences for Men's and Women's Perceived Health. *Journal of Health and Social Behavior*, 35(2), p.161. doi:https://doi.org/10.2307/2137363.

Ross, C.E. and Wu, C. (1995). The Links Between Education and Health. *American Sociological Review*, [online] 60(5), p.719. doi:https://doi.org/10.2307/2096319.

Rytina, N.F. (1982). Earnings of Men and Women: A Look at Specific Occupations. 105(4), pp.25–31.

Veall, M.R. and Zimmermann, K.F. (1994). Evaluating Pseudo-R2's for binary probit models. *Quality & Quantity*, 28(2), pp.151–164. doi:https://doi.org/10.1007/bf01102759.

Verbrugge, L.M. (1989). The twain meet: empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior*, [online] 30(3), pp.282–304. Available at: https://pubmed.ncbi.nlm.nih.gov/2778300/.

Appendix

Table A1 – Health Satisfaction Variable Categories

healthsat	Frequency	Percent	Cumulative
Not Satisfied (0)	**9803**	**35.73**	**35.73**
Satisfied (1)	**17634**	**64.27**	**100.0**
Total	**27437**	**100.00**	

Table A2 – Variable Definitions and Codes

Variable Name	Original Variable Names	Coding	Base Group	Expected Sign of Coefficient
heathsat	m_sclfsat1	*not satisfied, satisfied*	*N/A*	*N/A*
age	m_dvage	renamed to "age"	N/A	(+ve) until local max.
male	m_sex	female, **male**	male	(+ve)
educ	m_isced11_dv	**secondary**, UG, PG, other	secondary	(+ve)
race	m_racel_dv	**white**, asian, black, mixed & other	white	(-ve)
post	m_gor_dv	NEng, MidEng, SEng, Wales+NI, Scotland, **London**	london	(-ve)
drinks	m_auditc4	**0-2 drinks**, 3-4 drinks, 5+ drinks	**0-2 drinks**	(-ve)
hrsworked	m_jbhrs	N/A	N/A	(+ve)
othrs	m_jbot	N/A	N/A	(-ve)
wage	m_fimnlabgrs_dv	multiplied by 12 to get the annual income gross	N/A	(+ve)
sleep	m_hrs_slph & m_hrs_slp	minutes of actual sleep	N/A	(+ve)
smoker	m_smoker	not smoker, **smoker**	smoker	(+ve)
vegedays	m_wkvege	never, 1-3 days, 4-6 days, **everyday**	everyday	(-ve)
dayswalked	m_wday	N/A	N/A	(+ve)
carer	m_aidhrs	not carer, **carer**	carer	(+ve)
jobsat	m_jbsat	Inapplicable, **dissatisfied**, neutral, satisfied	dissatisfied	(+ve)
marbi	m_mastat_dv	**Not legally binded**, legally binded	not legally binded	(+ve)

jobty	m_jbnssec3_dv	unemployed, **management**, intermediate, routine	management	(+ve)
incsat	*m_sclfsat2*	***Dissatisfied**, satisfied, unemployed*	*Dissatisfied*	*(+ve)*
m_mday	m_mday	**No days with moderate activity**, days with moderate activity	No days with moderate activity	(+ve)

Variables that weren't included in the primary regression analyses.

Table A3 – Descriptive Statistics of dependent and explanatory variables

Variable	Num. of Observations	Mean	Standard Deviation	Min Value	Max Value
healthsat	27,437	.6427088	.4792104	0	1
Male	27,995	.4432934	.4967828	0	1
Educ	27,529	1.837408	.9640109	1	4
Race	27,905	1.250349	.6420568	1	4
Post	27,979	3.020229	1.527532	1	6
Drinks	27,740	1.369178	.6437168	1	3
Hrsworked	12,753	32.50256	10.82156	.1	97
Othrs	12,741	3.914999	8.197852	0	90
wage	27,965	15554.05	21258.65	0	199992
Sleep	22,476	416.7523	76.27019	0	1380
Smoker	27,836	.1057982	.3075847	1	1
Vegedays	27,758	3.069782	.8862988	1	4
Dayswalked	27,447	4.616825	2.527508	0	7
Carer	27,835	.1543381	.361279	0	1
Jobsat	27,838	2.358395	.6590603	1	3
Marbi	27,882	.5526146	.4972329	0	1
jobty	25,682	2.961218	1.205322	1	4
incsat	*27,512*	*1.663347*	*.4816399*	*1*	*3*
m_mday	27,259	.5117209	.4998718	0	1

Variables that weren't included in the primary regression analyses.

