CONTRIBUTI DI RICERCA CRENOS

CRENOS
CENTRO RICERCHE
ECONOMICHE NORD SUD
Università di Cagliari
Università di Sassari

MORTALITY, LIFESTYLE AND SOCIO-ECONOMIC STATUS

Silvia Balia
Andrew M. Jones

WORKING PAPERS

CUEC

2004/16

Mortality, Lifestyle and Socio-Economic Status

Silvia Balia Andrew M. Jones*
Department of Economics and Related Studies
University of York
York YO10 5DD
UK

October 2004

Abstract

This paper uses the *British Health and Lifestyle Survey* (1984-1985) data and the longitudinal follow-up of May 2003 to investigate the determinants of premature mortality risk in Great Britain. A behavioral model, which relates premature mortality to a set of observable and unobservable factors, is considered. We focus on unobservable individual heterogeneity and endogeneity affecting the mortality equation. A MSL approach for a multivariate probit (MVP) is used to estimate a recursive system of equations for deaths and lifestyles. This model is then compared with the univariate probit models that include or exclude lifestyles. In order to detect inequality in the distribution of health within the population and to calculate the contribution of socio-economic factors, we compare the range measure of health inequality to the Gini coefficient for overall health inequality. A Gini decomposition analysis for predicted premature mortality shows that endogenous lifestyles and unobservable heterogeneity strongly contribute to inequality in mortality, reducing the role of socio-economic status.

JEL codes I1 C0
Keywords: Mortality; Lifestyle; Socio-economic status; Health and Lifestyle Survey; Multivariate Probit; Simulation-based inference; Health Inequality.

*The authors wish to thank Paul Contoyannis for his suggestions and comments. Data from the Health and Lifestyle Survey (HALS) were supplied by the ESRC Data Archive. Any mistakes in the analysis and interpretation of the data presented in this paper are responsibility of the authors.

1 Introduction

Health inequalities have been of growing interest for the economics literature in recent years (see e.g., Deaton, 2003). Inequalities in health, for example between social groups, are partly explained by differences in lifestyle and living conditions, and lifestyle can vary between groups depending on economic circumstances (see Wagstaff, 1986). To better explain inequalities in health, it is appropriate to use a behavioral model, which contains socio-economic characteristics but also individual health decisions, and all potential determinants of health. Such a model will be helpful for the policy-maker, whose purpose is to improve the overall health status of the population and to reduce inequalities in health[1].

A recent contribution by Contoyannis and Jones (2004) contradicts the view, that is widely held in the epidemiological literature, that differences in lifestyle make a relatively minor contribution to the observed socio-economic gradient in health. This paper is an extension of their analysis of the relationship between individual socio-economic characteristics and health. The key point is that unobservable heterogeneity deserves particular attention.

The importance of inequality in health stems from evidence that substantial differences in health do exist across individuals. The best way to investigate the determinants of health status in different groups of the population is to work with individual data: the British Health and Lifestyle Survey (HALS) gives us the scope for an analysis at the individual level. While Contoyannis and Jones (2004) use a measure of self-assessed morbidity, we extend the analysis to use a measure of mortality. This exploits a longitudinal follow-up that records deaths of respondents to the original HALS study. We use data from the fourth revision of the deaths data, released in May 2003.

In this paper we refer to lifestyle as a set of behaviours which can influence health. We assume that individuals choose these health-related behaviors. However, it is not our intention to delve into the concept of *free choice* and debate the potential reduction of freedom, due to economic, environmental and social circumstances. Our approach recognizes that individual decisions reflect both the constraints of personal circumstances and consumption preferences.

As well as contextual factors, the presence of potential risky behaviours should be considered when dealing with individual choice. People can make an unhealthy risky decision because the cost of being fully informed about the health consequences is too high. Incomplete or incorrect information may characterize the decision process. Moreover, the rate at which people discount the future, especially when the decision is made at a time far from the time when it will have its effect on health, can influence choices. Fuchs (1982) suggests that the individual rate of time preference has an important role in individual health decisions. When the discount rate is low, people tend to invest more in both education and behaviours enhancing health. Hence, lifestyle and education are not the only two variables influencing health, but

[1]Much of the literature is concerned with the debate about the best normative framework to achieve equity and efficiency in health. (see Hurley, 2000, for a review of the literature).

the rate of time preference should be considered as a potential "third variable". Information problems and time preferences are likely to be unobservable to the researcher.

The plan of the paper is as follows. The next section surveys the previous economic literature in this area. Section 3 gives a brief overview of the microeconomic concepts underlying the empirical model. Section 4 describes the Health and Lifestyle Survey dataset. Section 5 describes the variables of main interest in the analysis. Section 6 gives a simple description of the sample, considering socio-economic status, lifestyle and mortality. Section 7 describes the estimation strategy and the main results. Section 8 introduces the Gini measure of overall inequality in health and shows the results of the decomposition of this index. The last section contains a short conclusion.

2 Previous economic literature

In the last thirty years, the health economics literature has used the concept of a technology through which individuals produce and modify their own health status. Health is reckoned to be both a consumption commodity and the outcome of a production process, which involves medical care and depends on individual behaviours and choices. If we want to identify factors that influence good health, we should recognize the important role played by individual behaviours.

Economic theory recognizes human capital as a determinant of investment and growth, and several studies consider health not only as a good that enhances individual satisfaction, but also as a capital stock. In Becker's theory (1965) of the allocation of time, households consume and produce commodities, using goods and time as inputs. Investments in human capital might be measured in terms of market goods and of the opportunity cost of time lost for competing and alternative uses. Following this idea, Grossman (1972) emphasized the fundamental distinction between human capital and health: health can increase the possibility of consuming market goods[2]. People invest in health to reduce spells of illness and be able to consume more. Becker and Grossman's works provided the framework for further studies, which have developed the concept of the health production function, using the time of the consumer, health-related behaviours and environmental factors as inputs.

Empirical analysis in health economics faces problems with the identification of the effects that socio-economic characteristics, health-related behaviours and utilization of medical services have on health. Three serious econometric problems arise: heterogeneity, endogeneity and selection-bias. Unobservable factors might exist, leading to biased estimates and to spurious relationships with the dependent variable. Regressors are not always exogenously

[2]In particular, although health is not a tradable good, it especially affects the quantity of time available to individuals for working.

determined, but rather correlated with the error term. If the sample is not representative of the underlying population, on account of self-selection [3], then bias will arise.

Many empirical analyses of the determinants of health have relied on single equation instrumental variable approaches, such as 2SLS and GMM, to address the potential biases (see Auster et al., 1969; Mullahy and Portney, 1990; Mullahy and Sindelar, 1996). Our approach is different. Following Contoyannis and Jones (2004), a wider concept of lifestyle is used: six, not just one, health-related behaviours are considered. They are described by reduced form equations and appear as potential endogenous regressors in the health equation. A multivariate probit model, whose specification is justified by the underlying economic theory, is implemented to allow for unobservable heterogeneity. The structure of the empirical model will be illustrated below.

3 Microeconomic foundations of the model

This section gives a brief overview of the economic theory underlying Contoyannis and Jones's (2004) model. The model is based on simple key assumptions: health is a fundamental commodity produced by health-related behaviors and other inputs; and the individual knows the marginal productivity of these behaviors and all the parameters of the decision process.

Health-related behaviors are particular individual choices, which help to explain the variation in the distribution of health in the population. Nevertheless, other factors that influence an individual's decisions, although known to the individual, are hidden to the researcher who does not have, therefore, a complete knowledge of the decision process. These unobservable factors make the desired consumption level differ for each individual. Contoyannis and Jones (2004) propose a model that permits the researcher to control for individual heterogeneity.

The problem of the decision maker is the maximization of the following utility function, under a budget and a time constraint [4]:

$$\max_{C,H} U = U(C, H, X_U, \mu_U) \tag{1}$$

[3] Grossman and Joyce (1990) focused on the first infant health production function, controlling for self-selection in the resolution of pregnancy as live births or induced abortions and in the use of prenatal medical care services.

[4] The money budget constraint and the time constraint are given by $\sum_{j=1}^{M} p_j C_j \leq y_0 + \omega L$ and $\sum_{j=1}^{M} \tau_j C_j = T - L$, respectively, where y_0 is exogenous income, ω is the wage rate, τ_j is the amount of time necessary to consume a unit of the commmodity C_j, T is the total time available and L represents the hours of labour. Taking $L = \tau_j C_j - T$ and combining the two constraints above, gives the full income constraint: $\sum_{j=1}^{M}(p_j + \omega\tau_j)C_j \leq y_0 + \omega T$. The maximization problem, which consists in solving the first order conditions from a Lagrangian function, would give the Marshallian demands for the M goods of our system and the marginal utilities of health and full income at the optimum.

The utility function depends on the measure of future health (H), on a vector of lifestyles C , and on X_U and μ_U, which are respectively vectors of observables and unobservables factors that influence utility.

The health production function is

$$H = h(C, X_H, \mu_H) \tag{2}$$

where X_H is a vector of exogenous factors that influence utility and μ_H is a vector of unobservable influences on health.

A recursive empirical model will be developed. The basic idea is that reduced form equations for lifestyle determinants (3) and for the health production function (4) do not distinguish between preferences and technological parameters, and do not permit estimation of the direct effect of lifestyle on health.

$$C_i = f_i(X, p, w, y_0, \tau, \mu) \qquad \text{where} \quad i = 1, ..., M \tag{3}$$

$$H = f(X, p, w, y_0, \tau, \mu) \tag{4}$$

where C_i is the vector of commodities desired by the individual; X is a vector that combines factors that influence both utility (X_U) and health (X_H); p, w and y_0 are elements of the money budget constraint; τ is an element of the time constraint and μ includes unobservable factors which influence both the individual utility function (μ_U) and the health outcome (μ_H).

The recursive empirical model can be estimates jointly, by Full Information Maximum Likelihood (FIML). It consists of the reduced form equations for lifestyle (3) and the structural form equation for health (2). With this specification the researcher can control for unobservable heterogeneity in the population allowing for correlation between the marginal utilities of health and lifestyle choices. Technical complications in the maximization of the likelihood arise on account of binary endogenous regressors, especially if there are more than three dependent variables. The error term in each reduced form equation depends on μ; assuming that random components in (3) and (2) have a multivariate normal distribution, the integrals in the likelihood function have no closed-form. The Maximum Simulated Likelihood (MSL) approach permits computation of the model and is used to estimate a multivariate probit model.

4 Data

In this paper, data from the first wave of the Health and Lifestyle Survey (HALS1) are used to measure lifestyle. They were collected between Autumn 1984 and Summer 1985, in two home visits (the second one by a nurse). The questionnaire was designed and piloted by a study team at the University of Cambridge School of Clinical Medicine and funded by the Health Promotion Research Trust. The sample design permits inferences about the British population, aged 18 and over in 1984-85.

Table 1: Flagging Status in May 2003

Flagging Status	Frequency	Percent
On file[5]	6506	72.26
Not NHS *registered*[6]	86	0.96
Deceased[7]	2171	24.11
Reported dead to HALS not on NHS Register[8]	1	0.01
Embarked - abroad[9]	43	0.48
Not yet flagged[10]	196	2.18

While Contoyannis and Jones (2004) use a morbidity measure, namely self-assessed health (SAH), we use a binary indicator of death as our measure of health. This allow us to measure health outcomes as recently as 2003 and the accuracy of the mortality data offsets concerns about measurement error in self-reported health.

Most of the 9003 individuals interviewed in HALS1, have been *flagged* on the NHS Central Register. In May 2003 the fourth deaths revision and the first cancer revision were completed[11]. The flagging process was quite lengthy because it required several checks in order to be sure that the flagging registrations were related to the person previously interviewed. As reported in Table 1, 97.8 per cent of the sample has been flagged. Deaths account for some 24 per cent of the original sample. The length of the follow-up period, nineteen years, and the checks on the official registers of deaths ensure the reliability of mortality data in HALS. A recent paper by Gardner and Oswald (2004), using longitudinal data from the British Household Panel Survey (BHPS) to study the influence of socio-economic dynamics on longevity, suffers the limitation of a shorter follow-up period (only ten years) and the absence of a double check on the official register.

The idea is to test the same kind of framework used on SAH, but exploiting the newly available information on deaths. We propose an analysis of the relationship between death and individual characteristics, measured nineteen years before. Although the analysis covers a relatively long follow-up period, increased risk of mortality may reflect the cumulative effect of poor health (it is not surprising that people in the poorest health status are more likely to die). Hence, at this stage of the work, the probability of death is studied to explain to what extent initial conditions (measured in 1984) determine subsequent health.

[5]Currently alive and flagged on the NHS Register

[6]But not known to be dead

[7]Known dead and death certificate information recorded on file

[8]May be alive

[9]Identified on NHS Register but currently out of country

[10]Not currently flagged for various reasons (no name etc.)

[11]For further information see the Working Manual for HALS available at the UK Data Archive Health and Lifestyle Survey - University of Cambridge Clinical School (2003).

5 Variables and Sample

The lifestyle variables indicate whether the individual is a non-smoker, a prudent consumer of alcohol, eats breakfast, sleeps the "optimal" numbers of hours, is not obese, and did sufficient physical activity in the last fortnight[12]:

1. Smoking is defined in terms of number of cigarettes smoked per day, using an indicator for current smokers who smoke one or more cigarettes per day.

2. Drinking is measured by a binary variable which indicates prudent alcohol consumption. The indicator is gender specific and is based on the number of drinks consumed in the past seven days before the interview[13].

3. Breakfast is an indicator of diet: we assume that eating breakfast within one hour of waking is a healthy behavior.

4. Sleep has been recognized as potential determinant of health status. Hence, we created an indicator which splits the sample into two groups according to their sleep pattern[14].

5. The indicator for obesity is measured using the Body Mass Index (BMI) reported at the nurse visit and it is gender specific [15].

6. Since sporting activities are known to be healthy and to help people in stress or depression, we also use an indicator of physical activity, measured in the last fortnight. It is created for each individual by summing the time involved in each of fourteen types of exercise.

The other variables describing socio-economic characteristics, geographical position, marital status, housing are reported in A.1.

We use a sample limited to 3670 individuals aged 40 and over from the original sample of 9003 observations. All observations containing missing values have been dropped out. Both descriptive statistics and econometric analysis are conditioned upon this restriction[16].

[12]We use lifestyle indicators close to the categories of healthy behaviours found in epidemiological and health economic studies, such as Belloc (1973), Belloc and Breslow (1972), Kenkel (1995) studies on the data from the Alameda County survey carried out in California in 1965.

[13]Our indicator of alcohol consumption cannot discriminate between different styles of drinking: it does not capture differences among people who are in an abstinence period, and could be heavy drinkers, and those who are completely non-drinkers. Hence, the interpretation of the impact of this variable on mortality can be difficult if these two drinking styles are likely to have different effects. In general, the interpretation of the impact of alcohol consumption on health is not easy because there is evidence that moderate consumption gives some positive effects on health.

[14]We define the optimal sleeping level as sleeping between seven and nine hours per night.

[15]Height, which is a component of the BMI, is also included as a continuous exogenous variable in the econometric model, because it is known to be a good predictor of mortality and morbidity risks.

[16]Persons who are younger than 40 are more likely to change their educational qualification over time, but the cross-sectional nature of the analysis does not allow to control for this

Table 2: Variable means by sub-samples defined by number of *a priori* "healthy behaviors"

Variable	Full N=3670	0/1/2 N=373 (*Exp*=234.88)	3/4/5 N=2945 (*Exp*=3181.89)	6 N=352 (*Exp*=253.23)
death	0.359	0.399	0.374	0.19
sc12	0.316	0.214	0.306	0.5
sc3	0.466	0.496	0.473	0.378
sc45	0.219	0.29	0.221	0.122
lhqdg	0.126	0.078	0.119	0.236
lhqhndA	0.125	0.094	0.124	0.168
lhq0	0.094	0.059	0.093	0.142
lhqnone	0.608	0.716	0.619	0.401
lhqoth	0.047	0.054	0.046	0.054
part	0.132	0.115	0.124	0.216
unemp	0.03	0.054	0.03	0.009
sick	0.033	0.054	0.033	0.011
retd	0.339	0.19	0.372	0.219
keephse	0.102	0.126	0.098	0.108
male	0.455	0.504	0.451	0.44

6 Descriptive analysis

6.1 Lifestyle and socio-economic status

A simple descriptive analysis is reported in Table A.2, which presents sample means for the most relevant variables that describe the full sample and the separate sub-samples for men and women. Since morbidity measures usually suggest poorer health status for women, and women have different characteristics and behaviours than men, we differentiate the analysis by gender.

It is worth noting that the full sample (that comprises 46 per cent men and 54 per cent women) is made up of individuals whose behaviours are mostly healthy, particularly among women. 88 and 85 per cent of the sample is prudent in the consumption of alcohol and is not obese. Only 30 per cent of individuals are smokers, while 32 per cent of them devote time to physical activities, 71 per cent usually eat breakfast and 58 per cent sleep a healthy number of hours.

The social class classification is derived from the Registar General's Social Class (RGSC), based on occupation. For our purposes, the most convenient way to aggregate individuals in the sample is to collapse social classes in three macro groups: a top class (SC12) including students, professional, managerial and intermediate workers, a middle class (SC3) including skilled workers

change. The interpretation of the effect of education on mortality risk could be biased. For the same reason, respondents in HALS 1984 who are older than 40 but are still classified as students are dropped out.

and armed services, and a bottom class (SC45) including partly skilled and unskilled. Individuals are largely concentrated in SC3; the extreme classes are smaller. Around 61 per cent of the sample does not have a formal education or qualifications, and only 13 per cent has a university degree. Men are mostly full time workers (56 per cent) or retired (31 per cent), whereas women are mostly retired (36 per cent) and part-time and full time workers (22 and 20 per cent).

