The Knowledgeable Patient

Investigating the role of lay knowledge in the production of health

Martin Karlsson*
Oxford Institute of Ageing

November 29, 2007

Abstract

In this paper, we analyse the role of knowledge in an individual's production of health. We investigate the importance of various factors which potentially influence an individual's decision to seek health related knowledge, and then estimate what impact such knowledge has on individual health. Using a cross-sectional dataset of middle-aged and old UK residents, we estimate a bivariate probit model which accounts for the possible endogeneity of knowledge in the health production function. We find that knowledge has a strong effect on self-assessed health, and that this variable tends to dominate other typical covariates with health, such as education and gender. Furthermore, we find that the parameter estimates for individuals with chronic conditions are systematically different from those of others, but no results are qualitatively different in these two subgroups.

JEL Classification C5, I12

Keywords health production, health knowledge

*martin.karlsson@ageing.ox.ac.uk. The author gratefully acknowledges helpful comments on previous versions of the paper from Bleddyn Davies and George Leeson, as well as seminar participants at the Health Economics Research Centre, University of Oxford; the Department of Economics, City University and the Faculty for Actuarial Science and Insurance at Cass Business School. The dataset analysed in this paper was made available by the Picker Institute. The research was funded by a first grant from the ESRC.

1 Introduction

For various reasons, lay health knowledge enjoys increasing interest as a key factor in the maintenance of an individual's health. Over the last few decades we have witnessed gradual changes in the relationship between doctors and patients; a transition which has been branded the demise of a "golden age of doctoring" (McKinlay and Marceau, 2002). Accordingly, patients are no longer passive recipients of expert advice from physicians, but tend to take on an active role in seeking out information concerning their condition and available treatments (cf. Coulter, 2002). This trend is probably partly due to an increase in the level of education in the population, but the epidemiological transition (i.e. the increasing importance of degenerative diseases, cf. Omran, 1971; Olshansky and Ault, 1986) has also, presumably, made it more worthwhile for the patient to build up a knowledge base. Furthermore, the emergence of new information technologies has significantly reduced the cost of acquiring health related information.

The overall objective of this paper is to investigate the role of knowledge and education in the individual's production of health. Drawing on the theoretical perspectives of Grossman's (1972) model of production of health, we seek to analyse empirically the decisions people take in the maintenance of their own health. We aim to analyse, firstly, how people obtain health knowledge to guide that process, and secondly, how this health knowledge contributes to their general health. For this purpose, we use a household survey conducted by the Picker Institute. To date, not much research has been done in this field – despite its importance for public policy. Economic theorising on the topic also remains in its infancy.

Changes in the general health knowledge of patients and their self-efficacy in obtaining and using this knowledge, have given rise to a great degree of optimism concerning what patient education can do for public health. One example is the Expert Patient Programme in the UK, where patients with chronic conditions participate in lay-lead courses which help them manage their own condition. Despite great expectations, there is a lack of systematic assessments of the actual effects of this type of programmes. Besides, many of the evaluations that do exist seem to suffer from a poor research design (Taylor & Bury, 2007). However, our analysis could enable policy makers to assess what can be expected from patient education programmes.

It must be kept in mind, however, that some patients run the risk of being left behind when the emphasis on patient knowledge increases. For example, surveys suggest that older patients are much less likely to challenge the knowledge of physicians (Chapple et al, 2002; Angus et al, 2005; Beaver et al, 2007). Furthermore, it is very likely that individuals who are marginalised in other respects also have worse access to health information (Ellins & Coulter, 2005). By means of illustration, *Table 1* shows that the use of internet as a source of information is much more widespread in the highest social grade (AB) than in the lowest (DE). This means that an increased reliance on patient knowledge in health policy could potentially sharpen inequalities. A randomised controlled trial can assess the effects of patient knowledge on health, but it does not address the issue of why patients seek for knowledge. By studying the actual behaviour of individuals, we can explain their knowledge seeking behaviour with reference to education, age and socioeconomic status. Hence, one of our main aims in this paper is to identify the main determinants of people's health information acquisition.

Table 1: Social Class and the Internet as Source of Health Information.

When you seek out health information how likely are you to look at Internet health websites?	**Social Class**			
	AB	C1	C2	DE
	Upper and Middle Class	*Lower Middle Class*	*Skilled Working Class*	*Working Class & Underclass*
Very Unlikely	28	26	37	47
Unlikely	21	24	24	30
Likely	22	25	19	12
Very likely	28	24	19	10
Don't Know	1	*	1	2
N	**548**	**806**	**475**	**1,021**

As mentioned initially, there is a paucity of research in this field. From our point of view, what is missing is a general economic model of why patients gather knowledge as opposed to other types of 'health investments' such as exercise, diet or seeking medical advice – and furthermore how knowledge interacts with these other inputs in the health production function. Given that these behaviours will to some degree be influenced by the individual's health, we have a problem of potential reverse causality as well. Besides, incorrect information could in fact be harmful or useless to the individual (Bessell et al, 2002; Crocco et al, 2002; Eysenbach et al, 2004). In order to address these issues, we will seek to control for endogeneity and pre-existing conditions in our estimates.

The role of individual knowledge has previously been analysed in the context of smoking (Carbone et al, 2005; Jones and Kirigia, 1999; Hsieh et al, 1996) where it was found that individuals learn from shocks to their own (but not their partner's) health and that public information campaigns can improve public health. Patient knowledge has also been analysed in the context of obesity (Tsou & Liu, 2006; Kan & Tsai, 2004) where it was found that health knowledge had an impact on body mass index amongst overweight people. Furthermore, research has shown that general literacy and health literacy (cf. Nutbeam, 2000) are linked with poor health and health outcomes (De Walt et al, 2004). A related study showed that persons with inadequate health literacy incur higher medical costs and use an inefficient mix of services (Howard et al, 2005). Finally, it has been suggested that the differences in health between Eastern and Western Europe is to some extent attributable to differences in risk knowledge (Steptoe & Wardle, 2001). However, there is as yet no unifying approach which analyses patient knowledge in the context of a behavioural model.

This paper is organised as follows. In Section 2, we give a brief summary of the economic literature – including the theoretical perspective as well as the most common empirical findings. Section 3 gives an overview of the Picker Institute survey and presents the econometric approach. Section 4 provides main estimation results and sensitivity analysis. Section 5 concludes and points out important issues for future research.

2 Theoretical Perspectives

A natural starting point for research on knowledge in the production of health is Grossman's (1972) model of health. Drawing on human capital theory, Grossman suggested that health is to be treated as a capital stock which determines the future choice set available to the individual, but also as a 'consumption good' which directly contributes to individual well-being. In the spirit of household economics, it is assumed that the individual uses time and material resources to produce 'health investments', which counteract the gradual deterioration of health that comes with age.

Grossman's model has since been extended in various directions, allowing for, amongst other things, uncertainty in the deterioration of health and endogenous 'choice' of length of life. The model has also been subjected to empirical testing using a wide range of techniques and datasets. For an overview, see Grossman (2004).