Table B1(i) – Base Regression Results with income satisfaction included

Regression	**(1)**	**(2)**	**(3)**
Model	**LPM**	**Probit**	**Logit**
age	-0.00561**	-0.0223***	-0.0378***
	(0.00222)	(0.00830)	(0.0144)
c.age#c.age	0.0000425*	0.000176*	0.000293*
	(0.0000250)	(0.0000926)	(0.000160)
Female	-0.00406	-0.0166	-0.0297
	(0.00937)	(0.0341)	(0.0594)
Male	0	0	0
	(.)	(.)	(.)
Secondary	0	0	0
	(.)	(.)	(.)
UG	-0.000359	0.000512	-0.00916
	(0.0105)	(0.0374)	(0.0649)
PG	0.00826	0.0281	0.0504
	(0.0128)	(0.0468)	(0.0817)
Other	0.0295	0.105	0.191
	(0.0282)	(0.109)	(0.191)
White	0	0	0
	(.)	(.)	(.)
Asian	-0.00349	-0.00927	-0.0178
	(0.0174)	(0.0609)	(0.104)
Black	0.0657**	0.230**	0.399**
	(0.0263)	(0.0922)	(0.160)
Mixed or Other	-0.000723	-0.00648	-0.0189
	(0.0269)	(0.0909)	(0.154)
North England	0.0153	0.0524	0.0853
	(0.0185)	(0.0676)	(0.118)
Midlands	-0.0162	-0.0578	-0.109
	(0.0162)	(0.0585)	(0.102)
South England	-0.0318*	-0.113*	-0.204*
	(0.0172)	(0.0618)	(0.107)
Wales and Northern Ireland	0.0201	0.0832	0.132
	(0.0187)	(0.0701)	(0.122)
Scotland	0.0170	0.0638	0.103
	(0.0201)	(0.0747)	(0.130)
London	0	0	0
	(.)	(.)	(.)
0-2 drinks	0	0	0
	(.)	(.)	(.)
3-4 drinks	0.0100	0.0306	0.0646
	(0.0103)	(0.0377)	(0.0662)

5+ drinks	-0.0243*	-0.0915*	-0.153*
	(0.0142)	(0.0496)	(0.0859)
Hours worked	-0.000276	-0.000629	-0.00119
	(0.000541)	(0.00196)	(0.00341)
Hours of overtime work	-0.000884	-0.00301	-0.00524
	(0.000546)	(0.00189)	(0.00329)
lnwage	0.00592	0.0187	0.0251
	(0.00862)	(0.0320)	(0.0558)
Sleep time	0.000439***	0.00160***	0.00273***
	(0.0000681)	(0.000243)	(0.000424)
Don’t smoke	-0.00681	-0.0273	-0.0481
	(0.0150)	(0.0515)	(0.0881)
Do smoke	0	0	0
	(.)	(.)	(.)
Never eat vegetables	-0.0618**	-0.198**	-0.332**
	(0.0290)	(0.0974)	(0.164)
Eat vegetables 1-3 days	-0.0266**	-0.0869**	-0.143**
	(0.0118)	(0.0404)	(0.0696)
Eat vegetables 4-6 days	0	0	0
	(.)	(.)	(.)
Eat vegetables everyday	0.0256**	0.0978***	0.171***
	(0.0100)	(0.0368)	(0.0641)
Number of days walked	0.0100***	0.0350***	0.0603***
	(0.00189)	(0.00655)	(0.0113)
Not a carer	0.0243*	0.0847*	0.141*
	(0.0130)	(0.0437)	(0.0757)
Carer	0	0	0
	(.)	(.)	(.)
Dissatisfied with Job	0	0	0
	(.)	(.)	(.)
Satisfied with Job	0.110***	0.351***	0.594***
	(0.0125)	(0.0383)	(0.0655)
Not legally binded	0	0	0
	(.)	(.)	(.)
Legally binded	0.00524	0.0179	0.0269
	(0.00996)	(0.0350)	(0.0605)
Job type – Management	0	0	0
	(.)	(.)	(.)
Job type – Intermediate	0.00601	0.0209	0.0393
	(0.0120)	(0.0431)	(0.0755)
Job type – Routine	0.0174	0.0594	0.0989
	(0.0113)	(0.0410)	(0.0708)
Not satisfied with Income	0	0	0

	(.)	(.)	(.)
Satisfied with income	0.395***	1.149***	1.916***
	(0.0112)	(0.0332)	(0.0564)
No days of activity	0	0	0
	(.)	(.)	(.)
Has days of activity	0.0648***	0.233***	0.401***
	(0.00890)	(0.0314)	(0.0547)
_cons	0.159*	-1.044***	-1.679***
	(0.0893)	(0.334)	(0.581)
R^2	0.230		
pseudo R^2		0.190	0.190
N	9067	9067	9067

Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Other variables included in the model: age, gender, ethnicity, geographical location, number of drinks daily, hours worked, hours of work overtime, hours of sleep, if they smoke, number of vegetables a week, days walked, if they are a carer, job satisfaction, marital status, job type, income satisfaction and whether they have days where they do moderate activity.

Table B1(ii) – Base Regression Results with income satisfaction not included

Regression	**(1)**	**(2)**	**(3)**
Model	**LPM**	**Probit**	**Logit**
age	-0.0149***	-0.0475***	-0.0804***
	(0.00242)	(0.00787)	(0.0133)
c.age#c.age	0.000144***	0.000458***	0.000777***
	(0.0000273)	(0.0000882)	(0.000149)
Female	-0.00236	-0.00677	-0.0137
	(0.0102)	(0.0323)	(0.0547)
Male	0	0	0
	(.)	(.)	(.)
Secondary	0	0	0
	(.)	(.)	(.)
UG	0.00837	0.0227	0.0415
	(0.0114)	(0.0354)	(0.0596)
PG	0.0172	0.0518	0.0934
	(0.0139)	(0.0442)	(0.0749)
Other	0.0460	0.133	0.249
	(0.0310)	(0.103)	(0.175)
white	0	0	0
	(.)	(.)	(.)
Asian	-0.0355*	-0.108*	-0.181*
	(0.0193)	(0.0575)	(0.0954)
Black	0.0363	0.114	0.195
	(0.0281)	(0.0862)	(0.145)
Mixed or Other	-0.0394	-0.122	-0.202
	(0.0300)	(0.0885)	(0.148)