Table 2 shows the mean values of some variables of interest for the full sample considered in our analysis. The same statistics are reported for sub-samples, according to the intensity of healthy behaviours. Since splitting the full sample in different groups on the basis of every combination of lifestyle choices would have been prohibitive, the table only presents the characteristics of three sub-groups[17].

Although sixty-four (2^6) possible combinations of health-related behaviours can be analyzed, we are only interested in the probability of having a totally healthy lifestyle $Pr(6)$, the probability of following a totally unhealthy lifestyle, one or two healthy behaviours $Pr(0/1/2)$, and the probability of three, four, or five healthy behaviours $Pr(3/4/5)$[18].

We find $Pr(0/1/2) = 0.064$, $Pr(3/4/5) = 0.867$ and $Pr(6) = 0.069$. These probabilities, multiplied by the full sample size, give the expected frequency of each group, reported between brackets in the table. The expected frequencies for the most unhealthy and the most healthy groups are underestimated. The expected frequency for the group of persons who follow three, four or five healthy practices is estimated to be 1.08 times bigger than observed frequency.

The subdivision of the sample shows that the number of deaths decreases as we move from the more unhealthy group to the healthiest lifestyle. Social position, work environment and education can have a strong role in the determination of one's health status and, consequently, of the odds of dying. The healthier individual behaviours are, the bigger the proportion of persons belonging to the higher social classes. The number of individuals in the bottom classes decreases moving from the most unhealthy lifestyles to the healthiest. A strong association of schooling with health-related choices is shown by the table: people have more healthy behaviours if they are more educated. Individuals with no educational qualifications have more unhealthy lifestyles. The role of schooling on health has been emphasized in Grossman (1972) and deeply investigated in Kenkel's empirical work (1991).

This evidence leads to the same conclusion as Contoyannis and Jones's (2004) study of SAH : health-related behaviours are not randomly distributed

[17]We do not report in the table the case in which individuals have a totally unhealthy lifestyle because the size of this group is too small to make inferences.

[18]These probabilities are computed using the following formulas: $Pr(6) = \prod_{i=1}^{6} P(Y_i = 1) = \{1,1,1,1,1,1\}$, $Pr(0/1/2) = \prod_{i=1}^{6} P(Y_i = 0) + \sum_{i=1}^{6} P(Y_i = 1) \prod_{k \neq i} P(Y_k = 0) + \sum_{i=1}^{6} \sum_{k>i} P(Y_i = 1) P(Y_k = 1) \prod_{j \neq k \neq i} P(Y_j = 0)$, $Pr(3/4/5) = 1 - Pr(0) - Pr(1/2) - Pr(6)$, where $Pr(0) = \prod_{i=1}^{6} P(Y_i = 0) = \{0,0,0,0,0,0\}$. Since $Pr(0)$ is very small (0.0005), people having either no healthy behaviour, or at least one or two, compose the most unhealthy group.

but cluster together in certain categories of individuals, and the relationship with social class and education must be taken into account. However we still need evidence about the extent to which these factors influence the health outcome. We do not know if their impact on health is subject to change depending on individual propensities to behave in a healthy way.

6.2 Deaths and socio-economic status

Table 3 shows that, in our sample, the most frequent causes of death are diseases of the circulatory system ($C7$), neoplasms ($C1$) and diseases of the respiratory system ($C8$). Deaths in the three classes are mainly due to diseases of the respiratory system, with a maximum of 48 percent of deaths due to this cause in the middle class (SC3). The incidence of respiratory diseases is higher for the SC45 class (around 17 persons out of 100 die for this cause, while in the other classes this percentage is a bit lower).

A crude way to see if mortality varies with the characteristics of the population is to use the simple death rate[19]. Table 4 records death rates for some variables of interest. It is not surprising that the highest social classes have lower mortality rates. The death rate increases from the highest social class (SC12) to the lowest classes. About 43 individuals die for every 100 persons in the population in the SC45 class. Only 27 individuals, for every 100, die in the highest social class. Premature mortality (death before age 65), is higher among people who are unskilled. To measure the difference of death rate between social classes a relative measure of mortality, which is the simple ratio of the death rates in different social classes, has been constructed. Table 5 shows the relative measure of mortality for some socio-economic characteristics of interest. The death rate is almost double in the lowest two classes with respect to the highest. The relative measure decreases comparing death rates between lower social classes.

Tables 4 and 5 also show the association between mortality and education. The role that a person can play in the labour market depends on their educational qualifications. The attained level of education is, in such a way, related to socio-economic status, income, housing, health status and health inequalities. Kenkel (1991) finds that better educated persons are also more likely to have a good knowledge of what a person should do to be more healthy. The death rate is bigger for people who have no qualifications: it is about twice the death rate of those who obtained a University degree or an O level.

[19]The death rate is calculated in each class as the proportion of deaths in that class. The number of deaths in a certain category, times 100, divided by the total number of individuals in that category.

Table 3: Causes of death and Percentage by social class(N = 1318)

Causes		Full sample		SC12	SC3	SC45
		Freq.	Percent	Percent	Percent	Percent
C1	*Infectious and parasitic diseases*	9	0.68	1.29	0.45	0.58
C2	*Neoplasms*	368	27.92	28.62	28.77	25.66
C3	*Endocrine, nutritional and metabolic diseases and immunity disorders*	17	1.29	1.61	0.75	2.04
C4	*Diseases of the blood and blood-forming organs*	2	0.15	0.64	-	-
C5	*Mental disorders*	20	1.52	3.22	0.90	1.17
C6	*Diseases of the nervous system and sense organs*	15	1.14	0.96	1.51	0.58
C7	*Diseases of the circulatory system*	598	45.37	40.84	48.04	44.31
C8	*Diseases of the respiratory system*	164	12.44	11.25	10.84	16.62
C9	*Diseases of the digestive system*	43	3.26	2.89	2.56	4.96
C10	*Diseases of the genitourinary system*	15	1.14	1.93	1.20	0.29
C12	*Diseases of the skin and subcutaneous tissue*	4	0.30			
C13	*Diseases of the musculoskeletal system and connective tissue*	12	0.91	-	0.60	-
C16	*Symptoms, signs and ill-defined conditions*	13	0.99	1.61	0.75	0.87
C17	*Injury and poisoning*	6	0.46	0.96	0.45	-
C18	*Supplementary classification of external causes of injury and poisoning*[20]	16	1.21	1.93	0.75	1.46
m. v.		28	2.12	2.25	2.41	1.46
total		28	100	100	100	100

[20] This includes also accidents and suicides.

Table 4: Death Rate in different socio-economic groups

Variable	Death Rate
sc12	26.86
sc3	38.83
sc45	42.77
lhqdg	24.68
lhqhndA	22.49
lhqO	22.32
lhqnone	43.12
male	42.97
female	30.01

Table 5: Relative mortality measures between some socio-economic group

Variable	Relative Mortality
sc45 - sc12	1.59
sc45 - sc3	1.10
sc3 - sc12	1.45
lhqnone - lhqdg	1.75
lhqhndA - lhqdg	0.91
lhqO - lhqhndA	0.99
lhqnone - lhqO	1.93
male - female	1.43

7 Models and Results

7.1 Univariate Probit Models

We start to analyze the impact of lifestyle variables on health outcome by means of a univariate probit model. For comparison, two versions of the model are estimated. The first includes lifestyle variables in the vector of regressors, whereas the second excludes them. The goal is to compare the difference in the impact of variables describing the individual's socio-economic position between these two models. Does lifestyle have a real effect on health? Or, do inequalities in health depend only on socioeconomic factors?

When health-related behaviors are included in the model social position and schooling may have less influence on the probability of dying, because lifestyles capture part of the effects that are not controlled for when they are excluded from the model. The aim is to assess how much of the socio-economic gradient in mortality is attributable to lifestyle.

The inclusion of health-related behaviors relies on the basic assumption of exogeneity: that is the lifestyle variables are not correlated with the error term. Methodological problems may arise on account of this assumption. A model that ignores the issue of endogeneity of some covariates and the potential unobservable heterogeneity in the sample is likely to produce biased estimates

of the parameters. However, this step of the analysis provides a benchmark for later models that do allow for endogeneity.

Our exercise leads to stronger results than those obtained by most of epidemiological researches on health inequalities. These usually use measures such as odds ratios, logistic regression or Agresti's Alpha to quantify the relationship between social position, work environment, education and health[21]. These measures assume a fixed ordering of social position and often do not consider intermediate categories. To evaluate the effect of including lifestyle variables on the social class gradient Borg and Kirstensen (2000) used a measure that compares odds ratios for the lowest social class relative to the highest for a model that includes lifestyles and a model that excludes them[22]. We construct an indicator quite close to this. It compares the variation of social class and education gradients, in terms of partial effect on mortality, between the model that incorporates lifestyles and the model that excludes lifestyles. Comparing the differences of social class and education impacts on mortality, particularly between extreme classes, in the two models, is a way to calculate the change in a range measure of inequality in health. What we expect is a reduction of the range once lifestyle variables are added to the deaths equation.

The range is the most basic and simple measure of inequalities in health[23], expressed in terms of a difference, or a ratio, and based on the comparison of the experience of the top and the bottom classes in the socio-economic distribution. Despite the fact that ease of the calculation makes the range appealing, it is limited for at least three reasons: it fails to consider the intermediate classes (even if the gap between extreme groups remains unchanged, inequality might be bigger in the middle groups), it does not take account of the size of the groups, and finally it can be interpreted only if the underlying distribution has a clear monotonic gradient.

Better measurements of inequality in health, which still take account of the socio-economic dimension, can be computed. In the next section, the Gini coefficient of total health inequality is computed and decomposed in order to assess the contribution of socio-economic variables to the overall health inequality. The results of the decomposition of the Gini are compared with the results of the analysis of inequality based on the range (see Wagstaff et al., 1991).

The regressors in the deaths equation are indicators of lifestyle habits, social class, education, work status, type of area, ethnicity, gender and age. Information criteria have been used to balance statistical fit of the model and sufficient parsimony in the parametrization. The Akaike information criterion (AIC), the consistent AIC (CAIC), and the Bayesian information criterion (BIC), all favour the same specification. These criteria suggest a cubic polynomial of age and the exclusion of the variables height and marital status. In

[21] A good review of the indicators of social position used in the epidemiologic literature and methods used to measure inequality in health is offered by Manor et al. (1997)

[22] They measure $\frac{OR_I - OR_E}{OR_E - 1}$, where OR_I is calculated for the model that includes lifestyles, and OR_E is calculated for the models that includes lifestyles and other control variables.

[23] See Wagstaff et al., 1991, for a discussion of the methods employed to measure inequalities in health.

Kenkel (1995), marital status is excluded from the health equation because it is considered a determinant in the input demand equation. Marital status is reckoned to influence health-related behaviours rather than the mortality risk directly (married people show lower death rates because they exhibit more positive health attitudes, are more likely to be wealthier and are linked to tighter-knit social support networks[24]). Concerning the choice of a polynomial representation of age, it depends on the fact that variations in the other covariates can be attributed to age. Hence, we need to separate variations across different ages and variations of other variables on the risk of mortality. We use a third order polynomial to define the probability of death as a smooth and flexible function of age. Information criteria suggest that a quartic representation would not improve the fit of the model.

The RESET tests, on the full sample and on the samples of men and women, suggest that the model is not misspecified[25].

Moving to the results of our comparative analysis, Table 6 and Table 7 show respectively social classes and education partial effects on the probability of dying and the effect of lifestyle on the social gradients on health, calculated as $\frac{PEdiff_E - PEdiff_I}{PEdiff_E}$, where $PEdiff_E$ is the difference between the partial effect for the lowest social class and the highest in the model that excludes lifestyles, whereas $PEdiff_I$ is relative to the model that includes lifestyle exogenous variables. We find a clear social class gradient on mortality[26]: the probability of dying increases, as expected, moving from the top classes to the bottom classes. For educational qualifications there is not a clear gradient.

The impact of socio-economic status on the probability of death increases: the variable sc12, for the top social group, is statistical significant and indicates a lower probability of dying once lifestyle is controlled for; while for sc45 mortality risk is positive and bigger but not statistically significant. The measure of the range decreases by around 29 per cent if lifestyle variables are included in the model. Also the impact of education becomes smaller, with a reduction in the range of around 45 per cent. This suggests that researchers should consider the close connection between health-related behaviours and health status: quite a high proportion of the socio-economic gradient is associated with differences in lifestyle. Inequalities in health cannot be explained only on the basis of income differences, education levels and social position. Indeed, the important role of individual preferences in the decision process of how healthy their life should be, has to be considered.

The separate models for men and women give stronger results. The reduc-

[24]See chapter 10 in the Handbook of Population, Poston and Micklin (2005).

[25]Wooldridge (2002, p.124-125), advises not to use the Reset to test the presence of omitted variables, because it can only test whether or not the expected value of the dependent variable conditional on the set of regressors is linear in the regressors. The result of the test confirms the validity of the cubic polynomial. The χ^2 test statistic, computed for the deaths equation that includes lifestyles, is equal to 0.25, 0.17 and 1.50, respectively for pooled sample, men and women, with p-value above the conventional significance level (p=0.61, p=0.68 and p=0.22).

[26]We collapse individuals into three macro social classes, in order to identify a clear gradient. However, using only three categories of social classes implies a reduction of information.

Table 6: Selected Partial Effects for Exogenous and Excluded Models

	Full sample		**Men**		**Women**	
	Exog	**Exclud**	**Exog**	**Exclud**	**Exog**	**Exclud**
Social Classes						
sc12	-0.055*	-0.067**	0.003	-0.010	-0.087**	-0.096**
sc45	0.001	0.012	0.034	0.050	-0.026	-0.016
Education						
lhqdg	-0.000	-0.010	0.024	0.012	-0.038	-0.043
lhqhndA	-0.035	-0.035	-0.012	-0.001	-0.055	-0.059
lhqnone	0.039	0.061†	0.100†	0.127*	-0.007	0.011
lhqoth	-0.016	0.005	-0.012	0.038	-0.005	-0.014
Significance levels:	† : 10%	∗ : 5%	∗∗ : 1%			

Table 7: The Reduction in Social Class and Education Gradients

	Gradient Reduction
	Social Class
Full sample	29.18
Men	48.22
Women	23.36
	Education
Full sample	44.65
Men	34.07
Women	42.24

tion of the social class gradient for men is around 48 per cent but it is much smaller for women, only 23 per cent. If the social class variables are not statistically significant for men, sc12 is highly significant for women. Inequality measured by the range with respect to education decreases only by 34 per cent for men, so less than for the full sample; while for women the reduction is higher, about 42 per cent.

We can summarize the findings from the univariate probit model, while recognizing that the model does not correct for endogeneity due to unobservable heterogeneity. For the full sample, the reference individual in the deaths equation is female, aged forty years old or more, is a skilled worker, with O level educational qualifications or equivalent, is white European, and lives in an inner city. In the pooled sample all lifestyle variables -with the exception of prudent drinking- are highly statistically significant and their partial effects on the dependent variable lead to the conclusion that healthy behaviours reduce the probability of death (see Tables A.3 and A.4). When the models are split by gender, only three healthy behaviours are statistically significant for men: that is smoking, obesity and breakfast. For women, only drinking and obesity are not statistically significant.

The model predicts a higher probability of death for men and white Europeans. People in the top social classes, shift workers, people living in the

countryside are less likely to die with respect to the reference individual. Educational qualification and work status are statistically significant for men, indicating a higher mortality risk for less educated people and a lower risk for shift workers. While for women only the area and social class indicators are statistically significant. Women from the top social classes and from the countryside bear a lower mortality risk.

7.2 Multivariate Probit Models

Efficient and consistent estimation of the parameters in the health production function requires a model that takes account of the nature of the variables used. The multivariate probit model is appropriate because it considers unobservable heterogeneity. The model consists of a recursive system of equations: for lifestyles and death. Its most important feature is that the random components of the lifestyle equations are allowed to be freely correlated with the random component of the deaths equation. If there are unobservable individual characteristics, influencing both individual's healthy behaviors and their probability of death, the model is able to take them into account.

Endogeneity can arise with the inclusion of lifestyles as regressors in the deaths equation, due to potential correlation between the error terms in the lifestyle equations and the error term in the deaths equation. If endogeneity is proven to exist, then estimates from the univariate probit version of the deaths equation will be biased.

We would like also to control for another issue, associated with unobservable heterogeneity: the potential measurement error in the indicators of lifestyle. Contoyannis and Jones's (2004) approach, used in this paper, overcomes the limitation of the classical approach of epidemiological studies (odds ratio, logistic regressions) to measure the reduction of the social class and education gradient in health. One of the limitations of some epidemiological studies is that they fail not only to consider the problem of unobservable heterogeneity but also the problem of measurement errors. Lynch et al. (1996), and Lynch et al. (1997), shed light on the possibility of measurement error in risk factors (behavioral, biological, psychological, social risks) used in the analysis. To obtain consistent and efficient estimates, Wooldridge (2002, p. 470-478) suggests maximum likelihood estimation when, in a system of two equations, an explanatory variable indicating participation (for example, a binary variable for smoking or drinking) is measured with error. After normalization of the error term in the equation with measurement problems, the ML procedure is used to calculate the average partial effect of the mis-measured variable on the response. Then it should be straightforward to test if there is measurement error by a simple asymptotic t-test, or LR-test, on the null hypothesis that the correlation between the errors of each lifestyle equation and the error of the deaths equation is zero. However, in this context, this test has a limitation due to the difficulty in discriminating between mis-measurement and unobservable heterogeneity in the data.

