Exploring variation in the demand for care using Grade of Membership analysis

Paul McNamee

Department of Epidemiology & Public Health
School of Health Sciences
University of Newcastle UK

Abstract

An important focus in empirical demand studies of health care concerns estimation of the relationship with health status. A particular issue relates to the characterisation of health status in terms of 'objective' measures of physical or mental functioning. This paper describes the use of a nonparametric technique, the Grade of Membership procedure, to provide summary measures of health. Using cross-sectional data from a stratified randomised sample of frail older people, demand for care is estimated with health treated as an endogenous variable. The results show that health can be usefully classified using Grade of Membership techniques, and that health status is an important factor in explaining variation in service use.

Acknowledgements

The study was funded by the Medical Research Council and the Department of Health. I am grateful to James C Wagner of the Center for Demographic Studies, Duke University, North Carolina for the supply of GoM software and technical help.

Introduction

The use of multiple observable indicators of health and their relationship to unobservable constructs, such as “true” health, is a focus of many studies which attempt to explore demand for health care. Multiple indicators are often used when more than one indicator is required to measure latent dimensions. The problem however with such indicators is that a method is required to summarize the data for subsequent multivariate analysis. Typically this can be handled by construction of an index which counts the number of health conditions present (Dwyer and Mitchell 1999, Borsch-Supan et al. 1992) or uses pre-existing scaling methods (Katz et al. 1963). While these can perform as well as more sophisticated measures (Borsch-Supan et al. 1989), the statistical properties of these methods depend on certain assumptions regarding the distribution of latent constructs.

A method which does not make any parametric assumptions about the distribution of variables is the Grade of Membership (GoM) procedure developed by Woodbury and colleagues (Clive et al. 1983; Woodbury et al. 1978; Woodbury and Clive 1974). This approach uses observable indicators to generate latent discrete classes which are defined as combinations of multiple constructs. An output of this method is the assignment of values for sample observations and indicator variables to represent the degree of proximity to each latent class. At a conceptual level, this procedure differs from Factor Analysis, which uses indicators to calculate latent continuous variables that represent one dimensional constructs. Statistically, the major difference is that values are simultaneously determined in GoM, whilst classification using Factor Analysis is conducted conditionally upon structural pararmeters using Bayes’ theorem (Manton et al. 1992).

A priori, the GoM model seems an appropriate tool for the study described in this paper. The measurement of the indicators is based on discrete values, and the multidimensional nature of health is retained within the profiles. In addition, individual heterogeneity is explicitly modelled through the degree of proximity to the identified profiles. Finally, the indicators combine non-additively to form the profiles.

The rest of the paper is organized as follows. The next section provides a description of the dataset, a summary of the GoM procedure, the variables used to calculate GoM parameters and the econometric model. This is followed by presentation of output from the GoM analysis and use of the identified profiles in the econometric specification. The concluding section summarizes the results.

Methods

Dataset

Data presented here were collected in a five year longitudinal multi-centre Resource Implications Study (RIS), which formed part of the MRC study of Cognitive Function and Ageing (Parker et al. 1997). The aim of RIS was to assess the resource implications to families, health and personal social services of mental and physical frailty among people aged 65 years of age or over using an age stratified randomised sample. Service use data were collected over two years between 1991-1995 for each subject living in private households or resident in continuing care accommodation. Mentally frail subjects were defined as those obtaining a level of three or more on the organic section of the Automated Geriatric Examination Computer Assisted Taxonomy (AGECAT, Copeland et al. 1986), which is approximately equivalent to moderate (or more) clinical dementia. Physically frail subjects were identified as those with a score of eleven or more on the Clackmannan Disability Scale (Bond and Carstairs 1982).

Data were collected from four geographical areas in England: Cambridgeshire, Newcastle, Nottingham and Oxford. These centres were chosen in order to capture the main national variations in urban-rural differences, socio-economic levels and rates of chronic disease, such as stroke, ischaemic heart disease and cancers of lung, stomach and breast. Analysis of response rates, extent of institutionalisation and socio-demographic variables suggest that the sample is nationally representative. This paper reports data for 1,283 individuals who were diagnosed as either mentally, physically or mentally and physically frail.