North England	0.0366*	0.113*	0.193*
	(0.0201)	(0.0637)	(0.108)
Midlands	-0.000296	-0.00156	-0.00128
	(0.0176)	(0.0550)	(0.0927)
South England	-0.0128	-0.0398	-0.0675
	(0.0186)	(0.0582)	(0.0981)
Wales and Northern Ireland	0.0587***	0.193***	0.323***
	(0.0203)	(0.0663)	(0.112)
Scotland	0.0435**	0.141**	0.239**
	(0.0218)	(0.0708)	(0.120)
London	0	0	0
	(.)	(.)	(.)
0-2 drinks	0	0	0
	(.)	(.)	(.)
3-4 drinks	0.00743	0.0222	0.0352
	(0.0110)	(0.0353)	(0.0599)
5+ drinks	-0.0296*	-0.0939**	-0.159**
	(0.0152)	(0.0466)	(0.0782)
Hours worked	-0.00154**	-0.00463**	-0.00806**
	(0.000601)	(0.00187)	(0.00321)
Hours of overtime work	-0.00134**	-0.00414**	-0.00694**
	(0.000593)	(0.00177)	(0.00295)
lnwage	0.0539***	0.164***	0.278***
	(0.00990)	(0.0307)	(0.0533)
Sleep time	0.000711***	0.00221***	0.00376***
	(0.0000742)	(0.000232)	(0.000398)
Don’t smoke	0.0237	0.0714	0.110
	(0.0166)	(0.0488)	(0.0814)
Does smoke	0	0	0
	(.)	(.)	(.)
Never eat vegtables	-0.0861***	-0.246***	-0.408***
	(0.0323)	(0.0921)	(0.150)
Eat vegetables 1-3 days	-0.0387***	-0.110***	-0.180***
	(0.0129)	(0.0380)	(0.0633)
Eat vegetables 4-6 days	0	0	0
	(.)	(.)	(.)
Eat vegetables everyday	0.0455***	0.148***	0.255***
	(0.0108)	(0.0348)	(0.0590)
Number of days walked	0.0133***	0.0404***	0.0681***
	(0.00207)	(0.00622)	(0.0104)
Not a carer	0.0410***	0.123***	0.204***
	(0.0140)	(0.0410)	(0.0683)
Carer	0	0	0
	(.)	(.)	(.)
Dissatisfied with job	0	0	0

	(.)	(.)	(.)
Satisfied with job	0.201***	0.567***	0.933***
	(0.0130)	(0.0352)	(0.0579)
Not legally binded	0	0	0
	(.)	(.)	(.)
Legally binded	0.0252**	0.0770**	0.127**
	(0.0108)	(0.0330)	(0.0552)
Job type - Management	0	0	0
	(.)	(.)	(.)
Job type - Intermediate	0.00161	0.00162	0.00715
	(0.0130)	(0.0405)	(0.0686)
Job type - Routine	0.00508	0.0127	0.0263
	(0.0125)	(0.0390)	(0.0658)
No days with activity	0	0	0
	(.)	(.)	(.)
Days with activity	0.0637***	0.197***	0.328***
	(0.00963)	(0.0296)	(0.0499)
_cons	-0.0875	-1.765***	-2.996***
	(0.1000)	(0.317)	(0.542)
R^2	0.091		
pseudo R^2		0.076	0.076
N	9072	9072	9072

Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Other variables included in the model: age, gender, ethnicity, geographical location, number of drinks daily, hours worked, hours of work overtime, hours of sleep, if they smoke, number of vegetables a week, days walked, if they are a carer, job satisfaction, marital status, job type and whether they have days where they do moderate activity.

Table B2 – Regression Results with educ == 1

Regression	**(1)**	**(2)**	**(3)**
Model	**LPM**	**Probit**	**Logit**
age	-0.0154***	-0.0495***	-0.0838***
	(0.00339)	(0.0110)	(0.0186)
c.age#c.age	0.000152***	0.000487***	0.000826***
	(0.0000387)	(0.000124)	(0.000209)
0.male	-0.0173	-0.0564	-0.0942
	(0.0164)	(0.0508)	(0.0858)
1.male	0	0	0
	(.)	(.)	(.)
1.educ	0	0	0
	(.)	(.)	(.)
white	0	0	0
	(.)	(.)	(.)
Asian	-0.0472	-0.142	-0.227