In Grossman's model, it is assumed that education increases the productivity of the individual's health investments. This is well in line with the well established empirical observation of a very strong relationship between educational attainment and health (cf. Fuchs, 2004). However, the theory is not very explicit on how education increases the impact of health investments and, quite naturally, this complicates the empirical analysis. In the literature, several different pathways between education and health have been suggested. Cutler and Lleras-Muney (2006), as well as Chandola et al (2006), consider a host of possible explanations:

- **Reverse causality** (e.g. children in poor health obtaining less schooling)
- **Confounding factors** (e.g. family background, genentic traits and other factors which can be expected to affect both health and education)
- **Income and access to health care** (individuals with higher education can afford to invest more in their health and might also have better access to health care)
- **Work environment** (for example if people with higher education work in less risky and physically demanding jobs)
- **Valuation of the future** (i.e. educated people can look forward to more pleasant future life and hence have stronger incentives to extend life)
- Better **access to information** and superior **cognitive skills** (e.g. higher ability to find relevant information or comply with complicated medical prescriptions)
- **Preferences** (e.g. education might influence people's time discount rates, which would in turn make them more likely to invest in their health)
- **Social rank** (i.e. education can be expected to change a person's relative position in society, and social rank might in itself affect health)

- Access to **social networks** (educated people have larger social networks, and social networks are good for health)

Cutler and Lleras-Muney (2006) conclude that none of these factors is likely to dominate but that all or most of them can explain some of the correlation between education and health. In particular, the authors are able to discard reverse causality and preferences, whereas on the other hand the importance of income and labour market conditions is well documented. They also conclude much less is known concerning the importance of rank, social networks and health related knowledge.

Our research has the potential to fill in one of the gaps in Grossman's theory, namely the role of knowledge in the production function. In the literature, there is a discussion as to whether education increases *allocative efficiency* (the combination of inputs used in the production of health) or *productive efficiency* (the amount of health 'produced' for a given combination of inputs). Knowledge is an input in itself, but it also has the potential to help the individual increasing the allocative efficiency. In order to develop a general theory of knowledge and health, it is furthermore important to assess whether different types of knowledge have a differential impact on an individual's health, and also whether one type of knowledge has different effects on different types of individuals (e.g. those with and without chronic conditions). We seek to address at least the second of these issues in this paper.

It is important to mention that insights could also be gained from perspectives brought in from other disciplines, which have addressed issues related to health and knowledge in some depth already. There are two natural candidates. Firstly, there is the *Patient Activation Measure* developed by Judith Hibbard, Professor of Health Policy at the University of Oregon (Hibbard et al, 2004). The patient activation measure focuses on self-efficacy in the patient's attitude to her own health, and recognises four different stages of patient awareness, which are supposed to contribute to an individual's health. Secondly, the psychological literature has suggested at least three different models to explain individual behaviour concerning health: the health belief model, which focuses on the perceived benefits of a certain behaviour (Rosenstock, 1966, 1974), the protection motivation theory, which focuses on fear appraisal (Rogers, 1975; Prentice-Dunn & Rogers, 1986), and the theory of planned behaviour (Ajzen, 1988).

In this research paper, we seek answers to some of the most relevant questions concerning the role of lay knowledge in the production of health. In particular, we analyse the following four issues:

1. Which variables influence the individual's choice to seek health knowledge?

2. How does health knowledge contribute to an individual's health?

3. Is there a reverse causality, so that individuals experience negative shocks to their health are more likely to seek health knowledge?

4. Does health knowledge have a differential impact on people with chronic and non-chronic conditions?

In the next section we outline the methodological approach used to analyse these issues.

3 Econometric Considerations

3.1 The Dataset

We use a dataset which has been provided by the Picker institute (cf. Ellins & Coulter, 2005). This dataset contains nearly 3,000 individuals aged 45 and older who have been approached through a telephone survey, covering a representative sample of the UK population in that age group. Respondents were asked more than 100 questions concerning their general health, chronic conditions, and their attitudes towards the maintenance of their own health, together with demographic and socioeconomic variables such as age, gender, ethnicity, income and education.

The dataset contains a wide range of health-related variables: self-assessed health (SAH); a long list of chronic conditions; and limitations to some specified daily activities. In this paper we use self-assessed health, which is arguably the most general of the various variables available. However, the prevalence of chronic conditions will prove useful in our attempt to control for previous health and hence control for confounding factors in our parameter estimates. There are two SAH variables in the dataset: one of them representing general health and the other specified to be the health status in the last four weeks. As indicated in *Table 2* there is some discrepancy between the two measures, and hence the general health variable was used for sensitivity analysis.

Table 2: Distribution of self-reported health.

	Self-reported health		**Past 4 weeks**	
	Frequency	*Percent*	*Frequency*	*Percent*
Poor	305	10.57	246	8.51
Fair	496	17.16	457	15.83
Good	964	33.36	767	26.53
Very good	735	25.43	833	28.84
Excellent	390	13.49	586	20.27
Total	2,889	100.00	2,889	100.00

Using self assessed health is obviously connected with the problem of potential individual heterogeneity in reporting behaviour. Different individuals may have different understandings of what a certain degree of health implies and this would probably bias the results. On the other hand, self-assessed health has been shown to be strongly correlated with more 'objective' measures of health (cf. Hernández-Quevedo et al, 2005) and has also been found to be a very good predictor of mortality (Huisman et al, 2007).

Concerning health related knowledge, we used the variable derived from the question *"How likely are you to seek out information to learn about how to cope with the problems of your health?"*. This question was asked to all respondents, irrespective of how they rated their own health or which diagnoses they reported. Individuals were asked to rate the probability in four steps – from *'Very Unlikely'* to *'Very Likely'*. *'Not applicable or don't know'* was allowed as a further alternative, although very few individuals (3.2 per cent) chose this option. This question is not the only one in the survey which relates to health knowledge – but the only one that is generic. All the other knowledge-related questions relate to specific situations such as a visit to the GP or a hypothetical visit to a hospital. These other variables were considered as possible instruments for the general knowledge variable instead.

For estimation, we dichotomised the knowledge variable, with *'Very likely'* (36.7 per cent of the responses) coded as 1 and all other answers coded as 0. In *Table 3*, we provide some summary statistics of the knowledge variable, as well as the likelihood to use a telephone helpline or the internet as a source. Obviously, relatively few of those who reported they were likely or very likely to seek health information would use any of these two sources.

Table 3: Likelihood to seek health related information from various sources.

	Seek Information		**Helpline**		**Internet**	
	Frequency	*Percent*	*Frequency*	*Percent*	*Frequency*	*Percent*
Very Unlikely	150	5.18	807	27.94	920	31.86
Unlikely	415	14.37	946	32.75	669	23.16
Likely	1173	40.60	736	25.47	500	17.31
Very Likely	1060	36.68	284	9.84	496	17.16
Not available			18	0.62	303	10.50
Don't know	91	3.16	98	3.38		
Total	2,889	100.00	2,889	100.00	2,889	100.00

In *Table 4*, we give an overview of some of the independent variables used in our analysis.

Table 4: Definition and mean of variables.