The multivariate probit model estimates a set of probabilities depending

on whether the i-th individual is dead or alive and has a more or less healthy lifestyle, according to the definition of healthy behaviors that we are using. There are 128 (2^7) combinations of successes and failures in our model, because we have seven equations and each response has got two possible outcomes.

We have a recursive system, which consists of a structural equation for the health production function and six reduced-form equations for lifestyles. We set the d^{th} equation as the deaths equation, so that the dependent variable is equal to zero if the individual is still alive. In the other equations the dependent variables take value one if that particular lifestyle is "healthy" (individuals do not smoke, do sleep well, and so on). The latent variables underlying each observed variable define the following equations:

$$\begin{array}{rcll} y^*_{il} & = & \gamma'_m W_{il} + \beta'_m Z_{ij} + \varepsilon_{il} & l = 1, \ldots, 6, \;\; i = 1, \ldots, n \\ y^*_{id} & = & \delta'_d Y_{il} + \beta'_d Z_{ij} + \varepsilon_{id} & j = 1, \ldots, J, \;\; h = 1, \ldots, H \end{array} \tag{5}$$

where $Y_{il} = \{y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6}\}$ is a vector of six lifestyles and each lifestyle is observed by the researcher as:

$$y_{il} = \begin{cases} 1 & \text{if } y^*_{il} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Using the matrix notation our recursive system of seven equations can be simplified by using the following matrix notation:

$$Y_{[(l+1)\times 1]} = \begin{pmatrix} \Gamma_{[(l+1)\times h]} & \Delta_{[(l+1)\times l]} & B_{[(l+1)\times j]} \end{pmatrix} \begin{pmatrix} W_{(h\times 1)} \\ Y_{(l\times 1)} \\ Z_{(j\times 1)} \end{pmatrix} + E_{[(l+1)\times 1]}$$

If Y_i is the random vector of the responses $\{Y_l, Y_d\}$, then the probability of observing a certain combination of responses on all seven variables, conditioned on parameters $\Theta = \{\Gamma, \Delta, B\}$, Ω is given by[27]:

$$P(\, Y_i = y_i \mid \Theta, \Omega) = \Phi_7(x_{i1d}, \ldots, x_{ild}) \tag{6}$$

where Φ_7 is the 7-dimensional multivariate standard normal distribution, $x_{ild} = d_{id} K_{il} \Theta' X$, $X = \{W, Y, Z\}$, and the matrix Ω has values of 1 on the leading diagonal and correlations between the error terms of the seven equations as off-diagonal elements. The errors terms of the latent equations have a multivariate normal distribution: $\varepsilon_i \sim MVN(0, \Sigma)$, where $\Sigma = \{\rho_{jk}\}$ is the correlation matrix, obtained considering the Choleski decomposition of the covariance matrix for the errors: $\Sigma = Cee'C'$, where e are independent standard normal random variables.

Chib and Greenberg (1997) suggest the use of the correlation matrix *for identifiability reasons*. The variances of the epsilon must be equal to one for the same reason, and the off-diagonal elements are symmetric. The parameters and the elements of Ω are not likelihood identified together, whereas the $J(J\text{-}1)/2$ parameters of Σ can be identified.

[27] In this formulation $K_{il} = 2y_{il} - 1$, for each i, l = 1, ..., 6, and $d_{id} = 2y_{id} - 1$.

Identification requires some exclusion restrictions. According to Schmidt (1981), simultaneous probit models suffer from identification problems. Imposing certain restrictions allows us to estimate a unique outcome of the latent y^*_{id}, for any value of the regressors and for any error term. This means that some exclusion restrictions have to be imposed in order to estimate all parameters[28]. The only restriction imposed is the recursivity of the model, given that the Z matrix in the deaths equation is just a portion of the regressor matrix in the lifestyle equations. Maddala (1983, p. 117-138) gives a good overview of simultaneous-equation and recursive models. Our model is of the type model 6, which is a simple recursive model[29]. If the random components of the latent equations are not independent and the matrix of regressors in the primary equation (mortality equation) includes all the regressors of the secondary equations (lifestyle equations), then the parameters in the equation of y^*_{id} are not identified. The identification restriction is that at least one variable in the lifestyle equations is not included in the deaths equations. In particular, we have chosen to exclude those variables which only indirectly influence the odds of dying through lifestyle variables. This holds with the assumption that not all factors determining the propensity to a healthy lifestyle influence also the death probability. Factors influencing both dependent variables may exist and may be hidden to the researcher: the multivariate probit model takes them into account. Occupational status, geographical location, drinking and smoking behaviors of the parents in the house, number of person living in the house, and characteristic of the house were considered to influence to a certain extent individual's preferences and decisions about health.

Evaluating the likelihood function, raises computational problems due to the fact that unobservable factors are jointly normal distributed. The log-likelihood function for our model has the form:

$$L = \sum_i \log \Phi_7(x_{i1d}, \ldots, x_{ild}) \tag{7}$$

Problems arise due to the numerical computation of multidimensional integrals. Here, the multivariate probit model is estimated in Stata using a GHK (Geweke-Hajivassilou-Keane) simulator for probabilities and a MSL procedure.

The GHK simulator exploits the Choleski decomposition of the covariance matrix, so that the joint probability originally based on unobservables can be written as the product of univariate conditional probabilities where the epsilon's are substituted by error terms, $u_i \sim \Phi_7(0, I_7)$, independent from each other by construction.

Although this simulation technique presents several advantages, the evaluation of the log-likelihood also requires another important stratagem to reduce

[28]In the multivariate probit model for SAH used by Contoyannis and Jones (2004), the sets of exogenous regressors are different since they are measured in two different waves: lifestyles of HALS1 are related to exogenous variables collected in the same survey, whereas SAH at HALS2 is related to exogenous variables at HALS2 and to lifestyles at HALS1. We cannot impose this restriction because we are using data collected in 1984, the first wave of the survey. Using the HALS2 data would restrict the sample size unnecessarily in this analysis of the 2003 deaths data.

[29]The responses (y^*_{il}, y^*_{id}) are observed as dichotomous variables, and y^*_{id} depends on y_{il}.

Table 8: Selected Partial Effects for alternative models of Mortality

	Full sample			**Men**			**Women**		
	MVP	**Exog**	**Excl**	**MVP**	**Exog**	**Excl**	**MVP**	**Exog**	**Excl**
S.C.									
sc12	-0.030†	-0.055*	-0.067**	0.012	0.003	-0.010	-0.037†	-0.087**	-0.096**
sc45	-0.015	0.001	0.012	0.002	0.034	0.050	-0.011	-0.026	-0.016
Education									
lhqdg	0.006	0.000	-0.010	0.031	0.024	0.012	-0.023	-0.038	-0.043
lhqhndA	-0.026	-0.035	-0.035	-0.023	-0.012	-0.001	-0.042	-0.055	-0.059
lhqnone	-0.006	0.039	0.061†	0.000	0.100†	0.127*	-0.012	-0.007	0.011
lhqoth	-0.024	-0.016	0.005	-0.057	-0.012	0.038	0.001	-0.005	-0.014
Significance levels: $\dagger$: 10% $*$: 5% $**$: 1%									

simulation bias. The simulated ML procedure using GHK at each iteration is numerically intensive. Indeed, even though $\check{P}_n(\theta)$ is unbiased for $P_n(\theta)$, ln $\check{P}_n(\theta)$ is not unbiased for ln $P_n(\theta)$. Only if the number of draws R grows at a rate that is faster than $\sqrt{N}$, the MSL estimator, which maximises the score function after plugging in the simulated probability, $\vartheta ln\check{P}_n(\theta)/\vartheta(\theta)$, is asymptotically consistent and efficient.

The MSL approach used in the Multivariate Probit Model is more efficient than the alternative procedure of separate ML estimation of the univariate probit models for the deaths equation and the lifestyle equations. The latter does not account for the correlation between the error terms, but rests on the assumption of exogeneity of the lifestyle covariates. Maddala (1983, p. 123) finds that if the error terms are not independent, the probit ML approach gives inconsistent estimates of the parameters. Maddala (1983) and Knapp and Seaks (1998) show that the log-likelihood function to be maximized in the multivariate probit model is equal to the sum of the log-likelihood functions obtained by the separate ML probit models when the restriction of independence of the errors is true. They propose two alternatives to the Hausman test for exogeneity of a dummy variable: a z test and a LR test[30]. The latter considers the likelihoods for the separate ML probit models as identical to equation (7) if the restriction $\rho_{jk} = 0$ holds. However, the LR test is not easy to calculate for each null of exogeneity because of the high number of marginal probabilities in the log-likelihood function. We use the statistic $z = \frac{\hat{\rho}}{S.E.(\hat{\rho})}$ for testing $H_o : \rho_{jk} = 0$. If the errors are independent, the MSL estimation is equivalent to the separate ML probit estimations. Hence, it is sufficient to test the unique restriction $\rho_{jk} = 0$, using the asymptotic standard errors provided by the MSL estimation.

The main results of the model (available from the authors upon request)

[30] An application can be found in Brown et al.'s paper (2004), which, analyzing the propensity to employment, considers the potential endogeneity of a binary variable indicating diabetes.

Table 9: Partial Effects of lifestyle for alternative models of mortality

	Full sample		**Men**		**Women**	
	MVP	**Exog**	**MVP**	**Exog**	**MVP**	**Exog**
nsmoker	-0.167**	-0.141**	-0.273**	-0.198**	-0.096†	-0.092**
breakfast	-0.045	-0.062**	-0.130	-0.061†	-0.002	-0.066*
sleepgd	-0.141*	-0.038*	-0.202*	-0.003	-0.231**	-0.066**
alqprud	-0.067	-0.030	0.090	-0.030	0.069	-0.040
nobese	-0.033	-0.072**	-0.008	-0.124*	-0.242**	-0.042
exercise	-0.230**	-0.046*	-0.161†	0.006	-0.040	-0.083**
Significance levels: † : 10% * : 5% ** : 1%						

can be summarized as follows[31]. The reference individual in the model is a married female, skilled full-time worker or student, with no formal qualification, she is a white European, living in an inner city somewhere in the south-east of England, she lives in rented accommodation, nobody else in the house smokes, and neither parent smoked. Table 8 reports a comparison between the partial effects of lifestyle computed for the multivariate probit and the univariate probit. Table 9 shows that the magnitude of the impact of lifestyle variables on mortality is bigger once we assume endogeneity, apart from breakfast and obesity[32]. These two lifestyles and drinking are not statistically significant. Not being a smoker, doing physical activity and sleeping well give a lower probability of dying: the estimated coefficients are highly statistically significant.

Concerning the other estimated coefficients we find similar results, except for the variable indicating the geographic area which is non significant in the endogenous model. The probability of death is a flexible function of age, increases for men and white Europeans and decreases for people in high social classes and shift workers. The impact of the variable sc12 is smaller in absolute value, the impact of sc45 changes direction. The impact of the higher level of schooling and of no schooling have another direction and are bigger in absolute value. Accounting for endogeneity alters substantially the impact of lifestyle on mortality and corrects for the bias due to the exogeneity assumption. The null of exogeneity is only rejected in two cases, as reported in Table A.5. A statistically significant correlation exists between the disturbances of the deaths equation and each of the disturbances of the equations for smoking and exercise. The correlation coefficients are positive, meaning that unobservables that increase the probability of being a non smoker and making physical activity also increase the probability of death. This result implies that frailer individuals select into non-smoking. We can easily argue that the probability of dying

[31]The MSL estimates of the coefficients and their statistical significance for our system of equations are computed by the software STATA, using the command *mvprobit* created by Cappellari and Jenkins. For more details about the algorithm see Cappellari and Jenkins, 2003.

[32]It is worth noting that obesity, even if not significant, has a negative sign. This result contradicts the results of the multivariate probit for the SAH model in Contoyannis and Jones (2004)that, counterintuitively, suggested that obesity causes higher SAH.

Table 10: Summary table for percentage reduction in social class and education gradient when lifestyles are included

Gradient Reduction	**Full sample**	**Men**	**Women**
Social Class			
Exogenous	29.18	48.22	23.36
MVP	80.97	116.11	66.55
Education			
Exogenous	44.65	34.07	42.24
MVP	116.69	126.80	79.53

increases because of unobservables such as a bad health history: mortality is higher for frailer individuals. The statistical significance of the correlation coefficients shows that frailer people tend to adopt healthier lifestyles (ie., they do not smoke or quit smoking, they usually do physical activities) than persons with a better health who, by contrast, are more likely to behave in less healthy ways. Our model, contrasting with Adda and Lechene (2004) who found that people in poor health select into smoking, suggests that people in poor health do not select into smoking and that non-smokers are less likely to die even if they are frailer[33]. Smoking and exercise are positively correlated and the correlation coefficient is statistically significant. These findings confirm the interpretation of the qualitative effect of these lifestyles on the odds of dying derived from the sign of the estimated coefficients.

Table 10 shows the comparison between the range measures obtained by the estimation of the deaths equation without lifestyles and the deaths equation with potentially endogenous lifestyles. The social class gradient reduces by around 81 percent in the MSL estimates. The reduction of the education gradient is even bigger: the range of inequality between extreme educational classes is reduced 1.17 times.

The multivariate probit for men predicts a more influential impact of lifestyle variables on mortality risk in term of the magnitude of their partial effects. However, the health-related behaviors that are statistically significant, smoking, good sleeping and exercise, are different from the univariate estimation: the diet variables and drinking do not contribute to explain the risk of mortality. It is worth noting that, even if it is not statistically significant, prudent alcohol consumption turns out to have a different impact on mortality with respect to the exogenous model. Its sign is positive, meaning that moderate drinking could have positive effects on health. The impact of the higher

[33]Adda and Lechene (2004), in an attempt to overcome the weaknesses of the medical, epidemiological and economic literature dealing with the health production function and the problems of individual heterogeneity and endogeneity, used a tobacco-free morbidity score in a duration model on BHPS data. Assuming an *a priori* correlation between smoking and mortality, they found that poor people are more likely to be smokers because they have a shorter life expectancy. The cost of smoking is not so high with respect to the number of years they expect to live given their health condition. More important, they found a strong cohort effect, which shows the role of information about smoking-related risks on the choice of smoking.

socioeconomic classes increases, while the impact of the lower classes tend to diminish in the MSL estimation. The null of exogenity is rejected in three cases: for smoking, sleeping well and exercise. The correlation coefficients are positive. The reduction of the range of inequality is very pronounced in this model: the social class gradient is 1.16 times smaller with respect to the deaths equation without lifestyle variables, and the education gradient about 1.3 times smaller.

For women, smoking, sleeping well and obesity are the statistically significant health-related behaviors. The impact of all lifestyles, except breakfast, is bigger than in the model with exogenous lifestyle variables. Also in this case, drinking is not statistically significant but its effect is anyway negative. The impact of socioeconomic indicators is generally smaller than in the model without lifestyle variables. Only sleeping and obesity are endogenous, with positive correlation coefficients. However, these two variables are not correlated with each other. Even in this case, the reduction of the range of inequality is much more important then in the exogenous model. The social class gradient is reduced by about 67 percent, while the education gradient decreases by around 80 percent when endogeneity is corrected for.

We can consider the recursive system estimated by MSL for the three samples as a 'general' model and then restrict it to a reduced system, which excludes the equations of the lifestyles variables for which the null hypothesis of exogeneity was not rejected. Moving from the 'universal' to the 'particular' case, we expect to get more efficient estimates. Table A.6 reports the estimates of the correlation coefficients and their significance level. For the pooled sample, smoking and exercise are still endogenous. For men, smoking and exercise are still endogenous but the null of exogeneity is not rejected for sleeping. For women, the null is rejected for obesity but not for sleeping. It is worth noting that, in the three cases, only lifestyles which were correlated between each other in the 'general' model remain endogenous.

When the null of exogeneity is rejected, we say that unobservables influence both mortality risk and the probability of having a certain health-related behavior. When more than one correlation coefficient is statistically significant, then we can say that the same unobservables influence the odds of dying and the probability of having health-related behaviors, irrespective of the particular behavior. The idea is that, for example, the same genetic characteristics or the same past experiences drive the individual to choose to be a non-smoker and to do sport. For this reason, it is not surprising that the correlation coefficient for the smoking and exercise equations is statistically significant. For women, the correlation coefficient between obesity and sleeping shows absence of correlation between the probability of having these health-related behaviors even in the 'general' system. We do not expect to find that both these variables are still endogenous once the recursive system is reduced to a smaller number of equations.

The measure of inequality given by the range predicts a notable reduction of the socioeconomic dimension of inequality in the distribution of health. However, the limits of this measure induced us to look for more robust and

meaningful indicators of inequality in health.

8 The Gini measure of total health inequality

In this section we apply one of the most robust measures of health inequality in order to overcome the drawbacks of the range used in the previous part of the paper. The literature on income distribution and income inequality has been of help to health researchers, who have adapted income inequality measurements to fit the distribution of health. Socioeconomic inequalities in health and pure inequality in health can be easily measured. The former can be thought of as a subset of the latter, where the population is ranked on the basis of individual income or social position and not of health.