	(0.0327)	(0.0949)	(0.157)
Black	0.0469	0.155	0.267
	(0.0486)	(0.147)	(0.244)
Mixed or Other	0.0221	0.0732	0.122
	(0.0479)	(0.151)	(0.257)
North England	0.0444	0.139	0.229
	(0.0354)	(0.106)	(0.179)
Midlands	0.0201	0.0669	0.104
	(0.0313)	(0.0934)	(0.156)
South England	0.00918	0.0326	0.0487
	(0.0331)	(0.0989)	(0.166)
Wales and Northern Ireland	0.0835**	0.270**	0.439**
	(0.0350)	(0.108)	(0.181)
Scotland	0.0840**	0.279**	0.459**
	(0.0379)	(0.119)	(0.201)
London	0	0	0
	(.)	(.)	(.)
0-2 drinks	0	0	0
	(.)	(.)	(.)
3-4 drinks	0.0192	0.0578	0.0973
	(0.0175)	(0.0547)	(0.0923)
5+ drinks	-0.0154	-0.0492	-0.0808
	(0.0218)	(0.0670)	(0.113)
Hours worked	-0.00314***	-0.00953***	-0.0164***
	(0.000949)	(0.00298)	(0.00509)
Hours of overtime work	-0.00265**	-0.00766**	-0.0126**
	(0.00115)	(0.00329)	(0.00559)
lnwage	0.0668***	0.196***	0.339***
	(0.0165)	(0.0523)	(0.0903)
sleep	0.000751***	0.00234***	0.00396***
	(0.000106)	(0.000336)	(0.000571)
Don’t smoke	-0.00352	-0.0146	-0.0230
	(0.0217)	(0.0647)	(0.108)
Does smoke	0	0	0
	(.)	(.)	(.)
1.vegedays	-0.103**	-0.299**	-0.503**
	(0.0444)	(0.127)	(0.205)
Eat vegetables 1-3 days	-0.0343*	-0.0988*	-0.160*
	(0.0184)	(0.0540)	(0.0899)
Eat vegetables 4-6 days	0	0	0
	(.)	(.)	(.)
Eat vegetables everyday	0.0588***	0.192***	0.322***
	(0.0176)	(0.0564)	(0.0954)
Number of days walked	0.0116***	0.0340***	0.0590***
	(0.00325)	(0.00955)	(0.0160)

Not a carer	0.0467**	0.137**	0.225**
	(0.0211)	(0.0607)	(0.101)
Carer	0	0	0
	(.)	(.)	(.)
Dissatisfied with job	0	0	0
	(.)	(.)	(.)
Satisfied with job	0.239***	0.661***	1.081***
	(0.0199)	(0.0540)	(0.0889)
Not legally binded	0	0	0
	(.)	(.)	(.)
Legally binded	-0.00119	0.000985	-0.00777
	(0.0170)	(0.0502)	(0.0837)
Job type - Management	0	0	0
	(.)	(.)	(.)
Job type - Intermediate	0.00828	0.0240	0.0425
	(0.0214)	(0.0656)	(0.110)
Job type - Routine	0.00297	0.00831	0.0214
	(0.0184)	(0.0566)	(0.0947)
No days with activity	0	0	0
	(.)	(.)	(.)
Days with activity	0.0517***	0.157***	0.256***
	(0.0153)	(0.0459)	(0.0771)
_cons	-0.178	-1.942***	-3.329***
	(0.157)	(0.502)	(0.859)
R^2	0.107		
pseudo R^2		0.087	0.087
N	3789	3789	3789

Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B3 – Regression Results with educ == 2

Regression	**(1)**	**(2)**	**(3)**
Model	**LPM**	**Probit**	**Logit**
age	-0.0169***	-0.0549***	-0.0924***
	(0.00457)	(0.0149)	(0.0254)
c.age#c.age	0.000169***	0.000546***	0.000923***
	(0.0000513)	(0.000167)	(0.000285)
0.male	0.00119	0.00771	0.0128
	(0.0170)	(0.0542)	(0.0921)
1.male	0	0	0
	(.)	(.)	(.)
1.educ	0	0	0
	(.)	(.)	(.)

2.educ	0 (.)	0 (.)	0 (.)
white	0 (.)	0 (.)	0 (.)
Asian	-0.00700 (0.0314)	-0.0250 (0.0952)	-0.0484 (0.159)
Black	0.0607 (0.0440)	0.171 (0.140)	0.302 (0.238)
Mixed or Other	-0.0768 (0.0482)	-0.226 (0.138)	-0.374* (0.226)
North England	0.0248 (0.0324)	0.0734 (0.105)	0.129 (0.178)
Midlands	-0.000839 (0.0279)	-0.00832 (0.0899)	-0.00430 (0.152)
South England	-0.0139 (0.0296)	-0.0474 (0.0947)	-0.0781 (0.159)
Wales and Northern Ireland	0.0386 (0.0342)	0.126 (0.114)	0.225 (0.194)
Scotland	0.0197 (0.0364)	0.0563 (0.118)	0.110 (0.201)
London	0 (.)	0 (.)	0 (.)
0-2 drinks	0 (.)	0 (.)	0 (.)
3-4 drinks	-0.0225 (0.0185)	-0.0740 (0.0587)	-0.128 (0.0996)
5+ drinks	-0.0471* (0.0266)	-0.147* (0.0818)	-0.252* (0.136)
Hours worked	-0.000340 (0.00101)	-0.000640 (0.00317)	-0.00154 (0.00542)
Hours of overtime work	-0.00133 (0.000896)	-0.00407 (0.00270)	-0.00712 (0.00447)
lnwage	0.0525*** (0.0165)	0.162*** (0.0505)	0.272*** (0.0876)
Sleep time	0.000791*** (0.000139)	0.00238*** (0.000419)	0.00411*** (0.000732)
Don’t smoke	0.0560* (0.0313)	0.164* (0.0910)	0.266* (0.151)
Does smoke	0 (.)	0 (.)	0 (.)
1.vegedays	-0.0465 (0.0549)	-0.129 (0.159)	-0.216 (0.261)
Eat vegetables 1-3 days	-0.0519** (0.0228)	-0.146** (0.0672)	-0.242** (0.112)
Eat vegetables 4-6 days	0	0	0