Variable	**Mean**	**Definition**
Age	61.07	Age in years in 2004
Male	0.38	Gender
Cohabiting	0.54	Married/living with partner
NPeople	2.12	Number of people in household
Edu1	0.55	Education: Secondary
Edu2	0.24	Education: Post secondary
Edu3	0.21	Education: Degree(s)
White	0.96	Ethnic: White
Conditions	0.72	Suffering from any chronic condition

The variables are largely self-explanatory. However, the high mean recorded for the variable 'Conditions' might require some further explanation. The dataset considers a wide range of chronic conditions. Four conditions have particularly high prevalence: *arthritis and rheumatism* (30 per cent), *hypertension* (29 per cent), *high cholesterol* (19 per cent) and *chronic pain including back pain* (19 per cent). Furthermore, 66 per cent suffer from more than one condition, whereas 28 per cent have none. The fact that 70 per cent of the sample suffer from some type of chronic condition looks very high. In order to get an estimate of whether the prevalence rates in the survey are representative for the general population, we compared them with results from the English Longitudinal Survey of Ageing (Marmot et al, 2003), which has a similar scope in terms of age albeit covering the English population only. A comparison for some conditions which are included in both datasets is provided in *Table 5*.

Overall, the total prevalence rates are similar in the two datasets, and it is difficult to identify a general bias in the Picker Institute data. For some conditions – such as angina and diabetes – our dataset seems to have slightly higher prevalence rates than ELSA, whereas the opposite holds for stroke and hypertension. However, the general agreements between the two surveys actually suggests that the Picker Institute survey provides us with a representative sample of the population as far as chronic conditions are concerned.

Table 5: Prevalence of chronic conditions in the dataset and in the ELSA survey.

Condition	Dataset	Gender	<55	55-59	60-64	65-69	70-74	75-79	80+	Total
Angina	ELSA	*Male*	*3.3%*	*7.9%*	*11.7%*	*11.7%*	*14.5%*	*20.8%*	*19.1%*	11.1%
		Female	*1.6%*	*2.9%*	*5.9%*	*8.7%*	*13.3%*	*12.7%*	*16.8%*	8.1%
	Picker	*Male*	*6.2%*	*7.4%*	*16.4%*	*18.8%*	*26.7%*	*25.4%*	*29.2%*	13.2%
		Female	*3.1%*	*7.7%*	*12.4%*	*13.1%*	*18.3%*	*24.4%*	*26.9%*	10.7%
Stroke	ELSA	*Male*	*1.2%*	*2.2%*	*3.8%*	*5.5%*	*6.8%*	*7.8%*	*13.3%*	4.8%
		Female	*1.0%*	*1.0%*	*2.3%*	*3.8%*	*5.9%*	*5.0%*	*9.3%*	3.7%
	Picker	*Male*	*0.7%*	*2.6%*	*3.3%*	*1.8%*	*1.0%*	*3.2%*	*4.2%*	1.8%
		Female	*0.6%*	*2.7%*	*2.8%*	*4.5%*	*2.2%*	*4.9%*	*5.8%*	2.5%
Hypertension	ELSA	*Male*	*25.8%*	*32.8%*	*37.5%*	*40.0%*	*44.3%*	*44.5%*	*38.5%*	36.1%
		Female	*24.7%*	*31.7%*	*35.4%*	*43.3%*	*49.1%*	*50.8%*	*46.7%*	38.8%
	Picker	*Male*	*21.9%*	*27.0%*	*35.5%*	*44.6%*	*40.0%*	*46.0%*	*18.8%*	30.0%
		Female	*16.0%*	*28.5%*	*37.2%*	*37.9%*	*46.1%*	*46.3%*	*40.4%*	30.1%
Diabetes	ELSA	*Male*	*4.6%*	*6.9%*	*8.1%*	*9.9%*	*13.0%*	*12.3%*	*9.7%*	8.5%
		Female	*2.1%*	*5.4%*	*5.7%*	*6.4%*	*9.9%*	*9.4%*	*6.5%*	6.1%
	Picker	*Male*	*5.5%*	*10.1%*	*9.2%*	*14.3%*	*21.9%*	*25.4%*	*8.3%*	10.6%
		Female	*3.1%*	*8.4%*	*13.2%*	*9.6%*	*12.2%*	*6.5%*	*8.7%*	7.6%

In order to take the differences between different conditions into account, while at the same time keeping the number of independent variables at a workable level, we introduced separate dummies for the four most prevalent chronic conditions mentioned above, and then added an additional dummy for individuals suffering from any other condition, but none of these four.

3.1.1 Sample stratification and self-selection

According to Ellins & Coulter (2005), the response rate in the survey was as low as 18.6 per cent. The authors point out that, to some extent, the reported response rate is artificially low, since some of the households which refused to participate would have been ineligible due to the cut-off age anyway. However, the response rate is so low that it would be desirable to assess whether non-random participation in the survey might have affected the results. Since we do not have information on the individuals who chose not to participate – apart from general population statistics, whenever available – the possibilities to address problems related to self-selection are limited.

Response rates by age, gender and region are provided in *Figure 1*. On the vertical axis, we have the ratio between the response rate within a certain demographic subgroup and the overall response rate (which was around 11.8 per 100.000 population in the relevant age bands). Obviously, response rates were to some degree biased due to the stratified sampling procedure, and this is why all Northern Ireland respondents (apart from the very old) have a significantly higher relative response rate than England. *Figure 1* suggests that the differences in response rates attributable to region are of a much greater importance than differences between age groups, or between men and women.

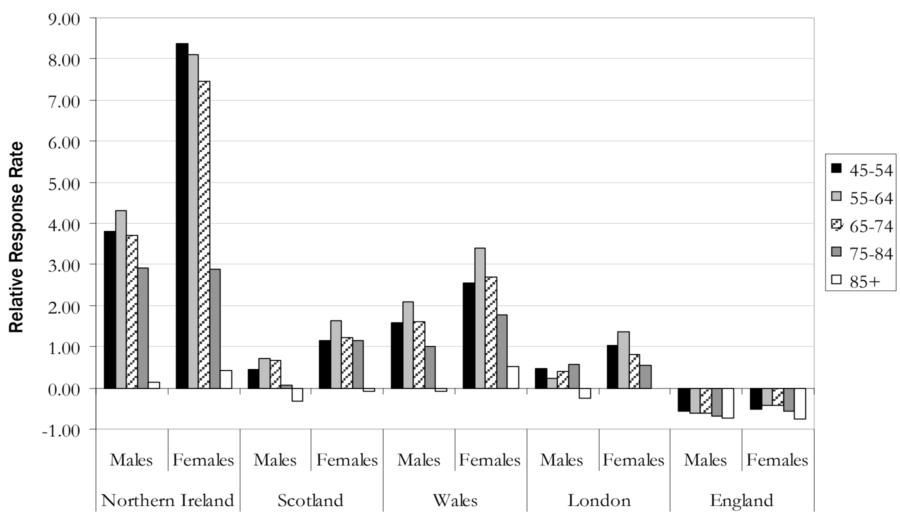

Figure 1: Relative response rates by age, gender and region.