Both the concentration index (CI) of income-related inequality and the Gini coefficient of health inequality could be used in this framework, however the particular nature of our data favours the Gini coefficient[34].

As Lerman and Yitzhaki (1989) show, the Gini can be expressed by:

$$G = \frac{2}{\bar{y}} cov[y, F(y)] \tag{8}$$

where $F(y)$ is the cumulative distribution of health and $\bar{y}$ is the mean of health. The estimator of $F(y)$ in a random sample is the rank of y divided by the sample size. Our health indicator is a binary variable. van Doorslaer and Jones (2003), dealing with an ordered categorical dependent variable for SAH from the Canadian National Population Health Survey, used the predictions from ordered probit or interval regressions to analyze total health inequality. We use predicted mortality, that is the linear index for death predicted after a probit estimation of the deaths equation presented in section 6. The advantage of using predictions is that they allow us to give a different value of health (in this case mortality) to each individual in the sample.

[34]The British HALS did not collect a continuous measure for income. Apart from bands of income, which suffer a high rate of item non-response, the only available information about the economic position of the individual is given by the social class classification. Respondents are classified according to the belonging socioeconomic group, but we cannot associate a continuous average value of income to groups. Since the indicator of social class is an ordinal categorical variable (it takes values from one to three if the individual belongs to SC12, SC3 or SC45), it is not possible to generate the concentration curve measuring the degree of income-related inequalities, by plotting the cumulative proportion of the population ranked by social class (x-axis) against the cumulative proportion of health (y-axis). Problems arising in the generation of the fractional rank, lead to unreliable estimates of the covariance between the measure of health and the fractional rank. The routine that generates the rank is very straightforward and it is available in the STATA through the user-written command –glcurve7– or through the command –egen rank (), unique–. These commands require the user to sort the dataset by the ranking variable. It can be shown that if this variable does not uniquely identify each observation, but only groups of observation, then the associated fractional rank will be generated as a different ordering each time.

Individuals in the sample are ranked by their predicted mortality. The estimated linear index allows a sufficient degree of individual variation in the measure of mortality but it associates to each individual either positive or negative values of mortality. The index is transformed in order to guarantee a positive support and to ensure that the Gini coefficient is positive[35]. The Gini coefficient for the excluded and the exogenous deaths equation, is 0.280 and 0.261 respectively. (These values are 0.292 and 0.276 for men, and 0.303 and 0.270 for women). The measure of pure inequality in health, irrespective of any socioeconomic dimension, is a bit smaller if health-related behaviours are considered as exogenous determinants of the health outcome. The MSL estimation predicts a lower level of inequality, around 0.236 for the pooled sample, 0.237 for men and a bit higher level, 0.276, for women.

8.1 Decomposition of total health inequality

The Gini measure of overall health inequality *per se* does not allow us to make a comparison with the measure of range presented in section 7, which captures the socioeconomic dimension of inequality. We are more interested in revealing the contribution of each determinant of mortality to inequality.

Wagstaff et al. (2001) stressed the importance of "unpacking" the causes of socioeconomic inequalities, by decomposing the CI. Inequalities depend both on the direct impact that the various determinants of health (e.g., lifestyles, parental factors, geography, income, education, ethnicity) have on the health outcome and on the distribution of these determinants across socioeconomic groups. Morris et al. (2003) decomposed both the CI and the Gini coefficient for the use of health care in England. Also van Doorslaer and Jones (2003) decomposed both indexes, using SAH and the Health Utility Index (HUI). They proposed a framework that reduces problems affecting decomposition due to non linearity in the determination of health.

For the linear regression model

$$y_i = \sum_k \beta_k x_{ki} + \varepsilon_i \tag{9}$$

the Gini can be written as an addictively decomposable form

$$G_{ini} = \sum_k \left(\frac{\beta_k \bar{x}_k}{\bar{y}_i} \right) C_k + \frac{GC_\varepsilon}{\bar{y}_i} = \sum_k \eta_k C_k + \frac{GC_\varepsilon}{\bar{y}_i} \tag{10}$$

The first component of (10) is the explained part of inequality: it is a weighted sum of the CI of the regressors, where the weights are the elasticity, η_k, of mortality with respect to each regressor x_k. The second component is the generalized concentration index for the error term, which can be computed as

[35]The definition of the new dependent variable requires that the linear index $x\hat{b}$ is transformed in $x\hat{b}^*$, where $x\hat{b}^* = x\hat{b} - min(x\hat{b})$, which satisfies the condition $x\hat{b}^* \geq 0$. van Doorslaer and Jones (2003) note that the percentage contributions in the regression-based decomposition analysis are invariant to linear transformation of y.

Table 11: Percentage contributions to overall inequality

Variable	Excl.	Exog.	MVP	MVP reduced
Full sample				
lifestyles	-	4.35	37.17	33.61
sc	2.39	1.68	0.74	0.71
edu	3.81	2.38	0.09	0.19
work	1.56	1.67	1.53	1.40
area	1.12	0.79	0.21	0.26
ethnicity	0.45	0.48	0.43	0.40
sex	5.57	5.21	3.46	3.82
age	85.10	83.44	56.38	59.60
Men				
lifestyles	-	1.96	42.44	37.85
sc	0.96	0.43	-0.18	-0.08
edu	5.40	3.91	0.11	0.40
work	2.43	2.50	2.85	3.13
area	0.38	0.27	0.02	0.03
ethnicity	0.76	0.77	0.75	-1.99
age	90.08	90.15	54.00	60.66
Women				
lifestyles	-	10.02	42.07	29.31
sc	4.89	3.82	1.72	2.50
edu	3.33	1.74	0.67	1.03
work	0.81	0.97	0.62	0.59
area	2.93	2.14	1.16	1.38
ethnicity	0.20	0.26	0.28	0.32
age	87.83	81.04	53.47	64.87

a residual. The C_k are the health-related CI of each regressor. The larger η_k and C_k are, the bigger is the importance of x_k in accounting for inequality in health. If C_k is small and elasticity is still big, the regressor x_k is important to explain mortality but not to explain inequality in mortality. Our dependent variable, predicted mortality, is additive in the regressors and it only permits us to estimate only the deterministic part of (10).

Table 11 presents the components of (10) as percentage contributions. The most important contributions to overall health inequality are attributable to age and gender. Apart from age and gender, education and social class make a relative high contribution (more than 2 per cent). These contributions turn out to be smaller if lifestyles are included in the model. Lifestyles contribute 4.35 per cent. As regards men, social class does not strongly contribute relative to education and work status, which contribute respectively by 5.4 per cent and around 2.4 per cent. Once lifestyles are in the model, the contribution of education decreases. Lifestyles have very little importance in the measure of overall health inequality for men. For women, we find high contributions of

social class, education and geography. However, these contributions decrease due to measured lifestyles factors, the latter making a contribution of 10 per cent.

Controlling for potential endogeneity of the lifestyle variables by means of the multivariate probit, gives very strong results in terms of the contribution of socioeconomic variables to overall inequality in mortality. For the pooled sample, social class's contribution diminishes by 30 per cent, falling from 2.39 without lifestyles to 0.74; education's contribution is reduced by more than 2 times, falling from 3.81 to 0.09. The extent of the contribution of lifestyles is very large in the endogenous model, which predicts a 37 per cent contribution to overall inequality, due in particular to the variable exercise. For men, it is worth noting the interesting result, illustrated in Table 11, about the contribution of social class. This decreases so much that it changes direction: social class has a negative contribution to inequality. This result derives from the negative health-related CI for sc12, meaning that the concentration of high social classes is in favour of people with low mortality risk, which is meaningful and consistent with the positive CI for sc45 indicating the concentration of the lowest social classes is pro-high mortality risk people. Both the estimated coefficients have a positive sign but the contribution to the overall Gini of sc12 offset the contribution of sc45. The work status variable, that can be seen as an indicator of social position by itself contributes almost 3 per cent, which is the biggest contribution after age and lifestyles.

Considering the estimation of the reduced recursive system shown in section 7, the Gini coefficient is slightly bigger (0.246 for the full sample, 0.242 for men and 0.375 for women) but the results of the decomposition analysis do not differ too much using the estimated coefficients from this model.

The measurement of total health inequality through the Gini coefficient and the estimation of the contribution of socioeconomic variables to its variation, capture the effect of including health-related behaviors in the deaths equation. Even though the Gini coefficient and the Lorenz curve do not measure the socioeconomic dimension of inequalities in health directly, they do shed light on the nature of inequality by means of the decomposition approach. We can conclude that health-related inequalities exist in the distribution of social class and education. The estimated C_ks for social classes and educational qualifications can be thought of as a "mirror image" of the CI for income-related inequality. These findings are stronger than the comparison between the range measures calculated with or without lifestyles.

9 Conclusion

We use the British Health and Lifestyle Survey (HALS, 1984-1985) data and the longitudinal follow-up of May 2003 to investigate the determinants of premature mortality risk in Great Britain.

A simple behavioral model, which relates premature mortality to a set of

observable and unobservable factors, is considered. Observable factors influencing mortality are socioeconomic and demographic characteristics, ethnicity, geography and individual health-related behaviors. Individuals' choices about their lifestyle may induce variations in health status and to affect premature mortality. We assume that the relationship between the socioeconomic environment and premature mortality is mediated by lifestyles. In order to assess the impact of lifestyles, we estimate probit models and compare models without lifestyles and models which include them.

Two main econometric issues arise in our analysis: unobservable individual heterogeneity and endogeneity of the discrete explanatory variables that affect the mortality equation. Factors hidden to the researcher, like the rate of time preference, biological or genetic characteristics and past experiences, may influence individual demand for health and for health inputs. These factors have an indirect effect on mortality by influencing the marginal utility of health, wealth, education and consumption of commodities other than health. We propose a MSL approach to estimate a recursive system of equations for deaths and lifestyles, in order to correct for heterogeneity and potential endogeneity of lifestyles. The multivariate probit model allows us to test whether unobservable characteristics influencing lifestyle also affect premature mortality. This model is then compared with a model without lifestyles and with a model that includes exogenous lifestyle variables.

The main economic concern is to detect inequality in the distribution of health within the population and to understand to what extent differences in social and economic characteristics contribute to inequality. We focus mainly on social class and schooling differences in the sample. We are critical of the range measure of inequality used by epidemiologists and by Contoyannis and Jones (2004). The range is compared to more robust measures of health inequality offered by the decomposition of the Gini coefficient for overall health inequality. We are able to decompose the Gini for predicted premature mortality and to compute health-related concentration indexes for all factors influencing mortality, including social class and education. Their contribution to the total Gini is then compared with the range measure. We find that endogenous lifestyles and unobservable heterogeneity strongly contribute to inequality in mortality, reducing the contribution of socioeconomic factors.

References

Adda, J. and Lechene, V. (2004). On The Identification Of The Effect Of Smoking On Mortality. Royal Economic Society Annual Conference 2004.

Auster, R., Levenson, I., and Sarachek, D. (1969). The production of health: an exploratory study. *Journal of Human Resources*, 4:411–436.

Becker, G. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 75(299):493–517.

Belloc, N. B. (1973). Relationship of Health Practices and Mortality. *Preventive Medicine*, 2:67–81.

Belloc, N. B. and Breslow, L. (1972). Relationship of Physical Health Status and Health Practices. *Preventive Medicine*, 1:409–421.

Borg, V. and Kirstensen, T. S. (2000). Social class and self-rated health: can the gradient be explained by differences in life style or work environment? *Social Science and Medicine*, 51:1019–1030.

Brown, H. S., Pagán, J. A., and Bastida, E. (2004). The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit. *Hispanic Health Research, The University of Texas*.

Cappellari, L. and Jenkins, S. P. (2003). Multivariate Probit regression using simulated maximum likelihood. *The Stata Journal*, 3(3):278–294.

Chib, S. and Greenberg, E. (1997). Analysis of Multivariate Probit. *Biometrika*, 85(2):347–361.

Contoyannis, P. and Jones, A. M. (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics*, 23(5):965–995.

Deaton, A. (2003). Health, Inequality and Economic Development. *Journal of Economic Literature.*, 41(1):113–158.

Fuchs, V. R. (1982). Time preferences and health: An explanatory study. in V. R. Fuchs, ed., Economic Aspect of Health, University of Chicago Press for NBER, Chicago, IL.

Gardner, J. and Oswald, A. (2004). How is mortality affected by money, marriage and stress? *Journal of Health Economics*. forthcoming.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *The Journal of Political Economy*, 80(2):223–255.

Grossman, M. and Joyce, T. J. (1990). Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City. *The Journal of Political Economy*, 98(5):983–1007.

Health and Lifestyle Survey - University of Cambridge Clinical School (2003). Deaths (4^{st} update) and Cancer (1^{st} Update - 2^{nd} listing) May 2003 - Working Manual. University of Cambridge, available at the web site http://www.data-archive.ac.uk/.

Hurley, J. (2000). An overview of the normative economics of the health sector. *in Culyer, A.J., and J. P. Newhouse (eds.) Handbook of Health Economics, Elsevier.*

Kenkel, S. D. (1991). Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, 99(2):287–305.

Kenkel, S. D. (1995). Should you eat breakfast? estimates from health production functions. *Health Economics*, 4:15–25.

Knapp, L. G. and Seaks, T. G. (1998). A Hausman test for a dummy variable in probit. *Applied Economics Letters*, 5:321–323.

Lerman, R. I. and Yitzhaki, S. (1989). Improving the accuracy of estimates of gini coefficients. *Journal of Econometrics*, 42:43–47.

Lynch, J. W., Kaplan, G. A., Cohen, R. D., Tuoomilehto, J., and Salomen, J. T. (1996). Do Cardiovascular Risk Factors Explain the Relation between Socioeconomic Status, Risk of All-Cause Mortality, Cardiovascualr Mortality, and Acute Myocardial Infraction? *American Journal of Epidemiology*, 144(10):934–941.

Lynch, J. W., Kaplan, G. A., and Salomen., J. T. (1997). Why do poor people behave poorly? *Social Science and Medicine*, 44(6):809–819.

Maddala, G. S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge Unversity Press.

Manor, O., Mattheus, S., and Power, C. (1997). Comparing measure of health inequality. *Social Science and Medicine*, 45(5):761–771.

Morris, S., Sutton, M., and Gravelle, H. (2003). Inequity and inequality in the use of health care in england: an empirical investigation. *CHE Technical Paper Series 27.*

Mullahy, J. and Portney, P. (1990). Air Pollution, Cigarette Smoking, and the Production of Respiratory Health. *Journal of Health Economics*, 9:193–205.

Mullahy, J. and Sindelar, J. (1996). Employment, Unemployment, and Problem Drinking. *Journal of Health Economics*, 15:409–434.

Poston, D. L. and Micklin, M. (2005). *Handbook of Population.* Handbooks of Sociology and Social Research Series. Kluwer Academic Publishers.

Schmidt, P. (1981). Constraints on the Parameters in Simultaneous Tobit and Probit Models. 12 in Structural Analysis of Discrete Data and Econometric Applications by Manski, C. F. and D. L. McFadden, The MIT Press, Cambridge.

van Doorslaer, E. and Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics*, 22:61–87.

Wagstaff, A. (1986). The demand for health: theory and applications. *Journal of Epidemiology and Community Health*, 40:1–11.

Wagstaff, A., Paci, P., and Joshi, H. (2001). Causes of inequalities in health. Who you are? Where you live? Or who your parents are? *Policy Research Working Paper 2713, World Bank.*

Wagstaff, A., Paci, P., and van Doorslaer, E. (1991). On the measurements of inequalities in health. *Social Science Medicine*, 33(5):545–557.