Grade of Membership analysis

Grade of Membership (GoM) analysis can be used as a procedure to determine latent profiles (pure types) of health and to show the degree to which individuals correspond to the identified profiles. More formally, assume there are *K* fuzzy states (or pure types) to be defined (*k* = 1,2,…,*K*). The study population comprises *I* individuals (*I* = 1,2,…,*I*), with *J* categorical variables, where the *j*th variable has *Lj* response levels. Each *Lj* response is coded as a binary variable *xijl*, so that if *xijl*=1 then the *i*th individual has the *l*th response to the *j*th variable. Assume further that *l kjl* is the probability that the *l*th response to the *j*th variable is associated with the *k*th fuzzy state and that *gik* is the weight for the *i*th individual for each of the *K* states, with 0<=*gik* <=1 and Σ*kgik* = 1.

The product *gik l kjl* therefore depicts the degree of involvement of an individual to a specific class *k* and the probability that an individual who belongs entirely in *k* (ie those who are only associated with a single profile or pure type) responds *l* to question *j*. This probability is given by:

$$P\,(xijl = 1.0) = \sum_{k} gik\,l\,kjl \qquad [1]$$

where the parameters *l kjl* and *gik* are estimated by maximising the likelihood function:

$$L = \prod_{i}\prod_{j}\prod_{l}\Big(\sum_{k} gik\,l\,kjl\Big)^{xijl} \qquad [2]$$

subject to: $0 <= gik <= 1, \quad \sum_{k} gik = 1,$

and $0 <= l\,kjl <= 1, \quad \sum_{l} l\,kjl = 1.$

Estimation of the model involves the simultaneous determination of *l kjl* and *gik* using a modified Newton-Raphson algorithm. Full details can be found in Manton and Stallard (1988). Briefly, with *l kjl* fixed, the likelihood function is maximized to estimate *gik*, then, fixing *gik*, the likelihood is maximized to update the *lkjl*, which is repeated until convergence. The optimal number of profiles is determined using likelihood ratio statistics. Based on an initial value of K, an optimal solution is found through calculation of χ^2 test statistics for K, K+1...n and K-1...n models. Differences in the value of the log likelihood are doubled to obtain these estimates for successive models, with degrees of freedom equal to the difference in the number of parameters to be estimated between models:

$$df = (I - 1) \times (K - 1) + (K-1) \times \sum_{j} L_j \qquad [3]$$

Estimation procedure

The problem faced is to summarize health status using the available indicator variables. Using the available dataset, a total of 27 physical and mental health variables were used to characterize latent health (J = 27).

The first set of items asked respondents whether they needed help, had difficulty with or had no problems with the following activities: cutting toenails, washing themselves, getting on a bus, going up and down stairs, heavy housework, shopping, cooking a hot meal, reaching a shelf, tying a knot and putting on shoes and socks.

A series of chronic health items asked the sample whether they currently suffer from or recently experienced the following conditions: leg pain, high blood pressure, stroke, diabetes, epilepsy, hearing problems, poor eyesight, asthma, arthritis, bronchitis, thyroid problems, peptic ulcer disease, anaemia and Parkinson’s disease.

The final physical functioning indicator variable consisted of a question which asked respondents to rate themselves as ambulatory, housebound, chairfast or bedfast.

Two cognitive measures were used: AGECAT, where a score of between 3 and 5 indicated the presence of likely dementia, and the Mini Mental State Examination

(MMSE), where values of 0-10, 11-22, 23-30 indicated severe, moderate/mild, and absence of cognitive impairment respectively.

In total there were therefore 10 Activities of Daily Living (ADL) items (with Lj = 3 for each), 14 chronic health condition items (with Lj = 2 for each), 1 mobility item (Lj = 4) and 2 cognitive function items (Lj = 2, Lj = 3), giving a total of 67 *xijl* response variables to be estimated for 1283 individuals.

Sample covariates

Four sociodemographic covariates were used to describe the characteristics associated with the pure types. These were age (coded into five categories: 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85+ years), sex (female=1, male=0), marital status (coded into five categories: married, cohabiting, single, widowed, divorced) and education (primary and secondary education=1, higher education=0).