It is of course a possibility that the decision to participate in the survey is somehow influenced by unobservable characteristics which are correlated with the error terms in our estimating equations. However, most techniques available to deal with this problem require that personal characteristics of non-participating individuals are known, and furthermore do not work well when some explanatory variables are endogenous (Wooldridge, 2002). Hence, our modelling approach will have to rely on the assumption that

$$D(\mathbf{y}|\mathbf{x}, s) = D(\mathbf{y}|\mathbf{x})$$

or in words: that the distribution of the dependent variables ($\mathbf{y}$) conditioning on the explanatory variables ($\mathbf{x}$) and participation in the sample ($s_i = 1$) equals the distribution of $\mathbf{y}$ conditioning on $\mathbf{x}$ only. This will be the case whenever s is a non-random function of $\mathbf{x}$, or when s is independent of $(\mathbf{x}, \mathbf{y})$.

However, since the sample has been obtained using standard stratification techniques – where stratification is based on region – we can actually improve the performance of the maximum likelihood estimator by applying the appropriate weights to each individual observation. The weighted exogenous sample MLE (WESMLE) will use weights defined as $w_j = Q_j/H_j$, where Q_j is the population frequency of region j and $H_j = N_j/N$ is the proportion of the sample in region j.

3.1.2 Estimator: Bivariate Probit

Given the nature of the dependent variables we are interested in, we estimate a bivariate model which is an ordered probit in the health part and a standard probit in the knowledge part. Hence, the model consists of the two estimating equations

$$y_h = \begin{cases} 1 & if \quad \alpha_0 < \mathbf{x}_h \boldsymbol{\delta}_h + \beta y_k + u_h \leq \alpha_1 \\ 2 & if \quad \alpha_1 < \mathbf{x}_h \boldsymbol{\delta}_h + \beta y_k + u_h \leq \alpha_2 \\ 3 & if \quad \alpha_2 < \mathbf{x}_h \boldsymbol{\delta}_h + \beta y_k + u_h \leq \alpha_3 \\ 4 & if \quad \alpha_3 < \mathbf{x}_h \boldsymbol{\delta}_h + \beta y_k + u_h \leq \alpha_4 \\ 5 & if \quad \alpha_4 < \mathbf{x}_h \boldsymbol{\delta}_h + \beta y_k + u_h \leq \alpha_5 \end{cases}$$

and

$$y_k = 1\left[\mathbf{x}_k \boldsymbol{\delta}_k + u_k \geq 0\right]$$

where $\boldsymbol{\alpha} = \begin{pmatrix} -\infty & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \infty \end{pmatrix}'$ are the cut-off points between different health states, $\mathbf{x}_i$ are the vectors of explanatory variables used in the two estimating equations. For the error terms u_i, we assume

$$\begin{pmatrix} u_h \\ u_k \end{pmatrix} \backsim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}.$$

Some variables appear in both estimating equations. These are: *Age* and *Age²*, *Sex, Edu2* and *Edu3* (post-secondary education and degrees, respectively, having secondary education as reference category), *Ethnic* (dummy taking on the value 1 for white individuals), *Cohabit* (married or living with partner), *NPeople* (number of people in household), and dummies for various chronic conditions.

Furthermore, we make use of exclusion restriction in order to increase the robustness of the estimates. Hence, we have included a couple of further variables in the knowledge equation which can be expected to be good instruments for knowledge. These are both question related to the person's self-efficacy in the maintenance of their own health. The first question we use – *Action* – is based on the respondent's reaction to the statement *'I am confident that I can take actions to prevent or minimize problems associated with my health condition'*. The second variable – *Qualifications* – is based on the statement *'If I went to a new doctor, I would find out as much as I could about his or her qualifications'*. The first of the two variables could be motivated from the alternative theories mentioned in Section 2 (particularly the Patient Activation Measure, which focuses on self-efficacy). The second of the two variables has the desirable property that it is based on a hypothetical situation, and thus answers are less likely to be affected by the individual's current health status. Since the preference for knowing a new doctor's qualifications is bound to be correlated with an individual's general preference for health related knowledge, this variable seems likely to be a good instrument for health related knowledge.

In both cases, answers range from *'Disagree strongly'* to *'Agree strongly'*, and to keep the number of parameters down, we dichotomised both by contrasting the highest response alternative ('Agree strongly') with the others. This approach seemed reasonable since the median tended to be in the second highest group. Since both variables are self-reported, just as the two dependent variables, their inclusion in our model could be expected to have the further advantage of controlling for heterogeneous reporting behaviour within the sample.

Our likelihood function is

$$L_i(\theta) = \sum_{i=1}^{5} 1\left[y_h = h\right] \begin{pmatrix} y_k \int\limits_{-\mathbf{x}_k \boldsymbol{\delta}_k}^{\infty} F\left(\mathbf{x}_h, y_k, u_k, \theta; h\right) \phi\left(u_k\right) du_k \\ +\left(1 - y_k\right) \int\limits_{-\infty}^{-\mathbf{x}_k \boldsymbol{\delta}_k} F\left(\mathbf{x}_h, y_k, u_k, \theta; h\right) \phi\left(u_k\right) du_k \end{pmatrix}^{w_{j_i}}$$

where θ is the vector of parameters to be estimated, $\phi(\cdot)$ is the density of the standard normal distribution, w_{j_i} is the weight attached to the stratum j which individual i belongs to, and

$$F\left(\mathbf{x}_h, y_k, u_k, \theta; h\right) = \Phi\left(\frac{x_h \delta_h + \beta y_k + \rho u_k - \alpha_{h-1}}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{x_h \delta_h + \beta y_k + \rho u_k - \alpha_h}{\sqrt{1-\rho^2}}\right)$$

is the probability of observing $y_h = h$, where Φ denotes the cumulative distribution function of the standard normal distribution.

However, when we use the WESMLE estimator, the covariance matrix for the estimated θ has to be adjusted to take the stratified sampling into account. Hence, the estimator used for the asymptotic variance of the parameter vector θ is (Wooldridge, 2001)

$$\widehat{Var}(\theta) = \left[\sum_{i=1}^{N} w_{j_i} \nabla_{\theta}^2 \ln L_i\right]^{-1} \left[\sum_{i=1}^{N} w_{j_i}^2 \left(\nabla_{\theta} \ln L_i - \overline{q_{j_i}}\right)' \left(\nabla_{\theta} \ln L_i - \overline{q_{j_i}}\right)\right] \times \left[\sum_{i=1}^{N} w_{j_i} \nabla_{\theta}^2 \ln L_i\right]^{-1}$$

where $\overline{q_{j_i}}$ is the average score for stratum j_i at the estimated values $\widehat{\theta}$.

3.1.3 Sensitivity Tests and Hypothesis Testing

In order to assess the robustness of our findings, we compare the baseline model outlined above with a number of alternative specifications. Firstly, we address the issue whether individuals with chronic conditions are systematically different from others by estimating separate parameters for individuals with and without chronic conditions. Secondly, we check the appropriateness of the exclusion restrictions by also estimating a full model with all explanatory variables included in both estimating equations. Thirdly, we look at the consequences of excluding knowledge from the estimates in an attempt to assess the extent to which the observed correlation between education and health is due to better knowledge.