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

Appendix A

Table A.1: Variable Definitions

Variable Name	Variable Definition
death	1 if has died at May 2003, 0 alive
Lifestyle	
nsmoker	1 if does not smoke, 0 if current smoker
breakfast	1 if does a healthy breakfast, 0 otherwise
sleepgd	1 if sleeps between 7 and 9 hours, 0 otherwise
alqprud	1 if consume alcohol prudently, 0 otherwise
nobese	1 if is not obese, 0 otherwise
exercise	1 if did physical exercise in the last fortnight, 0 otherwise
Social Class	
sc12	1 if professional/student or managerial/intermediate, 0 otherwise
sc3	1 if skilled or armed service, 0 otherwise
sc45	1 if partly skilled, unskilled, unclass. or partner never occupied, 0 otherwise
Education Level	
lhqdg	1 if University degree, 0 otherwise
lhqhndA	1 if higher vocational qualifications or A level or equivalent, 0 otherwise
lhqO	1 if O level/CSE, 0 otherwise
lhqnone	1 if no qualification, 0 otherwise
lhqoth	1 if other vocational/professional qualifications, 0 otherwise
Marital Status	
married	1 if married, 0 otherwise
widow	1 if widow, 0 otherwise
divorce	1 if divorced, 0 otherwise
seprd	1 if separated, 0 otherwise
single	1 if single, 0 otherwise
Occupational Status	
wkshft1	1 if shift worker, 0 otherwise
full	1 if full-time worker, 0 otherwise
part	1 if part-time worker, 0 otherwise
unemp	1 if the individual unemployed, 0 otherwise
sick	1 if absent from work due to sickness, 0 otherwise
retd	1 if retired, 0 otherwise
stdnt	1 if student, 0 otherwise
keephs	1 if housekeeper, 0 otherwise

continued on next page

Table A.1 – *continued from previous page*

Variable Name	Variable Definition
Area	
wales	1 if lives in Wales, 0 otherwise
north	1 if lives in North, 0 otherwise
nwest	1 if lives in North West, 0 otherwise
yorks	1 if lives in Yorkshire, 0 otherwise
wmids	1 if lives in West Midlands, 0 otherwise
emids	1 if lives in East Midlands, 0 otherwise
anglia	1 if lives in East Anglia, 0 otherwise
swest	1 if lives in South West, 0 otherwise
london	1 if lives in London, 0 otherwise
scot	1 if lives in Scotland, 0 otherwise
incity	1 if lives in the city, 0 otherwise
rural	1 if lives in the countryside, 0 otherwise
suburb	1 if lives in the suburbs of the city, 0 otherwise
Ethnicity	
ethipb	1 if Packistani or Bangladeshi
ethbawi	1 if Black, African or West Indian, 0 otherwise
ethothnw	1 if non-white, 0 otherwise
ethwheur	1 if White European, 0 otherwise
Physical	
male	1 if male, 0 otherwise
height	height in inches
age	age in years
age2	$age^2/100$
age3	$age^3/10000$
age4	$age^4/1000000$
Tenure	
housown	1 if own house, 0 otherwise
Household	
hou	number of other people in the house
Parental	
mothsmo	1 if only mother smoked, 0 otherwise
fathsmo	1 if only father smoked, 0 otherwise
bothsmo	1 if both parents smoked, 0 otherwise
smother	1 if anyone else in house smoked, 0 otherwise
alpa	father non to heavy drinker (0-4)
alma	mother non to heavy drinker (0-4)

Table A.2: Summary statistics

Variable	**Full sample** N=3670		**Men** N=1671		**Women** N=1999	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
death	0.359	0.48	0.43	0.495	0.3	0.458
nsmoker	0.700	0.458	0.683	0.465	0.713	0.452
breakfast	0.707	0.455	0.698	0.459	0.714	0.452
sleepgd	0.583	0.493	0.577	0.494	0.587	0.492
alqprud	0.88	0.326	0.808	0.394	0.939	0.239
nobese	0.854	0.353	0.912	0.283	0.805	0.396
exercise	0.323	0.468	0.33	0.47	0.318	0.466
sc12	0.316	0.465	0.303	0.46	0.326	0.469
sc3	0.466	0.499	0.481	0.5	0.454	0.498
sc45	0.219	0.413	0.216	0.412	0.221	0.415
lhqdg	0.126	0.332	0.136	0.343	0.117	0.322
lhqhndA	0.125	0.331	0.123	0.329	0.126	0.332
lhq0	0.094	0.292	0.089	0.284	0.099	0.298
lhqnone	0.608	0.488	0.583	0.493	0.629	0.483
lhqoth	0.047	0.213	0.069	0.253	0.03	0.169
full	0.364	0.481	0.557	0.497	0.203	0.402
part	0.132	0.338	0.027	0.162	0.219	0.414
unemp	0.03	0.171	0.052	0.222	0.012	0.109
sick	0.033	0.179	0.052	0.222	0.017	0.129
retd	0.339	0.473	0.309	0.462	0.364	0.481
keephse	0.102	0.303	0.004	0.06	0.185	0.388
wkshft1	0.057	0.232	0.086	0.28	0.034	0.18

Table A.3: Probit Model: Excluded Equation

Variable	**Full sample**		**Men**		**Women**	
	dF/dx	S.D.	dF/dx	S.D.	dF/dx	S.D.
sc12	-0.067**	0.022	-0.010	0.039	-0.096**	0.025
sc45	0.012	0.023	0.050	0.037	-0.016	0.026
lhqdg	-0.010	0.042	0.012	0.068	-0.043	0.048
lhqhndA	-0.035	0.040	-0.001	0.066	-0.059	0.045
lhqnone	0.061†	0.033	0.127*	0.054	0.011	0.040
lhqoth	0.005	0.052	0.038	0.075	-0.014	0.072
wkshft1	-0.106**	0.037	-0.121*	0.051	-0.062	0.067
rural	-0.076**	0.024	-0.037	0.040	-0.100**	0.027
suburb	-0.035†	0.020	-0.038	0.033	-0.031	0.025
ethwheur	0.096	0.059	0.148	0.095	0.048	0.076
male	0.159**	0.018	0.159**	0.018	0.159**	0.018
age	-0.004	0.063	0.063	0.127	-0.008	0.071
age2	0.043	0.104	-0.068	0.215	0.044	0.116
age3	-0.019	0.056	0.046	0.120	-0.020	0.062
N	3670		1671		1999	
Log-likelihood	-1656.8031		-776.44667		-871.29684	
χ^2	$\chi^2_{(14)} = 1478.79$		$\chi^2_{(13)} = 730.45$		$\chi^2_{(13)} = 700.15$	
Significance levels: † : 10% * : 5% ** : 1%						

Table A.4: Probit Model: Included Equation

Variable	**Full sample**		**Men**		**Women**	
	dF/dx	S.D.	dF/dx	S.D.	dF/dx	S.D.
nsmoker	-0.141**	0.021	-0.198**	0.033	-0.092**	0.027
breakf t	-0.062**	0.021	-0.061†	0.033	-0.065*	0.027
sleepgd	-0.038*	0.018	-0.003	0.029	-0.066**	0.022
alqprud	-0.030	0.029	-0.030	0.038	-0.040	0.049
nobese	-0.072**	0.026	-0.124*	0.051	-0.042	0.028
exercise	-0.046*	0.020	0.006	0.032	-0.082**	0.023
sc12	-0.055*	0.022	0.003	0.040	-0.087**	0.025
sc45	0.001	0.023	0.034	0.038	-0.026	0.026
lhqdg	-0.000	0.043	0.024	0.069	-0.038	0.049
lhqhndA	-0.035	0.040	-0.012	0.067	-0.055	0.045
lhqnone	0.039	0.034	0.100†	0.055	-0.007	0.041
lhqoth	-0.016	0.051	-0.012	0.076	-0.005	0.073
wkshft1	-0.122**	0.035	-0.136*	0.051	-0.080	0.061
rural	-0.059*	0.024	-0.020	0.041	-0.081**	0.028
suburb	-0.026	0.020	-0.032	0.033	-0.018	0.024
ethwheur	0.109†	0.057	0.165	0.092	0.066	0.071
male	0.161**	0.019	0.161**	0.019	0.161**	0.019
age	-0.022	0.063	0.046	0.128	-0.015	0.071
age2	0.075	0.104	-0.035	0.216	0.056	0.115
age3	-0.036	0.056	0.028	0.120	-0.026	0.061
N	3670		1671		1999	
Log-likelihood	-1613.2974		-750.7993		-846.126	
χ^2	$\chi^2_{(20)} = 1565.80$		$\chi^2_{(19)} = 781.74$		$\chi^2_{(19)} = 750.49$	
Reset test	$prob > \chi^2 = 0.6138$		$prob > \chi^2 = 0.6804$		$prob > \chi^2 = 0.2209$	
$\chi^2_{(1)}$	0.25		0.17		1.50	

Significance levels: † : 10% * : 5% ** : 1%

Table A.5: Correlation Coefficient in the MVP model for mortality

	Full sample		**Men**		**Women**	
rho	**coeff**	**s.d.**	**coeff**	**s.d.**	**coeff**	**s.d.**
rho21	0.247*	0.101	0.467**	0.157	0.040	0.144
rho31	0.101	0.134	0.334	0.220	-0.061	0.174
rho41	0.285	0.174	0.487*	0.202	0.432*	0.188
rho51	0.119	0.150	-0.163	0.162	-0.242	0.181
rho61	-0.010	0.143	-0.104	0.198	0.458**	0.143
rho71	0.486**	0.111	0.438*	0.188	-0.032	0.184
rho32	0.259**	0.029	0.220**	0.044	0.303**	0.040
rho42	0.030	0.029	0.001	0.043	0.071†	0.040
rho52	0.187**	0.038	0.177**	0.049	0.180**	0.062
rho62	-0.229**	0.035	-0.276**	0.060	-0.195**	0.045
rho72	0.105**	0.030	0.128**	0.046	0.071†	0.043
rho43	0.112**	0.028	0.140**	0.042	0.101**	0.039
rho53	0.269**	0.036	0.254**	0.048	0.302**	0.060
rho63	0.073*	0.035	0.037	0.062	0.093*	0.044
rho73	0.075*	0.030	0.084†	0.045	0.068	0.042
rho54	0.045	0.036	0.047	0.047	0.055	0.061
rho64	0.035	0.033	0.036	0.057	0.033	0.042
rho74	-0.011	0.028	0.001	0.042	-0.009	0.039
rho65	-0.046	0.049	0.017	0.067	-0.133†	0.074
rho75	-0.078*	0.037	-0.094†	0.049	-0.042	0.061
rho76	0.121**	0.036	0.232**	0.061	0.065	0.045

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%

Table A.6: Correlation Coefficient in the MVP model for mortality - reduced models

	Full sample		**Men**		**Women**	
rho	**coeff**	**s.d.**	**coeff**	**s.d.**	**coeff**	**s.d.**
rho21	0.270**	0.105	0.404**	0.132	0.190	0.222
rho31	0.530**	0.109	0.295	0.302	0.486**	0.174
rho41	0.565**	0.173	0.565**	0.173	0.000	0.000
rho32	0.108**	0.030	-0.000	0.043	0.041	0.042
rho42	0.142**	0.046	0.142**	0.046	0.000	0.000
rho43	-0.003	0.042	-0.003	0.042	0.000	0.000

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%

Appendix B

Full Results - Not for Publication

Table B.1: MVProbit Model for the full sample

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
nsmoker	-0.650**						
	(0.176)						
breakfast	-0.178						
	(0.227)						
sleepgd	-0.544*						
	(0.276)						
alqprud	-0.264						
	(0.280)						
nobese	-0.129						
	(0.256)						
exercise	-0.900**						
	(0.187)						
sc12	-0.120†	0.138*	0.053	0.025	-0.252**	0.191**	-0.017
	(0.066)	(0.060)	(0.059)	(0.055)	(0.076)	(0.069)	(0.057)
sc45	-0.060	-0.078	-0.044	0.023	-0.178*	0.012	-0.154*
	(0.062)	(0.059)	(0.059)	(0.055)	(0.077)	(0.067)	(0.060)
lhqdg	0.022	0.102	0.149	0.043	-0.012	0.121	-0.024
	(0.114)	(0.108)	(0.106)	(0.095)	(0.126)	(0.125)	(0.095)
lhqhndA	-0.105	-0.046	-0.057	-0.040	-0.101	-0.046	-0.021
	(0.111)	(0.102)	(0.101)	(0.093)	(0.123)	(0.117)	(0.093)
lhqnone	-0.025	-0.217**	-0.277**	-0.076	0.025	-0.040	-0.273**
	(0.096)	(0.085)	(0.084)	(0.078)	(0.105)	(0.097)	(0.079)
lhqoth	-0.097	-0.345**	-0.239†	0.057	-0.044	0.038	-0.059
	(0.140)	(0.128)	(0.129)	(0.122)	(0.156)	(0.159)	(0.123)
wkshft1	-0.446**	-0.189†	-0.158	-0.294**	-0.086	-0.214†	-0.203*
	(0.120)	(0.099)	(0.097)	(0.095)	(0.115)	(0.117)	(0.099)
rural	-0.059	0.103	0.108	0.052	0.161†	0.086	0.183**
	(0.072)	(0.069)	(0.068)	(0.064)	(0.088)	(0.077)	(0.067)
suburb	-0.020	0.018	0.150**	-0.050	0.086	0.195**	0.167**
	(0.058)	(0.055)	(0.054)	(0.051)	(0.069)	(0.062)	(0.054)
ethwheur	0.382†	0.086	0.473**	0.387*	-0.595*	0.186	0.159
	(0.199)	(0.165)	(0.153)	(0.153)	(0.292)	(0.182)	(0.167)
male	0.380**	-0.229**	-0.114	0.074	-0.705**	0.443**	0.115†
	(0.069)	(0.073)	(0.072)	(0.065)	(0.093)	(0.084)	(0.069)
age	-0.124	-0.001	0.243	0.346**	-0.175	-0.551**	-0.370*

continued on next page

Table B.1 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
	(0.168)	(0.170)	(0.156)	(0.134)	(0.189)	(0.170)	(0.169)
age2	0.280	-0.018	-0.382	-0.606**	0.333	0.853**	0.576*
	(0.278)	(0.287)	(0.262)	(0.223)	(0.315)	(0.281)	(0.286)
age3	-0.131	0.043	0.204	0.329**	-0.186	-0.420**	-0.318*
	(0.150)	(0.157)	(0.142)	(0.119)	(0.170)	(0.150)	(0.157)
widow		-0.219**	-0.194*	-0.196**	-0.053	-0.254**	-0.009
		(0.085)	(0.085)	(0.075)	(0.117)	(0.090)	(0.082)
divorce		-0.400**	-0.270*	-0.285*	0.095	-0.088	0.238*
		(0.121)	(0.120)	(0.113)	(0.158)	(0.141)	(0.114)
seprd		-0.251	-0.302†	-0.295†	-0.075	-0.068	0.170
		(0.175)	(0.173)	(0.164)	(0.212)	(0.211)	(0.165)
single		-0.146	-0.092	-0.183†	-0.060	-0.093	-0.256*
		(0.111)	(0.112)	(0.101)	(0.138)	(0.127)	(0.111)
part		-0.126	0.147†	0.319**	0.172	0.055	0.109
		(0.083)	(0.083)	(0.081)	(0.114)	(0.098)	(0.077)
unemp		-0.411**	-0.304*	0.184	-0.216	-0.126	-0.112
		(0.131)	(0.129)	(0.130)	(0.146)	(0.160)	(0.131)
sick		-0.321*	-0.214	-0.272*	0.374*	0.070	-0.808**
		(0.133)	(0.131)	(0.129)	(0.177)	(0.172)	(0.155)
retd		-0.269**	0.289**	0.242**	0.132	-0.206†	0.004
		(0.098)	(0.099)	(0.088)	(0.123)	(0.114)	(0.092)
keephse		-0.152†	-0.068	0.166†	0.161	-0.170†	-0.125
		(0.091)	(0.089)	(0.087)	(0.131)	(0.101)	(0.086)
wales		-0.191†	0.096	0.028	-0.223	-0.295*	-0.235*
		(0.111)	(0.108)	(0.101)	(0.143)	(0.122)	(0.106)
north		-0.327**	0.129	-0.164†	-0.527**	0.025	-0.103
		(0.105)	(0.106)	(0.098)	(0.133)	(0.126)	(0.101)
nwest		-0.331**	0.210*	0.104	-0.244*	-0.055	-0.099
		(0.084)	(0.087)	(0.079)	(0.112)	(0.101)	(0.080)
yorks		-0.246**	-0.013	0.000	-0.355**	-0.065	-0.043
		(0.095)	(0.093)	(0.087)	(0.121)	(0.112)	(0.089)
wmids		-0.246*	-0.009	-0.053	-0.142	-0.113	-0.296**
		(0.098)	(0.095)	(0.090)	(0.128)	(0.114)	(0.095)
emids		-0.080	-0.082	-0.091	-0.282*	-0.184	-0.064
		(0.103)	(0.097)	(0.090)	(0.126)	(0.115)	(0.091)
anglia		-0.090	-0.005	-0.110	-0.108	-0.172	-0.078
		(0.129)	(0.126)	(0.116)	(0.168)	(0.149)	(0.119)
swest		-0.084	0.028	0.008	-0.296*	-0.178	-0.101
		(0.098)	(0.095)	(0.088)	(0.122)	(0.111)	(0.089)
london		-0.165†	0.044	-0.043	-0.218†	0.070	-0.022

continued on next page

Table B.1 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
		(0.096)	(0.094)	(0.086)	(0.122)	(0.118)	(0.089)
scot		-0.391**	0.190*	-0.189*	-0.310*	-0.262*	-0.007
		(0.093)	(0.095)	(0.086)	(0.123)	(0.105)	(0.088)
height		0.006	0.017†	-0.001	-0.003	0.005	-0.001
		(0.009)	(0.009)	(0.008)	(0.011)	(0.010)	(0.008)
housown		-0.105	-0.090	-0.476**	0.168	-0.061	0.120
		(0.137)	(0.133)	(0.127)	(0.161)	(0.154)	(0.136)
hou		0.058*	-0.034	-0.014	0.111**	-0.034	-0.067**
		(0.025)	(0.024)	(0.023)	(0.032)	(0.029)	(0.024)
smother		-0.704**	-0.335**	-0.045	-0.262**	-0.081	-0.065
		(0.051)	(0.051)	(0.049)	(0.065)	(0.060)	(0.050)
mothsmo		-0.433**	-0.141	0.229†	-0.123	-0.037	-0.077
		(0.145)	(0.140)	(0.135)	(0.183)	(0.168)	(0.140)
fathsmo		-0.196*	0.085	0.091	0.042	-0.047	-0.009
		(0.078)	(0.074)	(0.066)	(0.102)	(0.083)	(0.071)
bothsmo		-0.282**	-0.123	0.062	-0.061	0.089	-0.008
		(0.088)	(0.084)	(0.078)	(0.113)	(0.100)	(0.081)
alpa		-0.046*	-0.052*	-0.022	-0.171**	-0.014	0.018
		(0.021)	(0.021)	(0.019)	(0.027)	(0.023)	(0.020)
alma		-0.046†	-0.022	0.007	-0.111**	0.070*	0.029
		(0.025)	(0.025)	(0.023)	(0.031)	(0.030)	(0.024)
cons	0.953	1.039	-5.796†	-5.793*	5.356	11.948**	7.808*
	(3.381)	(3.325)	(3.093)	(2.688)	(3.747)	(3.419)	(3.290)
N	3670						
Log-likelihood	-12612.083						
$\chi^2_{(272)}$	3393.75						