	(.)	(.)	(.)
Eat vegetables everyday	0.0443**	0.150***	0.260***
	(0.0177)	(0.0572)	(0.0973)
Number of days walked	0.0183***	0.0564***	0.0939***
	(0.00345)	(0.0104)	(0.0174)
Not a carer	0.0429*	0.135*	0.221*
	(0.0250)	(0.0731)	(0.123)
Carer	0	0	0
	(.)	(.)	(.)
Dissatisfied with job	0	0	0
	(.)	(.)	(.)
Satisfied with job	0.159***	0.461***	0.761***
	(0.0219)	(0.0605)	(0.0996)
Not legally binded	0	0	0
	(.)	(.)	(.)
Legally binded	0.0423**	0.129**	0.220**
	(0.0182)	(0.0561)	(0.0945)
Job type - Management	0	0	0
	(.)	(.)	(.)
Job type - Intermediate	0.00931	0.0251	0.0429
	(0.0202)	(0.0638)	(0.108)
Job type - Routine	0.00778	0.0166	0.0361
	(0.0214)	(0.0673)	(0.115)
No days with activity	0	0	0
	(.)	(.)	(.)
Days with activity	0.0824***	0.259***	0.431***
	(0.0162)	(0.0502)	(0.0847)
_cons	-0.132	-1.902***	-3.231***
	(0.183)	(0.570)	(0.982)
R^2	0.089		
pseudo R^2		0.075	0.075
N	3214	3214	3214

Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.0$

Table B4 – Regression Results with educ == 3

Regression	**(1)**	**(2)**	**(3)**
Model	**LPM**	**Probit**	**Logit**
main			
age	-0.0127*	-0.0434*	-0.0723*
	(0.00673)	(0.0225)	(0.0385)
c.age#c.age	0.000107	0.000367	0.000609
	(0.0000731)	(0.000242)	(0.000413)

0.male	0.0161	0.0537	0.0891
	(0.0215)	(0.0713)	(0.123)
1.male	0	0	0
	(.)	(.)	(.)
1.educ	0	0	0
	(.)	(.)	(.)
3.educ	0	0	0
	(.)	(.)	(.)
white	0	0	0
	(.)	(.)	(.)
Asian	-0.0784*	-0.236*	-0.417**
	(0.0407)	(0.121)	(0.204)
Black	-0.0367	-0.0860	-0.160
	(0.0597)	(0.176)	(0.293)
Mixed or Other	-0.0707	-0.231	-0.381
	(0.0643)	(0.186)	(0.315)
North England	0.0511	0.178	0.311
	(0.0391)	(0.135)	(0.237)
Midlands	-0.0279	-0.0921	-0.163
	(0.0349)	(0.112)	(0.193)
South England	-0.0344	-0.112	-0.198
	(0.0368)	(0.119)	(0.205)
Wales and Northern Ireland	0.0358	0.132	0.204
	(0.0399)	(0.137)	(0.237)
Scotland	0.0113	0.0499	0.0620
	(0.0409)	(0.138)	(0.240)
London	0	0	0
	(.)	(.)	(.)
0-2 drinks	0	0	0
	(.)	(.)	(.)
3-4 drinks	0.0503**	0.172**	0.294**
	(0.0234)	(0.0833)	(0.145)
5+ drinks	-0.0301	-0.102	-0.166
	(0.0400)	(0.123)	(0.209)
Hours worked	0.000247	0.000689	0.00159
	(0.00140)	(0.00440)	(0.00755)
Hours of overtime work	-0.000231	-0.000884	-0.00122
	(0.00116)	(0.00368)	(0.00629)
lnwage	0.0305	0.101	0.165
	(0.0209)	(0.0654)	(0.111)
Sleep time	0.000520***	0.00178***	0.00303***
	(0.000165)	(0.000571)	(0.000989)
Don’t smoke	0.0931*	0.288*	0.465*
	(0.0546)	(0.152)	(0.257)

Does smoke	0 (.)	0 (.)	0 (.)
1.vegedays	-0.0923 (0.108)	-0.264 (0.303)	-0.409 (0.522)
Eat vegetables 1-3 days	-0.0200 (0.0334)	-0.0543 (0.102)	-0.0796 (0.172)
Eat vegetables 4-6 days	0 (.)	0 (.)	0 (.)
Eat vegetables everyday	0.0340 (0.0231)	0.111 (0.0760)	0.196 (0.130)
Number of days walked	0.0104** (0.00456)	0.0334** (0.0144)	0.0561** (0.0245)
Not a carer	0.0183 (0.0301)	0.0575 (0.0948)	0.0975 (0.160)
Carer	0 (.)	0 (.)	0 (.)
Dissatisfied with job	0 (.)	0 (.)	0 (.)
Satisfied with job	0.186*** (0.0288)	0.544*** (0.0804)	0.906*** (0.133)
Not legally binded	0 (.)	0 (.)	0 (.)
Legally binded	0.0521** (0.0235)	0.162** (0.0743)	0.283** (0.126)
Job type - Management	0 (.)	0 (.)	0 (.)
Job type - Intermediate	-0.0245 (0.0300)	-0.0799 (0.0947)	-0.121 (0.164)
Job type - Routine	0.000813 (0.0342)	-0.000992 (0.109)	-0.00228 (0.183)
No days with activity	0 (.)	0 (.)	0 (.)
Days with activity	0.0628*** (0.0207)	0.202*** (0.0667)	0.343*** (0.114)
_cons	0.145 (0.241)	-1.187 (0.787)	-1.986 (1.345)
R^2	0.083		
pseudo R^2		0.072	0.072
N	1856	1856	1856