Since all these alternative specifications involve restricting some parameters of the full model to be equal to zero, there are two alternative ways to analyse model performance – one is to look at information criteria such as the Akaike Information Criterion, and the other is to perform a likelihood ratio test. We consider both approaches, and find that they tend to be consistent with each other.

4 Results

4.1 Descriptive Statistics

Before we proceed to the model estimates, a simple cross-tabulation is useful to explore patterns in the data. In *Figure 2*, we show how the likelihood to seek health related information varies with an individual's self-assessed health. In the highest health category, more than half of respondents (58.1 per cent) report that they are very likely to seek health information, compared to less than one third (29.1 per cent) in the worst health category. This figure seems to suggest that reverse causality – in the sense that people in poor health have stronger incentives to seek information – is of limited importance in comparison with the direct effect. However, the correlation patterns are far from straightforward. For example, amongst people reporting 'very good' health, 76.0 per cent claimed to be likely or very likely to seek health information. The corresponding figure for the 'fair' category is 80.9 per cent.

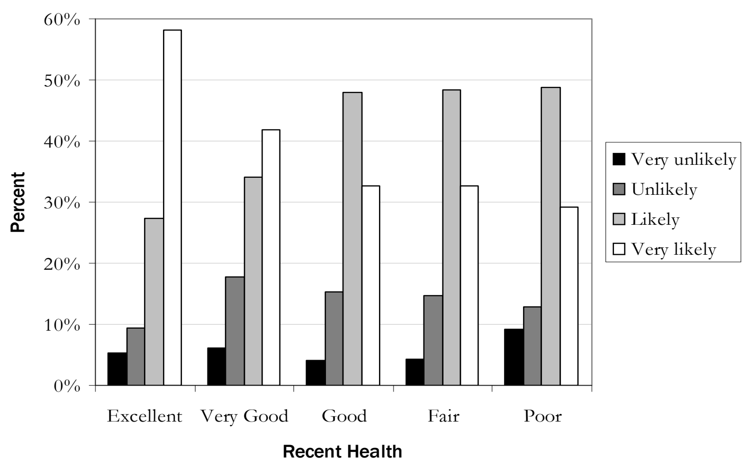

Figure 2: Likelihood to seek health information by SAH category.

The picture looks similar when recent health is replaced by general health. The percentage claiming to be 'very likely' to seek health information is now slightly less correlated with the health variable (ranging from 52.2 per cent amongst individuals in excellent health to 28.0 per cent amongst individuals in fair health; for the lowest health state the figure is 29.6). However, when the 'very likely' and 'likely' categories are merged, the correlation seems to just as weak as in the previous case.

4.2 Estimation Results

We now turn to the estimation results from the bivariate probit. We present results from the baseline model in *Table 6*. To the left, we present parameter estimates (coefficients and the corresponding p values) for the health equation – and to the right, we present the estimates for the knowledge equation. Concerning the **health** equation, it can be observed that knowledge has a remarkably strong impact on health, and this effect is also strongly significant. We also note a strong and significant negative correlation between the error terms in the two equations, as captured by the parameter ρ. This seems to suggest that the reverse causality from knowledge to health is negative, i.e., that individuals who experience bad health are more likely to seek health knowledge.

Concerning the usual suspects – age, gender, education, ethnicity and cohabitation – we note that they seem to have less explanatory power than in most other studies. Age is a significant determinant of health,[1] whereas education and gender come out insignificant. On the other hand, all the chronic conditions have the expected sign and are strongly significant. However, it should be noted that the parameter estimated for the effect of health knowledge is higher (in absolute terms) than the coefficients of some of the chronic conditions listed. Quite remarkably, the effect of health knowledge seems to be stronger than the effect of suffering from hypertension or cholesterol – and similar to the effect of suffering from arthritis. Hence, our results actually suggest that one important mechanism through which the control variables affect health is better access to knowledge. We will return to this issue in the sensitivity analysis below.

Another interesting finding is that the number of people in the household is significant and positive, whereas the cohabitation status has no significant influence on health. This finding seems somewhat at odds with the common result that the effect of cohabitation depends on the quality of the relationship.

In the **knowledge** equation, age, gender, education and cohabitation all come out significant, whereas ethnicity and household size seem to be of less importance. Interestingly, chronic conditions do not seem to increase people's tendency to seek health knowledge – rather to the contrary, hypertension, arthritis and the 'other' category significantly *reduce* an individual's propensity to seek health information. This finding clearly requires some further investigation. One possible explanation would be that there is persistence in the knowledge variable and that these chronic conditions have evolved as a consequence of lack of knowledge in previous periods. However, a more plausible explanation might be that the conditions in question reduce the individual's ability to seek health knowledge. Obviously, a longitudinal dataset would be required to analyse these issues in further detail.

Finally, the two instrumental variables (*'action'* and *'qualifications'*) come out positive and strongly significant, suggesting that an individual's self-efficacy matters beyond what is captured by demographic and socioeconomic factors – thus lending some support to the alternative models of knowledge acquisition

[1] Although strongly significant, age has only a limited effect on health within the age ranges included in the sample.

behaviour outlined in Section 2.

Table 6: Estimation Results, Baseline Model.

	Health		Knowledge	
Variable	*Estimate*	*pVal*	*Estimate*	*pVal*
Knowledge	0.679	0.00		
Age	9.366	0.00	0.250	0.02
Age^2	-6.386	0.00	-0.881	0.00
Male	0.067	0.23	-0.276	0.00
EDUpostsec	-0.001	0.99	0.189	0.02
EDUdegree	0.125	0.14	0.567	0.00
White	0.050	0.75	-0.020	0.90
Cohabit	0.066	0.28	0.210	0.02
NPeople	0.078	0.00	0.007	0.80
Hypertension	-0.401	0.00	-0.196	0.02
ChronicPain	-0.969	0.00	-0.102	0.25
Arthritis	-0.690	0.00	-0.125	0.00
Cholesterol	-0.375	0.00	0.115	0.20
Other	-0.703	0.00	-0.216	0.03
Action			0.494	0.00
Qualifications			0.428	0.00
a_1	1.448	0.00	-0.391	0.01
a_2	2.303	0.00		
a_3	3.193	0.00		
a_4	4.155	0.00		
ρ	-0.292	0.00		
N	2,811			
Logl	-5,413			
AIC	10,895			
BIC	11,103			

4.3 Hypothesis Testing

In the previous section, we found that knowledge seems to be of great importance in the individual's production of health. Furthermore, it was found that the presence of chronic conditions affects both the individual's general health and their acquisition of health knowledge. Hence, as mentioned initially, it can be hypothesised that individuals suffering from various chronic conditions have different incentives to acquire health knowledge than those who do not. In a first attempt to test this hypothesis, we estimated a model with separate parameter estimates for the two groups. Results are presented in *Table 7*. The top half of the table is identical to *Table 6*, and contains the joint estimates for the entire sample. The lower part of the table contains separate estimates for individuals who report having at least one chronic condition.

In the health equation, we find that the overall results are fairly similar to the baseline specification: Knowledge, age and chronic diseases all have a strong impact on health and the point estimates are very similar. Also in the knowledge equation the parameter estimates are similar. However, the estimates for postsecondary education and cohabitation both become insignificant in this specification.