In addition, six other covariates can be used to describe the pure types: accommodation (coded into six categories: own home, warden, local authority residential home, private residential home, private nursing home, long stay hospital), structure (living with others in own home=1, living alone in own home=0), contact (frequency of most frequent contact for living at home, coded into five categories: daily, 2-3 times per week, weekly, monthly, little), health (self-assessed health, coded into four categories: excellent, good, fair, poor), smoking (yes=1, no=0) and caretype (coded into four categories: informal care only, formal care in the community (<=£50 per week, with or without informal care), resource intensive formal care in the community (>£50 per week, with or without informal care) and long term care in residential/nursing homes or hospital).

Econometric model

Suppose there is a bundle of commodities, $\xi_{1\ldots n}$, which represent the type of care received. In the context of frail older people, the bundle can be composed of any number of different services. Assume however that these are classified into one of the four types identified above.

The demand for these commodities is assumed to depend upon latent health status η^*, a vector Z_1 of observable characteristics affecting demand (age, sex, marital status, household structure, informal carer contact, education) and a stochastic error term ε_1:

$$\xi_{1\ldots n}, \quad = \quad \eta^* + Z_1 + \varepsilon_1 \qquad [4]$$

Assume further that latent health status depends upon utilisation of commodities $\xi_{1\ldots n}$, a similar vector Z_2 of observable factors influencing health and a stochastic error term ε_2:

$$\eta^* \quad = \quad \xi_{1\ldots n} + Z_2 + \varepsilon_2 \qquad [5]$$

In this general formulation there is a potential simultaneous equations problem, which requires investigation and special estimation strategies.

Latent health status is proxied by grade of membership, *gik*. Following Portrait et al. (1999), it is conceivable that these could be influenced by unobservable characteristics, $\alpha_{1\ldots n.}$ Therefore the following measurement error model can be written:

$$gik^{1\ldots n} \quad = \quad \eta^* + \alpha_{1\ldots n} + \varepsilon_3 \qquad [6]$$

In [5], if unobserved characteristics not included in $\xi_{1\ldots n}$, and Z_2 influence both $gik^{1\ldots n}$ and the regressors, then the latter will be correlated with error terms $\varepsilon_{1,}\varepsilon_{2,}\varepsilon_3$. Parameter estimates will then be biased and inconsistent with Ordinary Least Squares (OLS) estimation.

It is apparent that this problem may be treated as a special case of endogeneity. Hausman exogeneity tests (1978) can therefore be employed to detect for the presence of unobservable heterogeneity. Specifically, $\xi_{1\ldots n}$, and elements within Z_2 suspected of being endogenous (dummy variables representing smoking, informal care contact and household structure) were each separately tested against a core set of exogenous variables (age, sex, marital status and education) using a two stage procedure. In the first stage, each suspected endogenous variable was regressed on the core set of exogenous variables to obtain instrumental variable (IV) estimates. In the second stage, the IVs were used in a regression which also included the suspected endogenous variable and the core set of exogenous variables.

The above general estimation strategies led to the development of two specific models:

1 Simultaneous structural equations model

This model described health and demand using a system of simultaneous equations. Four structural health and three structural demand equations were estimated. These used estimated values for health and demand from reduced form equations alongside socio-demographic variables. To ensure identification, zero restrictions were placed on variables found to be statistically insignificant in the first stage reduced form. Specifically, we have (see table 4 for coding of variables):

$$H_n = \alpha_1 + \beta_1 F_1^* + \beta_2 F_2^* + \beta_3 F_3^* + \beta_4 X_1 + \ldots + \beta_5 X_3 + \ldots + \beta_6 X_5 + \beta_7 X_6 + \beta_8 X_7 + \beta_9 X_8 + \beta_{10} X_9 + \beta_{11} X_{10} + \beta_{12} X_{11} + \beta_{13} X_{12} + \ldots + \varepsilon_n$$

thereby omitting mar1 (X_2), mar3 (X_4), smoke (X_{13}), and where n=1...3,5;

$$F_1, F_2 = \alpha_1 + \beta_1 H_1^* + \beta_2 H_2^* + \beta_3 H_3^* + \beta_4 H_5^* + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3 + \beta_8 X_4 + \ldots + \beta_9 X_6 + \beta_{10} X_7 + \beta_{11} X_8 + \beta_{12} X_9 + \ldots + \ldots + \beta_{13} X_{12} + \ldots + \varepsilon_{F1,2}$$

thereby omitting school (X_5), inform1 (X_{10}), inform2 (X_{11}) and smoke (X_{13});

$$F_3 = \alpha_1 + \beta_1 H_1^* + \beta_2 H_2^* + \beta_3 H_3^* + \beta_4 H_5^* + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3 + \beta_8 X_4 + \beta_9 X_5 + \ldots + \ldots + \beta_{10} X_8 + \beta_{11} X_9 + \ldots + \ldots + \ldots + \beta_{12} X_{13} + \varepsilon_{F3}$$

thereby omitting age1 (X_6), age2 (X_7), inform1 (X_{10}), inform2 (X_{11}) and struct (X_{12}).