Significance levels: † : 10% * : 5% ** : 1%

Table B.2: MVProbit Model for the full sample - reduced

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *exercise*
nsmoker	-0.720**		
	(0.175)		
breakfast	-0.144**		
	(0.054)		
sleepgd	-0.093*		
	(0.047)		
alqprud	-0.076		
	(0.072)		
nobese	-0.173**		
	(0.065)		
exercise	-0.961**		
	(0.184)		
sc12	-0.117†	0.132*	-0.019
	(0.063)	(0.060)	(0.057)
sc45	-0.065	-0.078	-0.152*
	(0.062)	(0.059)	(0.060)
lhqdg	0.020	0.117	-0.021
	(0.114)	(0.108)	(0.095)
lhqhndA	-0.095	-0.032	-0.017
	(0.112)	(0.102)	(0.093)
lhqnone	-0.015	-0.216*	-0.270**
	(0.095)	(0.085)	(0.079)
lhqoth	-0.106	-0.348**	-0.056
	(0.140)	(0.128)	(0.123)
wkshft1	-0.395**	-0.192†	-0.196*
	(0.118)	(0.100)	(0.099)
rural	-0.075	0.104	0.179**
	(0.071)	(0.069)	(0.067)
suburb	-0.016	0.018	0.164**
	(0.056)	(0.055)	(0.054)
ethwheur	0.346†	0.085	0.165
	(0.190)	(0.166)	(0.167)
male	0.413**	-0.226**	0.119†
	(0.053)	(0.073)	(0.069)
age	-0.152	0.036	-0.364*
	(0.168)	(0.172)	(0.169)
age2	0.327	-0.082	0.568*
	(0.277)	(0.290)	(0.286)
age3	-0.155	0.078	-0.315*
	(0.150)	(0.159)	(0.157)

continued on next page

Table B.2 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *exercise*
widow		-0.220**	-0.019
		(0.085)	(0.081)
divorce		-0.401**	0.224*
		(0.120)	(0.114)
seprd		-0.259	0.158
		(0.174)	(0.164)
single		-0.153	-0.262*
		(0.111)	(0.110)
part		-0.113	0.120
		(0.082)	(0.076)
unemp		-0.404**	-0.096
		(0.131)	(0.129)
sick		-0.323*	-0.821**
		(0.133)	(0.153)
retd		-0.248*	0.021
		(0.097)	(0.090)
keephse		-0.145	-0.118
		(0.091)	(0.085)
wales		-0.199†	-0.234*
		(0.110)	(0.105)
north		-0.343**	-0.117
		(0.104)	(0.100)
nwest		-0.334**	-0.097
		(0.084)	(0.079)
yorks		-0.253**	-0.049
		(0.095)	(0.089)
wmids		-0.253**	-0.298**
		(0.098)	(0.095)
emids		-0.086	-0.073
		(0.103)	(0.091)
anglia		-0.097	-0.084
		(0.129)	(0.118)
swest		-0.089	-0.104
		(0.099)	(0.088)
london		-0.170†	-0.033
		(0.096)	(0.088)
scot		-0.397**	-0.024
		(0.093)	(0.087)
height		0.006	-0.002
		(0.009)	(0.008)

continued on next page

Table B.2 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *exercise*
housown		-0.138	0.097
		(0.137)	(0.134)
hou		0.058*	-0.067**
		(0.025)	(0.024)
smother		-0.702**	-0.064
		(0.052)	(0.050)
mothsmo		-0.431**	-0.061
		(0.144)	(0.139)
fathsmo		-0.204**	-0.005
		(0.078)	(0.070)
bothsmo		-0.290**	-0.006
		(0.088)	(0.081)
alpa		-0.044*	0.015
		(0.021)	(0.020)
alma		-0.045†	0.028
		(0.025)	(0.024)
cons	1.142	0.398	7.744*
	(3.356)	(3.358)	(3.289)
N	3670		
Log-likelihood	-5689.6616		
$\chi^2_{(104)}$	2392.55		

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%

Table B.3: MVProbit Model for Men

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
nsmoker	-1.029**						
	(0.272)						
breakfast	-0.500						
	(0.378)						
sleepgd	-0.763*						
	(0.310)						
alqprud	0.353						
	(0.261)						
nobese	-0.031						
	(0.374)						
exercise	-0.605†						
	(0.315)						
sc12	0.045	0.063	0.064	0.013	-0.098	-0.036	-0.106
	(0.091)	(0.094)	(0.094)	(0.085)	(0.105)	(0.123)	(0.091)
sc45	0.008	-0.158†	0.020	-0.010	-0.162	-0.009	-0.059
	(0.088)	(0.089)	(0.089)	(0.083)	(0.099)	(0.115)	(0.090)
lhqdg	0.122	-0.003	-0.007	0.052	-0.123	0.382†	0.079
	(0.156)	(0.166)	(0.166)	(0.143)	(0.176)	(0.225)	(0.147)
lhqhndA	-0.090	-0.253	-0.389*	-0.027	-0.146	0.066	-0.014
	(0.155)	(0.156)	(0.156)	(0.138)	(0.169)	(0.200)	(0.141)
lhqnone	0.001	-0.336*	-0.438**	-0.022	-0.096	-0.028	-0.358**
	(0.134)	(0.132)	(0.134)	(0.116)	(0.144)	(0.165)	(0.120)
lhqoth	-0.224	-0.590**	-0.581**	0.078	-0.125	0.016	-0.108
	(0.179)	(0.174)	(0.176)	(0.161)	(0.193)	(0.228)	(0.166)
wkshft1	-0.383**	-0.110	-0.210†	-0.290*	-0.061	-0.307*	-0.167
	(0.137)	(0.126)	(0.122)	(0.117)	(0.134)	(0.155)	(0.125)
rural	0.045	0.155	0.170†	-0.004	0.191†	-0.148	0.131
	(0.096)	(0.102)	(0.101)	(0.094)	(0.114)	(0.132)	(0.100)
suburb	-0.025	0.042	0.226**	-0.069	0.066	-0.004	0.140†
	(0.079)	(0.081)	(0.081)	(0.074)	(0.089)	(0.109)	(0.080)
ethwheur	0.615*	0.388†	0.417†	0.287	-0.661†	0.069	0.062
	(0.251)	(0.236)	(0.230)	(0.223)	(0.360)	(0.328)	(0.237)
age	0.118	0.220	0.856**	0.260	0.085	-0.736*	-0.386
	(0.292)	(0.271)	(0.277)	(0.215)	(0.305)	(0.305)	(0.251)
age2	-0.137	-0.393	-1.411**	-0.431	-0.133	1.157*	0.586
	(0.490)	(0.458)	(0.470)	(0.359)	(0.519)	(0.503)	(0.424)
age3	0.106	0.243	0.762**	0.228	0.093	-0.584*	-0.315
	(0.270)	(0.251)	(0.258)	(0.193)	(0.287)	(0.268)	(0.232)
widow		-0.194	0.010	-0.171	-0.366*	-0.228	0.122
		(0.147)	(0.159)	(0.131)	(0.174)	(0.186)	(0.146)

continued on next page

Table B.3 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
divorce		-0.484*	-0.184	-0.022	-0.077	-0.317	0.088
		(0.204)	(0.204)	(0.183)	(0.225)	(0.252)	(0.200)
seprd		-0.277	-0.144	-0.245	0.198	-0.398	0.228
		(0.269)	(0.271)	(0.247)	(0.311)	(0.337)	(0.263)
single		-0.306*	-0.295†	-0.112	-0.254	-0.216	-0.038
		(0.149)	(0.154)	(0.144)	(0.167)	(0.197)	(0.151)
part		-0.330	0.271	0.190	-0.399	0.179	-0.128
		(0.227)	(0.238)	(0.208)	(0.246)	(0.371)	(0.232)
unemp		-0.517**	-0.438**	0.096	-0.046	-0.281	-0.087
		(0.149)	(0.148)	(0.145)	(0.164)	(0.184)	(0.154)
sick		-0.356*	-0.254	-0.270†	0.463*	0.198	-0.753**
		(0.152)	(0.157)	(0.147)	(0.191)	(0.224)	(0.185)
retd		-0.163	0.374*	0.146	0.129	-0.134	0.149
		(0.139)	(0.147)	(0.126)	(0.161)	(0.188)	(0.137)
wales		-0.196	-0.013	0.154	-0.386*	-0.409*	-0.274†
		(0.161)	(0.157)	(0.145)	(0.180)	(0.209)	(0.164)
north		-0.415**	0.256	-0.122	-0.726**	-0.187	-0.145
		(0.161)	(0.174)	(0.150)	(0.179)	(0.224)	(0.161)
nwest		-0.383**	0.110	0.087	-0.300*	-0.193	-0.070
		(0.123)	(0.127)	(0.113)	(0.145)	(0.173)	(0.119)
yorks		-0.396**	0.075	0.096	-0.496**	-0.148	0.081
		(0.142)	(0.144)	(0.130)	(0.158)	(0.202)	(0.138)
wmids		-0.282*	0.170	0.065	-0.079	0.032	-0.201
		(0.143)	(0.143)	(0.133)	(0.171)	(0.217)	(0.140)
emids		-0.207	-0.098	-0.058	-0.306†	-0.340†	-0.068
		(0.146)	(0.140)	(0.128)	(0.163)	(0.189)	(0.137)
anglia		-0.336†	0.133	-0.056	-0.115	-0.103	-0.246
		(0.181)	(0.183)	(0.162)	(0.220)	(0.260)	(0.181)
swest		-0.383**	-0.055	0.130	-0.411**	-0.297	-0.184
		(0.135)	(0.139)	(0.125)	(0.158)	(0.185)	(0.130)
london		-0.169	0.294*	-0.047	-0.142	-0.039	0.016
		(0.137)	(0.139)	(0.120)	(0.159)	(0.199)	(0.127)
scot		-0.413**	0.149	-0.144	-0.357*	-0.406*	0.095
		(0.137)	(0.143)	(0.126)	(0.163)	(0.180)	(0.131)
height		0.011	0.006	-0.002	0.015	-0.004	-0.010
		(0.012)	(0.013)	(0.011)	(0.014)	(0.017)	(0.012)
housown		-0.198	-0.095	-0.429*	0.051	-0.054	0.359†
		(0.188)	(0.188)	(0.184)	(0.200)	(0.231)	(0.191)
hou		0.002	-0.043	0.001	0.074†	0.001	-0.042
		(0.036)	(0.035)	(0.032)	(0.041)	(0.048)	(0.036)

continued on next page

Table B.3 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
smother		-0.699**	-0.310**	-0.002	-0.310**	-0.070	-0.109
		(0.079)	(0.080)	(0.074)	(0.087)	(0.106)	(0.079)
mothsmo		-0.474*	-0.425*	0.273	-0.419†	0.189	-0.098
		(0.215)	(0.214)	(0.198)	(0.240)	(0.342)	(0.225)
fathsmo		-0.071	-0.124	0.152	-0.037	-0.120	0.168
		(0.115)	(0.118)	(0.099)	(0.138)	(0.154)	(0.112)
bothsmo		-0.196	-0.309*	0.145	-0.160	-0.055	0.008
		(0.131)	(0.133)	(0.116)	(0.152)	(0.179)	(0.129)
alpa		-0.067*	-0.029	-0.018	-0.194**	0.030	0.008
		(0.030)	(0.031)	(0.027)	(0.035)	(0.040)	(0.029)
alma		-0.016	-0.074*	0.007	-0.090*	0.018	0.030
		(0.037)	(0.037)	(0.034)	(0.041)	(0.051)	(0.037)
cons	-3.676	-3.622	-16.583**	-4.613	-0.910	16.940**	8.784†
	(5.702)	(5.260)	(5.362)	(4.246)	(5.883)	(6.088)	(4.891)
N	1671						
Log-likelihood	-5770.1051						
$\chi^2_{(259)}$	1815.44						

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%

Table B.4: MVProbit Model for Men - reduced

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *sleepgd*	**Eq. 4** *exercise*
nsmoker	-0.972**			
	(0.218)			
breakfast	-0.094			
	(0.073)			
sleepgd	-0.480			
	(0.483)			
alqprud	-0.047			
	(0.081)			
nobese	-0.257*			
	(0.109)			
exercise	-0.854**			
	(0.293)			
sc12	0.008	0.055	-0.110	0.013
	(0.094)	(0.095)	(0.090)	(0.085)
sc45	-0.008	-0.153†	-0.058	-0.006
	(0.090)	(0.088)	(0.090)	(0.083)
lhqdg	0.137	0.027	0.075	0.070
	(0.160)	(0.168)	(0.145)	(0.144)
lhqhndA	-0.051	-0.232	-0.019	-0.022
	(0.157)	(0.157)	(0.140)	(0.138)
lhqnone	0.047	-0.330*	-0.360**	-0.013
	(0.138)	(0.132)	(0.119)	(0.117)
lhqoth	-0.169	-0.590**	-0.129	0.083
	(0.185)	(0.175)	(0.165)	(0.162)
wkshft1	-0.386**	-0.124	-0.163	-0.284*
	(0.139)	(0.126)	(0.125)	(0.118)
rural	0.039	0.152	0.123	-0.013
	(0.099)	(0.103)	(0.099)	(0.094)
suburb	-0.031	0.039	0.131†	-0.072
	(0.083)	(0.081)	(0.080)	(0.075)
ethwheur	0.490†	0.370	0.048	0.296
	(0.262)	(0.237)	(0.236)	(0.223)
age	-0.031	0.267	-0.303	0.268
	(0.294)	(0.273)	(0.249)	(0.218)
age2	0.111	-0.472	0.449	-0.447
	(0.496)	(0.462)	(0.419)	(0.363)
age3	-0.026	0.285	-0.244	0.237
	(0.274)	(0.253)	(0.228)	(0.196)
widow		-0.177	0.147	-0.181
		(0.149)	(0.143)	(0.137)

continued on next page

Table B.4 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *sleepgd*	**Eq. 4** *exercise*
divorce		-0.499*	0.102	-0.048
		(0.203)	(0.195)	(0.190)
seprd		-0.330	0.168	-0.263
		(0.269)	(0.258)	(0.253)
single		-0.307*	-0.032	-0.112
		(0.152)	(0.147)	(0.147)
part		-0.300	-0.049	0.197
		(0.229)	(0.227)	(0.216)
unemp		-0.525**	-0.111	0.087
		(0.151)	(0.151)	(0.151)
sick		-0.356*	-0.769**	-0.256
		(0.155)	(0.181)	(0.158)
retd		-0.139	0.194	0.158
		(0.141)	(0.133)	(0.130)
wales		-0.181	-0.256	0.161
		(0.163)	(0.163)	(0.150)
north		-0.396*	-0.142	-0.094
		(0.162)	(0.159)	(0.153)
nwest		-0.356**	-0.073	0.108
		(0.124)	(0.116)	(0.116)
yorks		-0.385**	0.093	0.104
		(0.143)	(0.134)	(0.134)
wmids		-0.267†	-0.195	0.092
		(0.145)	(0.138)	(0.135)
emids		-0.195	-0.073	-0.050
		(0.148)	(0.134)	(0.131)
anglia		-0.321†	-0.243	-0.055
		(0.183)	(0.176)	(0.168)
swest		-0.355**	-0.168	0.161
		(0.136)	(0.127)	(0.127)
london		-0.151	0.009	-0.021
		(0.139)	(0.123)	(0.124)
scot		-0.383**	0.077	-0.110
		(0.138)	(0.128)	(0.130)
height		0.012	-0.012	-0.000
		(0.013)	(0.012)	(0.012)
housown		-0.260	0.348†	-0.474**
		(0.191)	(0.188)	(0.183)
hou		-0.004	-0.042	-0.005
		(0.036)	(0.035)	(0.034)

continued on next page

Table B.4 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *sleepgd*	**Eq. 4** *exercise*
smother		-0.715**	-0.112	-0.034
		(0.077)	(0.076)	(0.075)
mothsmo		-0.460*	-0.060	0.260
		(0.216)	(0.221)	(0.205)
fathsmo		-0.076	0.166	0.155
		(0.116)	(0.111)	(0.103)
bothsmo		-0.203	0.010	0.142
		(0.132)	(0.127)	(0.120)
alpa		-0.063*	0.008	-0.014
		(0.030)	(0.029)	(0.028)
alma		-0.015	0.031	0.002
		(0.038)	(0.036)	(0.035)
cons	-0.594	-4.468	7.339	-4.809
	(5.748)	(5.300)	(4.845)	(4.295)
N	1671			
Log-likelihood	-3725.5601			
$\chi^2_{(139)}$	1331.10			