Concerning the parameter estimates for the differences between the two groups, we are unable to reject the hypothesis that knowledge has the same impact on health in the two groups. Indeed, only the parameters related to age seem to be different between the two. On the knowledge side, we note that individuals with chronic conditions seem to exhibit different age patterns in their knowledge acquisition, but also the effects of ethnicity seems to be different in the two groups.

Concerning the overall performance, there is some evidence to suggest that this model is preferred over the baseline model. The Akaike Information Criterion is 10,852 compared with 10,895 in the baseline specification. On the other hand, the BIC, which punishes the inclusion of extra parameters harder, takes on a value of 11,173 compared with 11,103 in the baseline model. Hence, according to that criterion the baseline model is preferred. Finally, a likelihood ratio test comparing the two models delivers a test statistic of 81.35, which is strongly significant ($p < 0.01$) at 19 degrees of freedom. In conclusion, we have found that although there is no significant difference between the two groups in the effect of knowledge on health, the parameters differ systematically between them in other respects.

4.4 Sensitivity Analysis

In this subsection, we present a number of alternative specifications, and compare their performance with the baseline model presented initially.

4.4.1 Including Instruments

A first change we consider is to include all explanatory variables in both equations. Results are presented in *Table 8*. In this specification, the two instrumental variables (*'action'* and *'qualifications'*) come out insignificant in the health equation, which seems to justify their exclusion in the baseline specification. In general, the parameter estimates in the full model are very similar to our baseline model. The point estimate for knowledge is somewhat higher than in the baseline specification. However, this specification does not perform as well as the baseline according to most indicators: a likelihood ratio test delivers a value of 5.51 which at 2 degrees of freedom is insignificant at the 5 per cent level and hence lends support for the baseline model. The BIC takes on a value of 11,114 which is higher than the baseline value of 11,103. The AIC is slightly lower in this specification, however.

4.4.2 Excluding Knowledge

Next, we analysed how the inclusion of the knowledge variable affects the parameter estimates of other variables. We have seen in the baseline specification that education and gender seem to be insignificant in the health equation, despite being very well known determinants of health. If health related knowledge is the main mechanism through which education affects health, we would expect the estimated education effects to become significant when the knowledge variable is omitted in the health function. If on the other hand education and gender remain insignificant in such a specification, we would instead conclude that the dataset is of poor quality, or that the baseline model has been poorly specified.

In *Table 9*, we compare the parameter estimates for the **health** equation in three different specifications. The two leftmost columns present results for the baseline model. The two middle columns present results for a model where the correlation between the error terms (ρ) has been suppressed. The two rightmost columns represent a specification where we have furthermore omitted knowledge in the health equation.

In general, we seem to get a worse fit when the knowledge variable is ignored. Some variables which turned out significant in the baseline specification now lose their significance - such as age, household size and one of the chronic conditions (arthritis). However, for education, the estimates move in the opposite direction: when knowledge is excluded from the model, both education parameters become significant. Hence, for university degrees we observe a point estimate of 0.29 in this alternative specification, compared with 0.13 in the baseline model. Also the estimate for other forms of postsecondary education is significant, although the point estimate is much lower than for university degrees. Similarly, cohabitation and gender – which were insignificant in the baseline specification – are now significant and with the expected signs. Hence, our results seem to suggest that health knowledge is an important mechanism through which education affects health and that the independent effect of education might be smaller than previously believed.

4.4.3 General health instead of recent health

Finally, we considered an alternative specification with general health - instead of recent health - as dependent variable. There are two reasons to believe that a specification with general health would fare worse than the corresponding model with recent health. Firstly, and most importantly, the endogeneity problem is likely to be less pronounced when recent health is considered. Secondly, the question concerning general health was asked at the beginning of the survey whereas the recent health question was asked after a host of very specific health-related questions had already been answered. This might imply higher reliability of the latter due to, for example, less recall problems or a reduced likelihood for individual reporting behaviour heterogeneity thanks to the framing that these other health-related questions provide. Furthermore, it could be expected that the actual relationship between knowledge and health is weaker when the general health is considered instead of recent health.

For the estimation of general health, the correlation coefficient ρ actually turned out insignificant and all available statistics suggested that the specification without a correlation parameter was to be preferred. Hence, we present results from such a specification in this case. Results are given in *Table 10.* Again, we find that knowledge has a strongly significant positive effect on health, but the point estimate is, just as expected, much lower. Furthermore, education and chronic diseases have a strong effect on health, whereas age and gender seem to be of limited importance. Also in the knowledge equation, education comes out strongly significant as well as cohabitation. Just as in the baseline specification, hypertension and 'other' chronic diseases seem to reduce an individual's likelihood to acquire health knowledge.

5 Conclusion

The aim of this paper is twofold: firstly, to explain individual behaviour in the decision to seek health related knowledge, and secondly, to analyse the effect of such individual knowledge on health. A further aim was to analyse to what extent health knowledge has a different impact on different types of individuals, and, in particular, whether individuals with chronic conditions benefit more from health related information than others. The latter issue is of great relevance given the current emphasis on lay-led education in the health policy of many developed countries. Our analysis has delivered a host of remarkable findings, which are listed below.

Firstly, we find that there is a strikingly strong relationship between knowledge and health. This effect is already positive and strongly significant when we disregard the possible endogeneity of the knowledge variable, but when we account for this problem the coefficient increases even further. Thus, the positive health effect of being 'very likely' to seek health information is comparable with the negative effect of suffering from arthritis, and it is actually higher – in absolute terms – than the effects of suffering from hypertension or elevated cholesterol.

Secondly, chronic conditions do not have the expected effect on the propensity to seek health knowledge. Indeed, individuals with arthritis or 'other' chronic diseases were estimated to be significantly *less* likely than others to seek health knowledge. This finding has at least two possible explanations: one is that these conditions evolve partly due to a previous lack of health knowledge, and that the parameter estimates pick up this effect. The alternative explanation is that individuals suffering from these conditions have their ability to seek health knowledge impeded. Clearly, the simple model used here does not enable us to identify the relative importance of these two explanations.

Thirdly, individuals suffering from chronic conditions are systematically different from those not suffering from any conditions, and that the parameters of the model should be estimated separately for the two groups. However, there are no remarkable qualitative differences between the two subsamples, so most conclusions hold for both. Besides, we found that there was no significant difference between the two groups in the estimates of the effect of knowledge.

Furthermore, knowledge seems to dominate some of the other covariates we consider in the production of health. This is of course the most remarkable finding in this study. In particular, it is interesting to note that the otherwise very significant relationship between education and health vanishes completely when knowledge is included in the model. The finding seems to suggest that the main channel through which education affects health is through superior health related knowledge. Such a result is apparently very informative towards the economic theorising on the production of health.