Robust Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

.do file (Stata Commands)

```
**Contents - 100379318**
*1. Cleaning the Data
*2. Tabulation of all Variables
*3. LPM, Probit and Logit Base Regressions
*4. Diagonstic tests for base
*5. Constructing predicted values range for LPM regression
*6. Regressions of subsamples
*7. Diagonstic tests of subsamples
*8. Predicted Values Range of subsamples
*9. Visualisation and Tabulations


/*1. Cleaning the data*

*Variables to clean:
*m_sclfsat1 - satisfaction with health /
*m_dvage - age /
*m_sex - sex /
*m_isced11_dv – highest educational qualification, short ISCED 2011 /
*m_cel_dv – ethnic group /
*m_gor_dv – Government Office Region /
*m_auditc4 - drinks on a typical day /
*m_jbhrs - number of hours normally worked per week/
*m_jbot - number of overtime hours in normal week/
*m_fimnlabgrs_dv - total monthly labour income gross (make annual)/
*m_hrs_slph & m_hrs_slpm - minutes of actual sleep (hrs & mins)/
*m_smoker - smoker/
*m_wkvege - days each week eat vegetables/
*m_wday - number of days individual walked (for >10 mins)/
*m_aidhrs - hours per week spent caring/
*m_jbsat - job satisfaction/
*m_sclfsato - lifesatisfaction/
*m_mastat_dv - marital status/
*m_sclfsat2 - satisfaction with income/
*m_vdhrs & m_vdmin - time doing healthy activities (convert to mins)/
*m_jbnssec3_dv - current job/

/*satisfcation with health*
tab m_sclfsat1
gen healthsat = m_sclfsat1
replace healthsat =. if healthsat<0
replace healthsat=0 if inlist(healthsat, 1, 2, 3, 4)
replace healthsat=1 if inlist(healthsat, 5, 6, 7)
label variable healthsat "Tells whether an individual is satisfied with their health or not"
label variable healthsat "0=not_satisfied, 1=satisfied"
tab healthsat


*/

*age variable*
rename m_dvage age

/*gender variable*
tab m_sex
gen male = m_sex
replace male =. if male <0
replace male =0 if male == 2
```

```
replace male =1 if male == 1
label define male 0 "Female" 1 "Male"
label values male male
label variable male "0 = female, 1 = male"
tab male
*/

/*educ variable - highest qualification*
tab m_isced11_dv
gen educ = m_isced11_dv
tab educ
replace educ =. if educ<0
replace educ = 1 if inlist(educ, 2, 3)
replace educ = 2 if educ == 6
replace educ = 3 if inlist(educ, 7, 8)
replace educ = 4 if educ==96
label define educ 1 "Secondary" 2 "Undergraduate" 3 "Postgraduate" 4 "Other"
label values educ educ educ educ
label variable educ "1=Secondary, 2=UG, 3=PG, 4= Other"
tab educ
*/

/*race variable*
tab m_racel_dv
gen race = m_racel_dv
replace race=. if race<0
replace race=1 if inlist(race, 1, 2, 3, 4)
replace race=2 if inlist(race, 9, 10, 11, 12, 13)
replace race=3 if inlist(race, 14, 15, 16)
replace race=4 if inlist(race, 5, 6, 7, 8, 17, 97)
label define race 1 "White" 2 "Asian" 3 "Black" 4 "Mixed or Other"
label values race race race race
label variable race "1 = white, 2 = asian, 3 = black, 4 = mixed&other"
*/

/*post variable - government office region*
tab m_gor_dv
gen post = m_gor_dv
tab post
replace post=. if post <0
replace post = 1 if inlist(post, 1, 2)
replace post = 2 if inlist(post, 3, 4, 5, 6)
replace post = 3 if inlist(post, 8, 9)
replace post = 4 if inlist(post, 10, 12)
replace post = 5 if post == 11
replace post =6 if post == 7
label define post 1 "North of Eng" 2 "Midlands of Eng" 3 "South of Eng" 4 "Wales and Northern Ireland" 5 "Scotland" 6
"London"
label values post post post post post post
label variable post "1 = NEng, 2 = MidEng, 3=SEng, 4=W+NI, 5=Scot, 6=LDN"
tab post
*/

/*drinks variable*
tab m_auditc4
gen drinks = m_auditc4
tab drinks
replace drinks=. if inlist(drinks, -1,-2,-7,-9)
replace drinks=1 if inlist(drinks, -8,1)
replace drinks=3 if inlist(drinks, 3, 4, 5)
```

```
label define drinks 1 "0-2 drinks" 2 "3-4 drinks" 3 "5+ drinks"
label variable drinks "1 = 0-2 drinks, 2 = 3-4 drinks, 3 = 5+ drinks"
label values drinks drinks
*/


/*hours worked a week
tab m_jbhrs
rename m_jbhrs hrsworked
replace hrsworked=. if hrsworked<0
*/

/*hours of overtime work a week
tab m_jbot
rename m_jbot othrs
replace othrs=. if inlist(othrs, -9, -7, -2, -1)
tab othrs
*/

/*annual wage variable - CHECK THIS VARIABLE AFTER RECEIVING FEEDBACK FROM GROUP PROJECT
//convert from monthly to annual wage: (monthly wage*12 months)
gen wage = (m_fimnlabgrs_dv*12)
replace wage=. if wage<0

//lnwage variable*
gen lnwage = log(wage)
label variable lnwage "log of wage"
*/

/*sleeptime variable (in minutes)
tab m_hrs_slph
tab m_hrs_slpm
replace m_hrs_slph=. if m_hrs_slph<0
replace m_hrs_slpm=. if m_hrs_slpm<0
gen sleep = (m_hrs_slph*60)+m_hrs_slpm
tab sleep
*/

/*smoke variable
tab m_smoker
gen smoker = m_smoker
replace smoker=. if smoker<0
replace smoker=0 if smoker == 2
label variable smoker "0 = Not Smoker, 1 = Smoker"
label values smoker smoker
*/

/*vegedays variable
gen vegedays = m_wkvege
replace vegedays=. if vegedays<0
label define vegedays 1 "Never" 2 "1-3 Days" 3 "4-6 Days" 4 "Everday"
label values vegedays vegedays
label variable vegedays "1 = Never, 2 = 1-3 Days, 3 = 4-6 Days, 4 = Everyday"
*/

/*dayswalked variable
gen dayswalked = m_wday
replace dayswalked=. if dayswalked<0