Significance levels: † : 10% * : 5% ** : 1%

Table B.5: MVProbit Model for Women

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
nsmoker	-0.393†						
	(0.238)						
breakfast	-0.010						
	(0.288)						
sleepgd	-0.877**						
	(0.305)						
alqprud	0.297						
	(0.347)						
nobese	-0.913**						
	(0.260)						
exercise	-0.167						
	(0.310)						
sc12	-0.155†	0.206**	0.027	0.040	-0.429**	0.281**	0.023
	(0.090)	(0.080)	(0.078)	(0.073)	(0.116)	(0.084)	(0.075)
sc45	-0.045	0.001	-0.104	0.059	-0.135	0.040	-0.216**
	(0.085)	(0.081)	(0.080)	(0.075)	(0.135)	(0.083)	(0.084)
lhqdg	-0.097	0.193	0.222	-0.004	-0.034	0.030	-0.091
	(0.161)	(0.145)	(0.142)	(0.131)	(0.190)	(0.153)	(0.130)
lhqhndA	-0.176	0.089	0.178	-0.068	-0.124	-0.072	-0.022
	(0.154)	(0.137)	(0.136)	(0.127)	(0.185)	(0.144)	(0.126)
lhqnone	-0.050	-0.139	-0.182	-0.114	0.169	-0.006	-0.238*
	(0.128)	(0.112)	(0.111)	(0.106)	(0.161)	(0.120)	(0.106)
lhqoth	0.005	0.015	0.281	0.069	0.001	0.149	0.089
	(0.224)	(0.210)	(0.222)	(0.198)	(0.298)	(0.235)	(0.198)
wkshft1	-0.400†	-0.282	-0.054	-0.309†	0.245	-0.116	-0.239
	(0.235)	(0.173)	(0.170)	(0.167)	(0.291)	(0.190)	(0.170)
rural	-0.185†	0.089	0.069	0.113	0.177	0.216*	0.272**
	(0.104)	(0.096)	(0.094)	(0.087)	(0.146)	(0.097)	(0.093)
suburb	0.002	0.028	0.109	-0.022	0.138	0.284**	0.212**
	(0.081)	(0.075)	(0.075)	(0.070)	(0.115)	(0.078)	(0.076)
ethwheur	0.333	-0.238	0.555**	0.440*	-0.142	0.302	0.241
	(0.273)	(0.249)	(0.212)	(0.216)	(0.445)	(0.225)	(0.239)
age	0.040	-0.276	-0.092	0.482**	-0.202	-0.446*	-0.537*
	(0.220)	(0.224)	(0.199)	(0.177)	(0.279)	(0.211)	(0.255)
age2	-0.002	0.473	0.178	-0.852**	0.399	0.679†	0.894*
	(0.361)	(0.377)	(0.332)	(0.293)	(0.461)	(0.350)	(0.436)
age3	0.020	-0.229	-0.096	0.465**	-0.239	-0.327†	-0.509*
	(0.192)	(0.205)	(0.178)	(0.157)	(0.245)	(0.187)	(0.242)
widow		-0.180†	-0.236*	-0.155†	0.314†	-0.291**	-0.022
		(0.108)	(0.107)	(0.092)	(0.176)	(0.103)	(0.107)

continued on next page

Table B.5 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
divorce		-0.312*	-0.307*	-0.402**	0.276	-0.018	0.338*
		(0.153)	(0.150)	(0.142)	(0.251)	(0.167)	(0.149)
seprd		-0.235	-0.390†	-0.327	-0.414	0.059	0.182
		(0.235)	(0.231)	(0.213)	(0.290)	(0.264)	(0.225)
single		0.008	0.161	-0.223	0.521†	-0.017	-0.515**
		(0.168)	(0.172)	(0.141)	(0.280)	(0.165)	(0.182)
part		-0.055	0.135	0.224*	0.246†	-0.040	0.101
		(0.100)	(0.099)	(0.094)	(0.146)	(0.110)	(0.094)
unemp		0.072	-0.092	0.409	-0.378	0.309	-0.086
		(0.298)	(0.276)	(0.287)	(0.350)	(0.379)	(0.282)
sick		-0.079	-0.178	-0.574*	4.615	-0.312	-0.672*
		(0.271)	(0.250)	(0.230)	(275.741)	(0.266)	(0.317)
retd		-0.431**	0.215	0.246*	0.024	-0.291*	-0.214
		(0.143)	(0.140)	(0.125)	(0.210)	(0.145)	(0.137)
keephse		-0.128	-0.039	0.049	0.156	-0.257*	-0.166
		(0.107)	(0.104)	(0.099)	(0.156)	(0.112)	(0.102)
wales		-0.139	0.181	-0.106	0.103	-0.316*	-0.160
		(0.153)	(0.150)	(0.140)	(0.256)	(0.148)	(0.149)
north		-0.262†	0.072	-0.275*	-0.238	0.041	-0.037
		(0.139)	(0.139)	(0.128)	(0.221)	(0.149)	(0.138)
nwest		-0.282*	0.302*	0.053	-0.025	-0.024	-0.065
		(0.118)	(0.122)	(0.110)	(0.188)	(0.123)	(0.114)
yorks		-0.102	-0.076	-0.113	-0.138	-0.062	-0.109
		(0.131)	(0.125)	(0.115)	(0.193)	(0.134)	(0.125)
wmids		-0.210	-0.151	-0.192	-0.211	-0.202	-0.372**
		(0.135)	(0.130)	(0.121)	(0.201)	(0.136)	(0.139)
emids		0.050	-0.072	-0.149	-0.297	-0.117	-0.079
		(0.145)	(0.136)	(0.124)	(0.198)	(0.143)	(0.132)
anglia		0.147	-0.154	-0.124	-0.095	-0.157	0.016
		(0.189)	(0.175)	(0.161)	(0.273)	(0.180)	(0.170)
swest		0.229	0.113	-0.164	-0.031	-0.165	-0.016
		(0.147)	(0.134)	(0.121)	(0.203)	(0.138)	(0.128)
london		-0.221	-0.212	-0.094	-0.267	0.061	-0.045
		(0.138)	(0.133)	(0.121)	(0.197)	(0.144)	(0.135)
scot		-0.414**	0.152	-0.272*	-0.181	-0.168	-0.053
		(0.129)	(0.130)	(0.115)	(0.204)	(0.131)	(0.125)
height		-0.000	0.026*	0.002	-0.025	0.004	0.013
		(0.013)	(0.013)	(0.011)	(0.019)	(0.013)	(0.012)
housown		0.027	-0.095	-0.447*	0.218	-0.106	-0.105
		(0.202)	(0.200)	(0.181)	(0.286)	(0.207)	(0.204)

continued on next page

Table B.5 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *nsmoker*	**Eq. 3** *breakfast*	**Eq. 4** *sleepgd*	**Eq. 5** *alqprud*	**Eq. 6** *nobese*	**Eq. 7** *exercise*
hou		0.139**	-0.023	-0.011	0.177**	-0.059	-0.076*
		(0.037)	(0.035)	(0.033)	(0.059)	(0.037)	(0.035)
smother		-0.707**	-0.314**	-0.060	-0.234*	-0.092	-0.024
		(0.071)	(0.069)	(0.066)	(0.107)	(0.074)	(0.069)
mothsmo		-0.383†	0.085	0.236	0.249	-0.148	-0.039
		(0.202)	(0.192)	(0.181)	(0.320)	(0.195)	(0.189)
fathsmo		-0.312**	0.206*	0.024	0.110	-0.044	-0.163†
		(0.107)	(0.097)	(0.087)	(0.156)	(0.098)	(0.097)
bothsmo		-0.358**	-0.035	-0.014	0.038	0.119	-0.019
		(0.120)	(0.112)	(0.103)	(0.173)	(0.120)	(0.111)
alpa		-0.032	-0.073**	-0.032	-0.147**	-0.032	0.027
		(0.029)	(0.028)	(0.026)	(0.046)	(0.029)	(0.028)
alma		-0.070*	0.024	0.001	-0.166**	0.091*	0.026
		(0.035)	(0.034)	(0.032)	(0.049)	(0.037)	(0.034)
cons	-2.142	6.132	-0.165	-8.080*	6.297	9.860*	9.867*
	(4.428)	(4.407)	(3.981)	(3.580)	(5.678)	(4.236)	(4.916)
N	1999						
Log-likelihood	-6652.1237						
$\chi^2_{(265)}$	1843.98						

Significance levels: † : 10% * : 5% ** : 1%

Table B.6: MVProbit Model for Women - reduced

Variable	**Eq. 1** *deaths*	**Eq. 2** *sleepgd*	**Eq. 3** *nobese*
nsmoker	-0.265**		
	(0.076)		
breakfast	-0.192*		
	(0.077)		
sleepgd	-0.487		
	(0.360)		
alqprud	-0.144		
	(0.140)		
nobese	-0.956**		
	(0.315)		
exercise	-0.257**		
	(0.078)		
sc12	-0.210*	0.044	0.272**
	(0.090)	(0.072)	(0.084)
sc45	-0.069	0.060	0.040
	(0.084)	(0.076)	(0.084)
lhqdg	-0.100	-0.007	0.039
	(0.164)	(0.131)	(0.154)
lhqhndA	-0.187	-0.056	-0.088
	(0.158)	(0.127)	(0.144)
lhqnone	-0.038	-0.117	-0.011
	(0.128)	(0.106)	(0.120)
lhqoth	0.011	0.054	0.137
	(0.232)	(0.198)	(0.238)
wkshft1	-0.339	-0.294†	-0.126
	(0.245)	(0.167)	(0.190)
rural	-0.205*	0.110	0.204*
	(0.103)	(0.088)	(0.097)
suburb	0.015	-0.025	0.277**
	(0.081)	(0.070)	(0.078)
ethwheur	0.339	0.435*	0.314
	(0.269)	(0.215)	(0.225)
age	-0.036	0.464**	-0.445*
	(0.225)	(0.179)	(0.211)
age2	0.140	-0.822**	0.678†
	(0.368)	(0.295)	(0.349)
age3	-0.058	0.449**	-0.326†
	(0.196)	(0.157)	(0.186)
widow		-0.146	-0.298**
		(0.096)	(0.103)

continued on next page

Table B.6 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *sleepgd*	**Eq. 3** *nobese*
divorce		-0.402**	-0.009
		(0.145)	(0.168)
seprd		-0.332	0.037
		(0.219)	(0.261)
single		-0.224	-0.051
		(0.144)	(0.165)
part		0.242*	-0.046
		(0.096)	(0.111)
unemp		0.449	0.313
		(0.291)	(0.385)
sick		-0.510*	-0.384
		(0.251)	(0.269)
retd		0.253*	-0.313*
		(0.128)	(0.145)
keephse		0.076	-0.261*
		(0.102)	(0.113)
wales		-0.066	-0.334*
		(0.143)	(0.149)
north		-0.240†	0.016
		(0.133)	(0.150)
nwest		0.084	-0.040
		(0.111)	(0.124)
yorks		-0.089	-0.071
		(0.118)	(0.134)
wmids		-0.174	-0.216
		(0.124)	(0.136)
emids		-0.138	-0.126
		(0.127)	(0.143)
anglia		-0.146	-0.149
		(0.165)	(0.182)
swest		-0.145	-0.181
		(0.124)	(0.139)
london		-0.074	0.074
		(0.125)	(0.146)
scot		-0.266*	-0.185
		(0.119)	(0.131)
height		0.004	0.006
		(0.011)	(0.013)
housown		-0.475**	-0.130
		(0.184)	(0.207)

continued on next page

Table B.6 – *continued from previous page*

Variable	**Eq. 1** *deaths*	**Eq. 2** *sleepgd*	**Eq. 3** *nobese*
hou		-0.009	-0.057
		(0.034)	(0.037)
smother		-0.060	-0.101
		(0.066)	(0.073)
mothsmo		0.250	-0.143
		(0.184)	(0.195)
fathsmo		0.041	-0.048
		(0.089)	(0.098)
bothsmo		0.007	0.122
		(0.106)	(0.119)
alpa		-0.033	-0.031
		(0.026)	(0.029)
alma		0.007	0.094*
		(0.032)	(0.037)
cons	-0.683	-7.895*	9.753*
	(4.501)	(3.602)	(4.223)
N	1999		
Log-likelihood	-3063.8894		
$\chi^2_{(101)}$	940.03		

Significance levels: † : 10% ∗ : 5% ∗∗ : 1%

Table B.7: Decomposition of the Gini: Excluded Deaths Eq. - Full Sample

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
sc12	-0.193	0.316	2.006	-0.030	-0.202	0.006	0.022	2.19
sc45	0.034	0.219	2.006	0.004	0.151	0.001	0.002	0.20
lhqdg	-0.028	0.126	2.006	-0.002	-0.256	0.000	0.002	0.16
lhqhndA	-0.100	0.125	2.006	-0.006	-0.290	0.002	0.006	0.65
lhqnone	0.174	0.608	2.006	0.053	0.160	0.008	0.030	3.01
lhqoth	0.013	0.047	2.006	0.000	-0.007	0.000	0.000	0.00
wkshft1	-0.325	0.057	2.006	-0.009	-0.471	0.004	0.016	1.56
rural	-0.221	0.219	2.006	-0.024	-0.112	0.003	0.010	0.96
suburb	-0.098	0.472	2.006	-0.023	-0.019	0.000	0.002	0.15
ethwheur	0.292	0.979	2.006	0.143	0.009	0.001	0.005	0.45
male	0.445	0.455	2.006	0.101	0.155	0.016	0.056	5.57
age	-0.010	57.468	2.006	-0.285	0.110	-0.032	-0.112	-11.24
age2	0.121	34.388	2.006	2.080	0.218	0.454	1.620	161.96
age3	-0.054	21.396	2.006	-0.576	0.320	-0.184	-0.656	-65.61
costant	-3.340					0.280		
$\sum_k \eta_k C_k$						0.280		
Total Gini						0.280		

Table B.8: Decomposition of the Gini: Included Deaths Eq. - Full Sample

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.390	0.700	2.276	-0.120	-0.036	0.004	0.016	1.65
breakfast	-0.174	0.707	2.276	-0.054	0.011	-0.001	-0.002	-0.22
sleepgd	-0.106	0.583	2.276	-0.027	-0.075	0.002	0.008	0.79
alqprud	-0.083	0.880	2.276	-0.032	0.002	0.000	0.000	-0.03
nobese	-0.200	0.854	2.276	-0.075	-0.011	0.001	0.003	0.32
exercise	-0.132	0.323	2.276	-0.019	-0.258	0.005	0.018	1.85
sc12	-0.159	0.316	2.276	-0.022	-0.197	0.004	0.017	1.66
sc45	0.003	0.219	2.276	0.000	0.148	0.000	0.000	0.02
lhqdg	0.000	0.126	2.276	0.000	-0.249	0.000	0.000	0.00
lhqhndA	-0.101	0.125	2.276	-0.006	-0.280	0.002	0.006	0.59
lhqnone	0.113	0.608	2.276	0.030	0.155	0.005	0.018	1.78
lhqoth	-0.045	0.047	2.276	-0.001	-0.006	0.000	0.000	0.00
wkshft1	-0.384	0.057	2.276	-0.010	-0.452	0.004	0.017	1.67
rural	-0.172	0.219	2.276	-0.017	-0.108	0.002	0.007	0.69
suburb	-0.075	0.472	2.276	-0.015	-0.018	0.000	0.001	0.11
ethwheur	0.341	0.979	2.276	0.147	0.009	0.001	0.005	0.48
male	0.457	0.455	2.276	0.091	0.149	0.014	0.052	5.21
age	-0.063	57.468	2.276	-1.582	0.108	-0.170	-0.651	-65.13
age2	0.213	34.388	2.276	3.219	0.213	0.686	2.627	262.66
age3	-0.101	21.396	2.276	-0.953	0.313	-0.298	-1.141	-114.09
costant	-1.768							
$\sum_k \eta_k C_k$						0.261		
Total Gini						0.261		

Table B.9: Decomposition of the Gini: Endogenous Deaths Eq. - Full Sample

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.650	0.700	2.690	-0.169	-0.084	0.014	0.061	6.15
breakfast	-0.178	0.707	2.690	-0.047	-0.013	0.001	0.003	0.25
sleepgd	-0.544	0.583	2.690	-0.118	-0.158	0.019	0.080	8.02
alqprud	-0.264	0.880	2.690	-0.086	-0.009	0.001	0.003	0.35
nobese	-0.129	0.854	2.690	-0.041	-0.007	0.000	0.001	0.13
exercise	-0.900	0.323	2.690	-0.108	-0.478	0.052	0.223	22.27
sc12	-0.120	0.316	2.690	-0.014	-0.164	0.002	0.010	0.99
sc45	-0.060	0.219	2.690	-0.005	0.123	-0.001	-0.003	-0.26
lhqdg	0.022	0.126	2.690	0.001	-0.201	0.000	-0.001	-0.09
lhqhndA	-0.105	0.125	2.690	-0.005	-0.238	0.001	0.005	0.50
lhqnone	-0.025	0.608	2.690	-0.006	0.129	-0.001	-0.003	-0.31
lhqoth	-0.097	0.047	2.690	-0.002	0.006	0.000	0.000	0.00
wkshft1	-0.446	0.057	2.690	-0.009	-0.375	0.004	0.015	1.53
rural	-0.059	0.219	2.690	-0.005	-0.090	0.000	0.002	0.18
suburb	-0.020	0.472	2.690	-0.004	-0.016	0.000	0.000	0.02
ethwheur	0.382	0.979	2.690	0.139	0.007	0.001	0.004	0.43
male	0.380	0.455	2.690	0.064	0.125	0.008	0.035	3.46
age	-0.124	57.468	2.690	-2.640	0.092	-0.244	-1.049	-104.86
age2	0.280	34.388	2.690	3.579	0.184	0.659	2.838	283.80
age3	-0.131	21.396	2.690	-1.045	0.272	-0.285	-1.226	-122.56
costant	0.953							
$\sum_k \eta_k C_k$						0.232		
Total Gini						0.236		