We found that our baseline model, albeit simple, performed reasonably well in comparison with some alternative specifications. However, results should be interpreted with some caution since the dataset has some obvious limitations. Firstly, all information we rely on is self-reported. This might be of less importance as far relatively 'objective' things such as educational attainment or chronic conditions are concerned, but in more subjective issues such as general health or likelihood to seek health information, there is a clear risk that reporting behaviour correlates with other personal attributes, thus biasing the results. Furthermore, the non-longitudinal character of the dataset is problematic since it restricts our ability to address different endogeneity problems in more depth. Hence, one main priority for future research should be to gather better data, which are more based on objectively measurable variables and longitudinal in character.

However, also the current dataset would allow for some further extensions, which should be considered in future research. One such is to use the information on various health related behaviours - e.g. exercise, diet, cigarette consumption - to estimate a structural model (cf. Balia and Jones, 2007) where the allocative relevance of knowledge is measured. Such an analysis would possibly give me the ultimate answer to the question whether knowledge is key to understand the education gradient in health.

Another extension of the current analysis would be to allow the parameters of the model to vary with the reported health status. It could be argued that the effect of knowledge on health should be very different in different health states, as suggested by previous research on obesity and knowledge (Kan & Tsai, 2004). Our finding that coefficients are different in the groups with and without chronic conditions suggests that such an analysis could offer valuable insights.

6 References

Ajzen, Icek (1988), *Attitudes, Personality, and Behavior.* Chicago: Dorsey Press.

Angus J., S. Evans, J. Lapum, E. Rukholm, R. St Onge, R. Nolan and I. Michel (2005), "'Sneaky disease': the body and health knowledge for people at risk for coronary heart disease in Ontario, Canada.", *Social Science and Medicine*, 60(9), pp. 2117-28.

Auld, M. C. and N. Sidhu (2005), "Schooling, cognitive ability, and health", *Health Economics* 14(10), pp. 1019-1034.

Balia, S. and A. M. Jones (2007), "Mortality, lifestyle and socio-economic status", forthcoming, *Journal of Health Economics.*

Beaver, K., O. Craven, G. Witham, M. Tomlinson, S. Susnerwala, D. Jones and K. A. Luker (2007), "Patient participation in decision making: views of health professionals caring for people with colorectal cancer", *Journal of Clinical Nursing*, 16:4, pp. 725-733.

Bessell, T.L., S. McDonald, C.A. Silagy, J.N. Anderson, J.E. Hiller and L.N. Sansom (2002), "Do Internet interventions for consumers cause more harm than good? A systematic review", *Health Expectations*, 5:1.

Carbone, J. C., S. Kverndokk, and O. J. Røgeberg (2005), "Smoking, health, risk, and perception", *Journal of Health Economics*, 24, pp. 631-53.

Chandola, T., P. Clarke, J. N. Morris, D. Blane (2006), "Pathways between education and health: a causal modelling approach", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169:2, pp. 337-359.

Chapple A., S. Campbell, A. Rogers and M. Roland (2002), "Users' understanding of medical knowledge in general practice", *Social Science and Medicine*, 54(8), pp. 1215-24.

Coulter, A. (2002), *The Autonomous Patient: Ending Paternalism in Medical Care*, The Nuffield Trust.

Crocco, A. G., M. Villasis-Keever and A. R. Jadad, (2002), "Analysis of Cases of Harm Associated With Use of Health Information on the Internet", *JAMA*, 287, pp. 2869-2871.

Cutler, D. M. and A. Lleras-Muney (2006), *Education and Health: Evaluating Theories and Evidence*, NBER Working Paper No. W12352.

Ellins, J. and A. Coulter (2005), *How engaged are people in their health care? Findings of a national telephone survey*, Research Report, The Health Foundation.

Eysenbach, G., J. Powell, M. Englesakis, C. Rizo and A. Stern (2004), "Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions", *BMJ*, 328:1166.

Fuchs, V. R. (2004), "Reflections on the socio-economic correlates of health", *Journal of Health Economics*, 23:4, pp. 653-661.

Grossman, Michael (1972), "On the concept of health capital and the demand for health", *Journal of Political Economy* 80:223-255.

Grossman, Michael (2004), "The Demand for Health, 30 years later: a very personal retrospective and prospective reflection", *Journal of Health Economics*, 23:4, pp. 629-636.

Hernandez-Quevedo C, A. Jones and N. Rice (2005), *Reporting bias and heterogeneity in self-assessed health: evidence from the British Household Panel Survey*. Health Econometrics and Data Group (HEDG) Working Paper 05/04: York.

Hibbard J. H., J. Stockard, E. R. Mahoney, and M. Tusler (2004), "Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patient and Consumers", *Health Services Research*, 39(4), pp. 1005-1026.

Howard D. H., J. Gazmararian, R. Parker (2005), "The impact of low health literacy on the medical costs of medicare-managed care enrolees", *American Journal of Medicine*, 118(4), pp. 371-7.

Hsieh, C.-R., L.-L Yen, J.-T. Liu and C. J. Lin (1996), "Smoking, health knowledge, and anti-smoking campaigns: An empirical study in Taiwan", *Journal of Health Economics*, 15, pp. 87-104.

Huisman, M., F. van Lenthe and J. Mackenbach (2007), "The predictive ability of self-assessed health for mortality in different educational groups", *International Journal of Epidemiology*, pp. 1-7.

Jones, A. M. and J. M. Kirigia (1999), "Health Knowledge and Smoking among South African Women", *Health Economics*, 8, pp. 165–169.

Kan, K and W.-D. Tsai (2004), " Obesity and risk knowledge", *Journal of Health Economics*, 23, pp. 907-34.

Kind P, B. Leese and G. Hardman (2005), "Measuring health status: information for primary care decision making", *Health Policy*, 71(3), pp. 303-13.

Marmot, M., J. Banks, R. Blundell, C. Lessof and J. Nazroo (eds.) (2003), *Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing*, London: IFS

McKinlay, J. B. and Lisa D. Marceau (2002), "The end of the Golden Age of Doctoring", *International Journal of Health Services*, 32: 2, pp. 379-416.

Nutbeam D. (2000), "Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century", *Health Promotion International*, 15(3), pp. 259-67.

Olshansky, S. J. and B. Ault (1986), "The fourth stage of the epidemiologic transition: The Age of Delayed Degenerative Diseases", *The Milbank Quarterly*, 64(3), p. 355-391.

Omran, A. R. (1971), "The epidemiologic transition. A theory of the epidemiology of population change", *The Milbank Memorial Fund Quarterly*, 49(4), p. 509-538.

Prentice-Dunn, S. and R. W. Rogers (1986), "Protection Motivation Theory and preventive health: beyond the Health Belief Model", *Health Education Research*, 1: 3, pp. 153-161.

Rogers, R. W. (1975) "A protection motivation theory of fear appeals and attitude change", *Journal of Psychology*, 91, pp. 93— 114.

Rosenstock, I. M. (1966), "Why people use health services", *Milbank Memorial Fund Quarterly*, 44, pp. 94–124.

Rosenstock, I. M. (1974), “The health belief model and preventive health behavior”, *Health Education Monographs*, 2, pp. 354 – 386.

Salomon, J. A., A. Tandon and C. J. L. Murray (2004), “Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes”, *BMJ* 328;258.