2 Recursive system

The second approach assumed a recursive system, with demand determined by health and a vector of co-variates, whilst health is dependent upon co-variates only. Specifically:

$$H_n = \alpha_1 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \varepsilon_n$$

$$F_1 = \alpha_1 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_5 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3 + \beta_8 X_4 + \beta_9 X_5 + \beta_{10} X_6 + \beta_{11} X_7 + \beta_{12} X_8 + \beta_{13} X_9 + \beta_{14} X_{10} + \beta_{15} X_{11} + \beta_{16} X_{12} + \beta_{17} X_{13} + \varepsilon_{F1}$$

$$F_2 = \alpha_1 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_5 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 X_3 + \beta_8 X_4 + \beta_9 X_5 + \beta_{10} X_6 + \beta_{11} X_7 + \beta_{12} X_8 + \beta_{13} X_9 + \beta_{14} X_{10} + \beta_{15} X_{11} + \beta_{16} X_{12} + \beta_{17} X_{13} + \varepsilon_{F2}$$

$$F_3 = \alpha_1 + \beta_1 H_1 + \beta_2 H_2 + \beta_3 H_3 + \beta_4 H_5 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \ldots + \ldots + \ldots + \beta_{10} X_{13} + \varepsilon_{F3}$$

thereby omitting inform1 (X_{10}), inform2 (X_{11}) and struct (X_{12}).

Results

Using the χ^2 test, the change in log likelihood was significant after adding a fifth dimension and not significant after adding a sixth dimension (Log Likelihood ratio= 6905 df=5504, Log Likelihood ratio= 7214 df=6880 with K=5 and K=6 respectively). Table 1 shows the relationship between indicator variables and pure types by the profile probabilitites λ_{kjl}. The first column depicts responses of the sample to each indicator variable and co-variate. The columns associated with each pure type display the λ_{kjl}, the probability that a particular response is associated with a particular profile.

Considering the indicator variables only, Type I is non-demented, although requires help with a range of activities of daily living. They tend to suffer from some chronic conditions, especially leg pain, high blood pressure and arthritis. They are housebound. They can be labelled as PHYSICALLY FRAIL.

Type II is likely to suffer from dementia. They do not need help with activities of daily living, but some can only be performed with difficulty. They do not tend to suffer from chronic conditions. They are ambulatory. They can be labelled as MENTALLY FRAIL.

Type III is non-demented but suffers from some moderate cognitive impairment. They are likely to require help with all activities of daily living and suffer from a range of conditions and diseases , in particular high blood pressure, previous stroke, diabetes, epilepsy and Parkinson's disease. They are either housebound or bedbound. They can be labelled as PHYSICALLY FRAIL, CHRONIC.

Type IV is the healthiest group. They are non-demented and tend to require help with only some activities of daily living, although they have difficulty with several others. They are likely to suffer more from arthritis and are ambulatory. Relative to the other types, they can be classed as HEALTHY.

Type V is the frailest group. They are either demented or suffer from moderate impairment. In addition, they need help with virtually all activities, tend to be either housebound or chairbound. However, they do not suffer from chronic health conditions. They can be described as VERY FRAIL.

Table 2 presents the association between sample co-variates and pure types. PHYSICALLY FRAIL tend to be female, widowed, living alone in their own home and have frequent informal care contacts. They rate their own health mostly as fair or poor, and receive some formal care services.

The MENTALLY FRAIL tend to live in their own home with a spouse or other relative or friend. They receive less visits from other informal carers. They rate their own health as excellent or good. A substantial minority receive no formal care services.

The PHYSICALLY FRAIL CHRONIC tend to be male, married, living in the community and aged less than 80 years. They receive a number of visits from informal carers. Self rated health is fair or poor and the majority receive formal care.