Table B.10: Decomposition of the Gini: Endogenous Deaths Eq. (reduced model) - Full Sample

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.720	0.700	2.525	-0.199	-0.090	0.018	0.073	7.27
breakfast	-0.144	0.707	2.525	-0.040	-0.003	0.000	0.000	0.04
sleepgd	-0.093	0.583	2.525	-0.022	-0.064	0.001	0.006	0.56
alqprud	-0.076	0.880	2.525	-0.027	0.002	0.000	0.000	-0.03
nobese	-0.173	0.854	2.525	-0.059	-0.010	0.001	0.002	0.23
exercise	-0.961	0.323	2.525	-0.123	-0.511	0.063	0.255	25.54
sc12	-0.117	0.316	2.525	-0.015	-0.169	0.002	0.010	1.00
sc45	-0.065	0.219	2.525	-0.006	0.126	-0.001	-0.003	-0.29
lhqdg	0.020	0.126	2.525	0.001	-0.210	0.000	-0.001	-0.08
lhqhndA	-0.095	0.125	2.525	-0.005	-0.243	0.001	0.005	0.46
lhqnone	-0.015	0.608	2.525	-0.004	0.133	0.000	-0.002	-0.19
lhqoth	-0.106	0.047	2.525	-0.002	0.002	0.000	0.000	0.00
wkshft1	-0.395	0.057	2.525	-0.009	-0.385	0.003	0.014	1.40
rural	-0.075	0.219	2.525	-0.007	-0.091	0.001	0.002	0.24
suburb	-0.016	0.472	2.525	-0.003	-0.016	0.000	0.000	0.02
ethwheur	0.346	0.979	2.525	0.134	0.007	0.001	0.004	0.40
male	0.413	0.455	2.525	0.074	0.126	0.009	0.038	3.82
age	-0.152	57.468	2.525	-3.466	0.094	-0.326	-1.327	-132.68
age2	0.327	34.388	2.525	4.457	0.188	0.837	3.405	340.46
age3	-0.155	21.396	2.525	-1.312	0.278	-0.364	-1.482	-148.17
costant	1.142							
$\sum_k \eta_k C_k$						0.246		
Total Gini						0.246		

Table B.11: Decomposition of the Gini: Excluded Deaths Eq. - Men

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
sc12	-0.026	0.303	2.033	-0.004	-0.143	0.001	0.002	0.19
sc45	0.126	0.216	2.033	0.013	0.166	0.002	0.008	0.76
lhqdg	0.030	0.136	2.033	0.002	-0.221	0.000	-0.002	-0.15
lhqhndA	-0.004	0.123	2.033	0.000	-0.284	0.000	0.000	0.02
lhqnone	0.330	0.583	2.033	0.094	0.173	0.016	0.056	5.61
lhqoth	0.097	0.069	2.033	0.003	-0.077	0.000	-0.001	-0.09
wkshft1	-0.323	0.086	2.033	-0.014	-0.520	0.007	0.024	2.43
rural	-0.096	0.218	2.033	-0.010	-0.027	0.000	0.001	0.09
suburb	-0.097	0.466	2.033	-0.022	-0.037	0.001	0.003	0.28
ethwheur	0.406	0.978	2.033	0.195	0.011	0.002	0.008	0.76
age	0.162	57.890	2.033	4.616	0.113	0.520	1.781	178.13
age2	-0.174	34.863	2.033	-2.991	0.221	-0.662	-2.268	-226.76
age3	0.117	21.799	2.033	1.257	0.322	0.405	1.387	138.71
costant	-6.600							
$\sum_k \eta_k C_k$						0.292		
Total Gini						0.292		

Table B.12: Decomposition of the Gini: Included Deaths Eq. - Men

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.505	0.683	2.286	-0.151	-0.043	0.007	0.024	2.37
breakfast	-0.156	0.698	2.286	-0.048	0.028	-0.001	-0.005	-0.49
sleepgd	-0.008	0.577	2.286	-0.002	-0.001	0.000	0.000	0.00
alqprud	-0.077	0.808	2.286	-0.027	0.027	-0.001	-0.003	-0.26
nobese	-0.314	0.912	2.286	-0.125	-0.011	0.001	0.005	0.52
exercise	0.016	0.330	2.286	0.002	-0.203	0.000	-0.002	-0.17
sc12	0.007	0.303	2.286	0.001	-0.136	0.000	0.000	-0.04
sc45	0.086	0.216	2.286	0.008	0.162	0.001	0.005	0.48
lhqdg	0.062	0.136	2.286	0.004	-0.213	-0.001	-0.003	-0.29
lhqhndA	-0.030	0.123	2.286	-0.002	-0.274	0.000	0.002	0.16
lhqnone	0.260	0.583	2.286	0.066	0.167	0.011	0.040	4.01
lhqoth	-0.031	0.069	2.286	-0.001	-0.070	0.000	0.000	0.02
wkshft1	-0.368	0.086	2.286	-0.014	-0.499	0.007	0.025	2.50
rural	-0.051	0.218	2.286	-0.005	-0.027	0.000	0.000	0.05
suburb	-0.083	0.466	2.286	-0.017	-0.036	0.001	0.002	0.22
ethwheur	0.460	0.978	2.286	0.197	0.011	0.002	0.008	0.77
age	0.119	57.890	2.286	3.023	0.109	0.331	1.200	119.99
age2	-0.091	34.863	2.286	-1.384	0.215	-0.298	-1.082	-108.18
age3	0.072	21.799	2.286	0.687	0.314	0.216	0.783	78.34
costant	-5.290							
$\sum_k \eta_k C_k$						0.276		
Total Gini						0.276		

Table B.13: Decomposition of the Gini: Endogenous Deaths Eq. - Men

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-1.029	0.683	2.696	-0.261	-0.136	0.035	0.148	14.78
breakfast	-0.500	0.698	2.696	-0.130	-0.052	0.007	0.028	2.78
sleepgd	-0.763	0.577	2.696	-0.163	-0.163	0.027	0.111	11.07
alqprud	0.353	0.808	2.696	0.106	0.042	0.004	0.019	1.86
nobese	-0.031	0.912	2.696	-0.011	0.001	0.000	0.000	-0.01
exercise	-0.605	0.330	2.696	-0.074	-0.387	0.029	0.120	11.95
sc12	0.045	0.303	2.696	0.005	-0.097	0.000	-0.002	-0.21
sc45	0.008	0.216	2.696	0.001	0.114	0.000	0.000	0.03
lhqdg	0.122	0.136	2.696	0.006	-0.162	-0.001	-0.004	-0.41
lhqhndA	-0.090	0.123	2.696	-0.004	-0.200	0.001	0.003	0.34
lhqnone	0.001	0.583	2.696	0.000	0.118	0.000	0.000	0.02
lhqoth	-0.224	0.069	2.696	-0.006	-0.070	0.000	0.002	0.17
wkshft1	-0.383	0.086	2.696	-0.012	-0.562	0.007	0.029	2.85
rural	0.045	0.218	2.696	0.004	-0.018	0.000	0.000	-0.03
suburb	-0.025	0.466	2.696	-0.004	-0.025	0.000	0.000	0.05
ethwheur	0.615	0.978	2.696	0.223	0.008	0.002	0.008	0.75
age	0.118	57.890	2.696	2.535	0.085	0.215	0.897	89.67
age2	-0.137	34.863	2.696	-1.772	0.170	-0.301	-1.254	-125.38
age3	0.106	21.799	2.696	0.856	0.252	0.215	0.897	89.71
costant	-3.676					0.240		
$\sum_k \eta_k C_k$						0.240		
Total Gini						0.237		

Table B.14: Decomposition of the Gini: Endogenous Deaths Eq. (reduced model) - Men

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.972	0.683	2.660	-0.250	-0.124	0.031	0.129	12.89
breakfast	-0.094	0.698	2.660	-0.025	0.002	0.000	0.000	-0.03
sleepgd	-0.480	0.577	2.660	-0.104	-0.097	0.010	0.042	4.21
alqprud	-0.047	0.808	2.660	-0.014	0.020	0.000	-0.001	-0.12
nobese	-0.257	0.912	2.660	-0.088	-0.011	0.001	0.004	0.39
exercise	-0.854	0.330	2.660	-0.106	-0.465	0.049	0.205	20.50
sc12	0.008	0.303	2.660	0.001	-0.107	0.000	0.000	-0.04
sc45	-0.008	0.216	2.660	-0.001	0.128	0.000	0.000	-0.03
lhqdg	0.137	0.136	2.660	0.007	-0.179	-0.001	-0.005	-0.52
lhqhndA	-0.051	0.123	2.660	-0.002	-0.220	0.001	0.002	0.22
lhqnone	0.047	0.583	2.660	0.010	0.129	0.001	0.006	0.56
lhqoth	-0.169	0.069	2.660	-0.004	-0.082	0.000	0.001	0.15
wkshft1	-0.386	0.086	2.660	-0.012	-0.604	0.008	0.031	3.13
rural	0.039	0.218	2.660	0.003	-0.022	0.000	0.000	-0.03
suburb	-0.031	0.466	2.660	-0.005	-0.027	0.000	0.001	0.06
ethwheur	0.490	0.978	2.660	0.180	-0.027	-0.005	-0.020	-1.99
age	-0.031	57.890	2.660	-0.680	0.091	-0.062	-0.258	-25.84
age2	0.111	34.863	2.660	1.454	0.182	0.265	1.104	110.42
age3	-0.026	21.799	2.660	-0.213	0.270	-0.057	-0.239	-23.92
costant	-0.594					0.240		
$\sum_k \eta_k C_k$						0.240		
Total Gini						0.242		

Table B.15: Decomposition of the Gini: Excluded Deaths Eq. - Women

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
sc12	-0.318	0.326	1.716	-0.060	-0.262	0.016	0.052	5.22
sc45	-0.053	0.221	1.716	-0.007	0.148	-0.001	-0.003	-0.33
lhqdg	-0.141	0.117	1.716	-0.010	-0.344	0.003	0.011	1.10
lhqhndA	-0.197	0.126	1.716	-0.014	-0.310	0.004	0.015	1.48
lhqnone	0.036	0.629	1.716	0.013	0.170	0.002	0.007	0.74
lhqoth	-0.044	0.030	1.716	-0.001	-0.062	0.000	0.000	0.02
wkshft1	-0.211	0.034	1.716	-0.004	-0.596	0.002	0.008	0.81
rural	-0.341	0.220	1.716	-0.044	-0.205	0.009	0.030	2.96
suburb	-0.098	0.477	1.716	-0.027	0.003	0.000	0.000	-0.02
ethwheur	0.160	0.980	1.716	0.092	0.007	0.001	0.002	0.20
age	-0.025	57.115	1.716	-0.822	0.112	-0.092	-0.304	-30.35
age2	0.139	33.990	1.716	2.752	0.222	0.610	2.018	201.78
age3	-0.063	21.058	1.716	-0.776	0.326	-0.253	-0.836	-83.60
costant	-2.565							
$\sum_k \eta_k C_k$						0.303		
Total Gini						0.303		

Table B.16: Decomposition of the Gini: Included Deaths Eq. - Women

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.286	0.713	2.074	-0.099	-0.026	0.003	0.010	0.951
breakfast	-0.207	0.714	2.074	-0.071	-0.003	0.000	0.001	0.090
sleepgd	-0.213	0.587	2.074	-0.060	-0.142	0.009	0.032	3.174
alqprud	-0.124	0.939	2.074	-0.056	0.003	0.000	-0.001	-0.063
nobese	-0.133	0.805	2.074	-0.051	-0.029	0.001	0.006	0.554
exercise	-0.279	0.318	2.074	-0.043	-0.335	0.014	0.053	5.311
sc12	-0.293	0.326	2.074	-0.046	-0.252	0.012	0.043	4.309
sc45	-0.085	0.221	2.074	-0.009	0.145	-0.001	-0.005	-0.488
lhqdg	-0.129	0.117	2.074	-0.007	-0.327	0.002	0.009	0.881
lhqhndA	-0.189	0.126	2.074	-0.011	-0.299	0.003	0.013	1.271
lhqnone	-0.023	0.629	2.074	-0.007	0.163	-0.001	-0.004	-0.414
lhqoth	-0.017	0.030	2.074	0.000	-0.064	0.000	0.000	0.006
wkshft1	-0.291	0.034	2.074	-0.005	-0.557	0.003	0.010	0.972
rural	-0.282	0.220	2.074	-0.030	-0.195	0.006	0.022	2.157
suburb	-0.057	0.477	2.074	-0.013	0.003	0.000	0.000	-0.014
ethwheur	0.233	0.980	2.074	0.110	0.006	0.001	0.003	0.265
age	-0.049	57.115	2.074	-1.348	0.108	-0.146	-0.540	-53.954
age2	0.182	33.990	2.074	2.976	0.215	0.640	2.372	237.161
age3	-0.086	21.058	2.074	-0.870	0.317	-0.276	-1.022	-102.168
costant	-1.468							
$\sum_k \eta_k C_k$						0.270		
Total Gini						0.270		

Table B.17: Decomposition of the Gini: Endogenous Deaths Eq. - Women

Variable	Coeff.	x_k	$\bar{x\hat{\beta}}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.393	0.713	2.129	-0.132	-0.030	0.004	0.014	1.44
breakfast	-0.010	0.714	2.129	-0.003	0.011	0.000	0.000	-0.01
sleepgd	-0.877	0.587	2.129	-0.242	-0.278	0.067	0.247	24.68
alqprud	0.297	0.939	2.129	0.131	0.019	0.002	0.009	0.90
nobese	-0.913	0.805	2.129	-0.345	-0.100	0.034	0.127	12.66
exercise	-0.167	0.318	2.129	-0.025	-0.262	0.007	0.024	2.40
sc12	-0.155	0.326	2.129	-0.024	-0.221	0.005	0.019	1.93
sc45	-0.045	0.221	2.129	-0.005	0.121	-0.001	-0.002	-0.21
lhqdg	-0.097	0.117	2.129	-0.005	-0.253	0.001	0.005	0.50
lhqhndA	-0.176	0.126	2.129	-0.010	-0.250	0.003	0.010	0.95
lhqnone	-0.050	0.629	2.129	-0.015	0.142	-0.002	-0.008	-0.78
lhqoth	0.005	0.030	2.129	0.000	-0.041	0.000	0.000	0.00
wkshft1	-0.400	0.034	2.129	-0.006	-0.270	0.002	0.006	0.62
rural	-0.185	0.220	2.129	-0.019	-0.166	0.003	0.012	1.16
suburb	0.002	0.477	2.129	0.000	0.002	0.000	0.000	0.00
ethwheur	0.333	0.980	2.129	0.153	0.005	0.001	0.003	0.28
age	0.040	57.115	2.129	1.085	0.091	0.098	0.362	36.16
age2	-0.002	33.990	2.129	-0.032	0.181	-0.006	-0.021	-2.11
age3	0.020	21.058	2.129	0.198	0.267	0.053	0.194	19.42
costant	-2.142							
$\sum_k \eta_k C_k$						0.272		
Total Gini						0.276		

Table B.18: Decomposition of the Gini: Endogenous Deaths Eq. (reduced model) - Women

Variable	Coeff.	x_k	$x\hat{\beta}$	η_k	C_k	$\eta_k C_k$	contribution	% contribution
nsmoker	-0.265	0.713	2.112	-0.090	-0.014	0.001	0.005	0.47
breakfast	-0.192	0.714	2.112	-0.065	-0.010	0.001	0.002	0.23
sleepgd	-0.487	0.587	2.112	-0.135	-0.198	0.027	0.099	9.94
alqprud	-0.144	0.939	2.112	-0.064	0.004	0.000	-0.001	-0.09
nobese	-0.956	0.805	2.112	-0.364	-0.106	0.039	0.143	14.35
exercise	-0.257	0.318	2.112	-0.039	-0.308	0.012	0.044	4.41
sc12	-0.210	0.326	2.112	-0.032	-0.237	0.008	0.028	2.85
sc45	-0.069	0.221	2.112	-0.007	0.131	-0.001	-0.003	-0.35
lhqdg	-0.100	0.117	2.112	-0.006	-0.271	0.001	0.006	0.55
lhqhndA	-0.187	0.126	2.112	-0.011	-0.270	0.003	0.011	1.12
lhqnone	-0.038	0.629	2.112	-0.011	0.152	-0.002	-0.006	-0.64
lhqoth	0.011	0.030	2.112	0.000	-0.043	0.000	0.000	0.00
wkshft1	-0.339	0.034	2.112	-0.005	-0.294	0.002	0.006	0.59
rural	-0.205	0.220	2.112	-0.021	-0.174	0.004	0.014	1.38
suburb	0.015	0.477	2.112	0.003	0.000	0.000	0.000	0.00
ethwheur	0.339	0.980	2.112	0.158	0.005	0.001	0.003	0.32
age	-0.036	57.115	2.112	-0.980	0.098	-0.096	-0.354	-35.41
age2	0.140	33.990	2.112	2.253	0.194	0.437	1.617	161.72
age3	-0.058	21.058	2.112	-0.581	0.286	-0.166	-0.614	-61.44
costant	-0.683							1.03
$\sum_k \eta_k C_k$						0.270		
Total Gini						0.275		