Smith, R.B. (2005), “An Alternative Perspective on Information Asymmetry; Implications for Consumer Authority in Physician Services Markets”, *Journal of Economics & Management Strategy*, 14: 3, pp. 665-699.

Steptoe, A and J. Wardle (2001), “Health behaviour, risk awareness and emotional well-being in students from Eastern Europe and Western Europe”, *Social Science & Medicine*, 53:12, pp. 1621-1630.

Taylor, D. and M. Bury (2007), “Chronic illness, expert patients and care transition”, *Sociology of Health & Illness*, 29:1, pp. 27–45.

Tsou, M.-W. and J.-T. Liu “Obesity, Schooling and Health Knowledge: An Empirical Study of Taiwanese Women”, *Education Economics*, 14:1, pp. 89 – 106.

De Walt, D., N. D. Berkman, S. Sheridan, K. N. Lohr and M. P. Pignone (2004), "Literacy and health outcomes: A systematic review of the literature", *Journal of General Internal Medicine*, 19(12), pp.1228-39.

Wooldridge, J. M. (2001), “Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples,” *Econometric Theory* 17, pp. 451-470.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Table 7: Estimation Results, Separate estimates for chronically ill.

	Health		**Knowledge**	
Variable	*Estimate*	*pVal*	*Estimate*	*pVal*
Knowledge	0.831	0.00		
Age	11.590	0.00	1.128	0.00
Age ^{2}	-8.569	0.00	-2.135	0.00
Male	0.115	0.31	-0.448	0.00
EDUpostsec	-0.046	0.84	0.092	0.55
EDUdegree	0.085	0.69	0.484	0.00
White	0.116	0.51	0.228	0.20
Cohabit	0.024	0.90	0.138	0.36
NPeople	0.094	0.24	0.019	0.72
Hypertension	-0.215	0.01	-0.162	0.31
ChronicPain	-0.837	0.00	-0.067	0.50
Arthritis	-0.511	0.00	-0.085	0.40
Cholesterol	-0.258	0.00	0.142	0.13
Other	-0.345	0.00	-0.138	0.33
Action			0.707	0.00
Qualifications			0.305	0.08
CKnowledge	-0.166	0.26		
CAge	-1.705	0.00	-0.432	0.00
CAge ^{2}	2.013	0.00	1.095	0.00
CMale	-0.057	0.68	0.249	0.12
CEDUpostsec	0.084	0.73	0.145	0.41
CEDUdegree	0.046	0.81	0.096	0.60
CWhite	-0.126	0.48	-0.366	0.00
CCohabit	0.077	0.72	0.098	0.58
CNPeople	-0.036	0.63	-0.018	0.77
CAction			-0.287	0.06
CQualifications			0.151	0.45
a_1	1.978	0.00	-0.591	0.00
a_2	2.825	0.00		
a_3	3.721	0.00		
a_4	4.701	0.00		
ρ	-0.315	0.01		
N	2,811			
Logl	-5,372			
AIC	10,852			
BIC	11,173			

Table 8: Estimation Results, Full Model.

	Health		**Knowledge**	
Variable	*Estimate*	*pVal*	*Estimate*	*pVal*
Knowledge	0.840	0.00		
Age	9.582	0.00	0.169	0.00
Age ^{2}	-6.503	0.00	-0.820	0.00
Male	0.081	0.15	-0.274	0.00
EDUpostsec	-0.011	0.87	0.187	0.02
EDUdegree	0.076	0.32	0.570	0.00
White	0.050	0.44	-0.023	0.87
Cohabit	0.051	0.34	0.211	0.00
NPeople	0.078	0.00	0.005	0.00
Hypertension	-0.381	0.00	-0.198	0.01
ChronicPain	-0.942	0.00	-0.108	0.14
Arthritis	-0.666	0.00	-0.124	0.12
Cholesterol	-0.380	0.00	0.112	0.18
Other	-0.678	0.00	-0.219	0.01
Action	0.047	0.47	0.473	0.00
Qualifications	-0.148	0.07	0.461	0.00
a_1	1.623	0.00	-0.357	0.05
a_2	2.461	0.00		
a_3	3.335	0.00		
a_4	4.279	0.00		
ρ	-0.391	0.00		
N	2,811			
Logl	-5,410			
AIC	10,894			
BIC	11,114			

Table 9: Estimation Results, Kowledge included or excluded.

	HealthA		HealthB		HealthC	
Variable	*Estimate*	*pVal*	*Estimate*	*pVal*	*Estimate*	*pVal*
Knowledge	0.679	0.00	0.228	0.00		
Age	9.366	0.00	9.806	0.00	9.911	0.95
Age ^{2}	-6.386	0.00	-6.839	0.00	-6.986	0.77
Male	0.067	0.23	0.015	0.79	-0.012	0.00
EDUpostsec	-0.001	0.99	0.037	0.57	0.055	0.02
EDUdegree	0.125	0.14	0.237	0.00	0.288	0.00
White	0.050	0.75	0.047	0.73	0.042	0.92
Cohabit	0.066	0.28	0.104	0.10	0.122	0.01
NPeople	0.078	0.00	0.082	0.00	0.083	0.77
Hypertension	-0.401	0.00	-0.440	0.00	-0.453	0.02
ChronicPain	-0.969	0.00	-1.008	0.00	-1.010	0.32
Arthritis	-0.690	0.00	-0.722	0.00	-0.726	0.12
Cholesterol	-0.375	0.00	-0.364	0.00	-0.353	0.15
Other	-0.703	0.00	-0.759	0.00	-0.772	0.04
a_1	1.448	0.00	1.335	0.15	-0.705	0.22
a_2	2.303	0.00	2.210	0.02	0.142	0.81
a_3	3.193	0.00	3.122	0.00	1.153	0.04
a_4	4.155	0.00	4.106	0.00	2.176	0.00
ρ	-0.292	0.00				
N	2,811		2,811		2,811	
Logl	-5,413		-5,416		-5,430	
AIC	10,895		10,900		10,925	
BIC	11,103		11,102		11,121	

Table 10: Estimation Results, General health.

	Health		Knowledge	
Variable	*Estimate*	*pVal*	*Estimate*	*pVal*
Knowledge	0.218	0.00		
Age	4.325	0.14	0.172	0.96
Age ^{2}	-2.686	0.23	-0.836	0.76
Male	-0.058	0.30	-0.281	0.00
EDUpostsec	0.200	0.00	0.190	0.02
EDUdegree	0.340	0.00	0.571	0.00
White	0.108	0.43	-0.015	0.92
Cohabit	-0.027	0.67	0.208	0.01
NPeople	0.055	0.06	0.009	0.77
Hypertension	-0.722	0.00	-0.191	0.02
ChronicPain	-0.880	0.00	-0.092	0.32
Arthritis	-0.645	0.00	-0.129	0.12
Cholesterol	-0.487	0.00	0.129	0.15
Other	-0.925	0.00	-0.208	0.04
Action			0.474	0.00
Qualifications			0.457	0.00
a_1	-0.406	0.66	-0.373	0.74
a_2	0.465	0.62		
a_3	1.561	0.09		
a_4	2.554	0.01		
N	2,811			
Logl	-5,362			
AIC	10,793			
BIC	10,995			