label variable dayswalked "Number of days walked (for >10 mins)"
*/
```

```
/*carer variable
tab m_aidhrs
gen carer = m_aidhrs
replace carer=. if inlist(carer, -1, -2, -7)
replace carer=0 if carer==-8
replace carer=1 if inlist(carer, 1, 2, 3, 4, 5, 6, 7, 8, 9, 97)
label variable carer "0 = not carer, 1 = carer"
label values carer carer
tab carer
*/

/*jobsatisfaction variable*
//keeping '-8' because inapplicable is a variable negative label here because it is inapplicable if the individual doens't have a job
gen jobsat = m_jbsat
replace jobsat=. if inlist(jobsat, -1,-2,-7)
replace jobsat = 1 if inlist(jobsat,1,2,3,4)
replace jobsat= 2 if inlist(jobsat, 5,6,7)
replace jobsat= 3 if jobsat == -8
label define jobsat 1 "dissatisfied" 2 "satisfied" 3 "assumed unemployed (N/A)"
label values jobsat jobsat
label variable jobsat "1=dissatisfied, 2=satisfied, 3=inapplicable,"
tab jobsat
*/

/*marriage variable*
//creating a binary marriage variable - indidivuals who are legally married: 0-not true, 1- true
tab m_mastat_dv
gen marbi = m_mastat_dv
replace marbi=. if marbi<0
replace marbi=0 if inlist(marbi, 1, 5, 6, 7, 8, 9, 10)
replace marbi=1 if inlist(marbi, 2, 3, 4)
label define marbi 0 "not legally binded" 1 "legally binded"
label values marbi marbi
label variable marbi "0 - not legally binded, 1 - legally binded"
tab marbi
*/

/*incsat variable
tab m_sclfsat2
gen incsat = m_sclfsat2
replace incsat=. if inlist(incsat, -1,-2,-7,-9)
replace incsat=1 if inlist(incsat, 1,2,3,4)
replace incsat=2 if inlist(incsat, 5,6,7)
replace incsat=3 if incsat==-8
label define incsat 1 "Dissatisfied" 2 "Satisfied" 3 "Unemployed"
label values incsat incsat
label variable incsat "1 = dissatisfed, 2 = satisfied, 3 = unemployed"
tab incsat

*/

/*activetime variable (in minutes)
tab m_vdhrs
tab m_vdmin
replace m_vdhrs=. if m_vdhrs<0
replace m_vdmin=. if m_vdmin<0
gen activetime = (m_vdhrs*60)+m_vdmin
```

```
*/

/*jobty variable
tab m_jbnssec3_dv
gen jobty = m_jbnssec3_dv
replace jobty=. if jobty==-9
replace jobty=4 if jobty==-8
label variable jobty "-8 = unemployed, 1 = management, 2 = intermediate, 3 = routine"
label define jobty 1 "Management" 2 "Intermediate" 3 "Routine" 4"Unemployed"
label values jobty jobty
tab jobty
*/

/*gambling variable
tab m_onlgam
gen gambling = m_onlgam
replace gambling=. if gambling<0
replace gambling=0 if gambling==2
label variable gambling "0 = don't gamble, 1 = do gamble"
label define gambling 0 "don't gamble" 1 "do gamble"
label values gambling gambling
*/

/*moderate activity variable
tab m_mday
replace m_mday =. if m_mday<0
replace m_mday = 1 if inlist(m_mday, 1, 2, 3, 4, 5, 6, 7)
tab m_mday
label values m_mday m_mdaylabel
*/

*/

/*2. Tabulation of all variables
tab healthsat
tab male
tab educ
tab race
tab post
tab drinks
tab smoker
tab vegedays
tab carer
tab jobsat
tab marbi
tab jobty
tab m_mday

//Descriptive Statistics
//Descriptive Statistics
sum healthsat male educ race post drinks hrsworked othrs lnwage sleep smoker vegedays dayswalked carer jobsat marbi
jobty m_mday
*/

/*3. Regressions - Raw*
//Base regression (OLS)
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday
margins, dydx(*) atmeans
estimates store base
```

```
//Base regression (OLS) with RSEs
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans
estimates store baserob

//Probit regression
probit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans
estimates store probit_a

//Logit regression
logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans
estimates store logit_a

//Logistic regression: USE THIS FOR INTERPRETATION
logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans
estimates store logistic

//Summary Table
esttab baserob probit_a logit_a using healthsat.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2  obslast compress

//Regression with and without income satisfaction
logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans
estimates store woincsat

logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday ib(1).incsat,
vce(robust)
margins, dydx(*) atmeans
estimates store wincsat

esttab woincsat wincsat using incomesat.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2  obslast compress
*/

/*4. Diagnostic tests (base)*


//Testing VIF Values (to determine whether to use RSEs)
*Base*
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday
estat vif

//VIF = 4.78, we have an average < 10, however there are high VIF values for age and age^2 (due to their collinearity
between each other)

//Test for heteroskedasticity (Breusch-Pagan Test)
*Base*
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday
estat vif
```

```
hettest

//chi2 = 0.0000.. and so our regression is heteroskedastic, but to mitigate any issues of homoskedasticity, use robust standard errors for logit and probit regression

//Testing the validity of count R-Squared value
quietly logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday, vce(robust)
margins, dydx(*) atmeans

*Correctly classified = 77.28%
*Total Frequency (Count R-Squared):
di (3428/4477)*100
*= 76.57%
*76.57 < 77.28 and so our Count R-Squared is significant
*/

/*5. Constructing the predicted values range for LPM regression
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 1, vce(robust)
estimates store baserobsecondary

predict baserobeduc
histogram baserobeduc, xline(0 1)
sum baserobsecondary

reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 2, vce(robust)
estimates store baserobundergrad

predict baserobeduc
histogram baserobeduc, xline(0 1)
sum baserobundergrad

reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 3, vce(robust)
estimates store baserobpostgrad

predict baserobpostgrad
histogram baserobeduc, xline(0 1)
sum baserobeduc
*/

/*6. Regressions (sub samples)
//For Secondary:
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 1, vce(robust)
estimates store baserobsec

probit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 1, vce(robust)
margins, dydx(*) atmeans
estimates store probitsec
```

```
logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 1, vce(robust)
margins, dydx(*) atmeans
estimates store logisticsec

esttab baserobsec probitsec logisticsec using healthsatsec.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2 obslast compress


//For Undergraduate:
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 2, vce(robust)
estimates store baserobug

probit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 2, vce(robust)
margins, dydx(*) atmeans
estimates store probitug

logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 2, vce(robust)
margins, dydx(*) atmeans
estimates store logisticug

esttab baserobug probitug logisticug using healthsatug.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2 obslast compress


//For Postgraduate:
reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 3, vce(robust)
estimates store baserobpg

probit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 3, vce(robust)
margins, dydx(*) atmeans
estimates store probitpg

logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ == 3, vce(robust)
margins, dydx(*) atmeans
estimates store logisticpg

esttab baserobpg probitpg logisticpg using healthsatpg.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2 obslast compress


//Male Regressions
logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if male == 1, vce(robust)
margins, dydx(*) atmeans
estimates store malehealth
```

```
//Female Regression
logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage sleep
ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if male == 0,
vce(robust)
margins, dydx(*) atmeans
estimates store femalehealth

esttab malehealth femalehealth using healthgender.rtf, replace star (* 0.1 ** 0.05 *** 0.01) se mtitles r2 pr2 obslast
compress


*/

/*7. Diagnostic tests (subsamples)
*educ variable
quietly reg healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ ==
//"1,2,3"
estat vif
hettest


quietly probit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ ==
//"1,2,3", vce(robust)
estat classification
di (1306/1763)*100 //Secondary
*70.49% sig (< 70.70% - Correctly Classified)

di (1248/1609)*100 //UG
*71.62% sig (< 72.99% - Correctly Classified)

di (801/1011)*100 //PG
*71.62% insig (< 72.93% - Correctly Classified)


quietly logistic healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday if educ ==
//"1,2,3", vce(robust)
estat classification
di(1306/1763)*100 //Secondary
*67.85% sig (< 70.73% - Correctly Classified)

di (1248/1609)*100 //UG
*77.56% sig (< 78.12% - Correctly Classified)

di (801/1011)*100
*79.23% sig (< 79.33% - Correctly Classified)

*/

/*8. Predicted Values Range (LPM's)
*Secondary
predict baserobsec
histogram baserobsec, xline(0 1)
sum baserobsec
*UG
predict baserobug
histogram baserobug, xline(0 1)
```

```
sum baserobug
*PG
predict baserobpg
histogram baserobpg, xline(0 1)
sum baserobpg
*/

/*9. Visualisation and Tabulations*
*ECDF Cumulative Distribution Graphs
predict baserob
cumul baserob, gen(base_cumu)
line base_cumu baserob, sort


predict probit_a
cumul probit_a, gen(probit_cumu)
line probit_cumu probit_a, sort


predict logit_a
cumul logit_a, gen(logit_cumu)
line logit_cumu logit_a, sort


*Bar chart of distribution of Health satisfaction variable
graph bar (count), over(healthsat)



//Relationship between age and health satisfaction
quietly logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday,
vce(robust)
margins, at(age=(16(5)101))
marginsplot


//Relationship between income and healthsatisfcation
quietly logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday,
vce(robust)
margins educ, at(lnwage=(0(2)12)) atmeans
marginsplot


//Relationship between gender and health satisfaction
quietly logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday,
vce(robust)
margins male, at(age=(16(5)101))
marginsplot


//Relationship between gender and health satisfaction with age (w/interaction term)
quietly logit healthsat c.age##c.age ib(1).male ib(1).educ ib(1).race ib(6).post ib(1).drinks c.hrsworked c.othrs lnwage
sleep ib(1).smoker ib(3).vegedays c.dayswalked ib(1).carer ib(1).jobsat ib(0).marbi ib(1).jobty ib(0).m_mday
i.male#c.age, vce(robust)
margins male, at(age=(16(5)101))
marginsplot
```