The HEALTHY tend to be female, reside in their own home alone and older than 80 years of age. They receive very few formal care services and rate their health as good or fair.

The VERY FRAIL also tend to be female, widowed or single and older than 80 years of age. They are mostly resident in long term care homes.

Hausman specification tests (Table 3) indicated the presence of simultaneity [1], implying the need for modelling strategies of the kind developed in the paper. First, reduced form logit regressions for health were conducted, with $H_n = 1$ if *gik* >0, and $H_n = 0$ otherwise.

Table 4 shows that subjects who were single, received primary and secondary education only, were aged 70 years or over, and did not live alone were less likely to be found in the Type I category. For Type II, subjects were more likely to be male, single, aged 70 years or over and be in receipt of regular informal care. Type III subjects were significantly more likely to be male and under 85 years of age. They were likely to have regular informal care support and live with others. For Type IV, the healthy category, there were no statistically significant relationships with any co-variates. The very frail category (Type V) were significantly more likely to be aged 75 years or over.

Logit regressions were also employed to estimate reduced form demand for care equations. As Table 5 shows, in contrast to the health reduced form health equations, very few statistical associations were revealed. Subjects who were single were less likely to receive infrequent formal care, whilst no variables were associated with the likelihood of receiving frequent formal care in the community. Care in long-term settings was positively associated with being single and negatively associated with marriage/co-habitation. Older age was also a significant predictor. A negative relationship was found with smoking.

[1] However, Hausman tests for exogeneity showed that the null hypothesis was accepted in all cases, so IV estimation was not undertaken.

The observation that few socio-demographic variables predicted type of care received suggests that health plays a large part in determining demand. Conversely, it may also be hypothesised that health status is to some extent dependent upon service use. To account for these possibilities, the modelling strategies described above were estimated using two-stage least squares. Tables 6 and 7 report estimates obtained from model 1. In Table 6, significant t-ratios for all service types suggest that utilisation is an important determinant of health. There were positive associations between individuals classed as physically frail, chronic or very frail and service receipt. Women were less likely to be classed as physically frail, chronic or very frail. Subjects who never married were more likely to be very frail. Tertiary educated individuals seemed less likely to suffer from mental or physical frailty only. Older age (>85 years) had a strong positive association with the very frail category, whilst younger age groups (70-79 years) were more likely to be healthy or suffer from cognitive impairment. Weekly/monthly informal care contact was positively associated with all health states other than the physically frail, chronic or very frail categories. Subjects with cognitive impairment were more likely to live with someone in their own home, whilst living alone was associated with the very frail category.

In Table 7, due to large standard errors found among the majority of variables, only health types I and II (physically frail and mentally frail respectively) were found to have a significant association with receipt of long-term care. Widowhood and primary/secondary education only also displayed positive relationships with receipt of long-term care.

Tables 4 and 8 displays model 2, which assumes health is a determining factor in the demand for care, with no causation running in the opposite direction. Subjects in poorer health (types I-III, V) were less likely to receive smaller amounts of formal care. Types III and V were more likely to receive long-term care. Subjects who were married were less likely to receive long-term care, whilst unmarried individuals were less likely to consume smaller amounts of formal care. People living with others were more likely to receive smaller levels of formal care. Finally, subjects who smoked were less likely to be resident in long term care.

Summary and concluding remarks

This paper described the use of Grade of Membership analysis as a technique for summarising health in multivariate modelling of the demand for care. Five main health types were produced, each of which appeared to capture different facets of physical and mental frailty. These descriptors of health were then used to assess their association with demand for care.

The key methodological issue of the cross sectional dataset was the need to explore a number empirical strategies to account for simultaneity bias. From the reduced form equations, an important observation was that few socio-demographic variables predicted type of care receipt. An obvious reason for this could be the omission of health as an endogenous variable. However, the two-stage estimation strategy appeared unsuccessful, as health was found to show no statistical relationship with demand in the structural equation. A likely reason for this relates to the lack of suitable instruments for health, as found elsewhere (Kenkel 1995).

A more appealing strategy, although more restrictive in terms of the assumption of zero contemporaneous correlation, may be found in the recursive model. Health was found to be significantly associated with demand, whilst a number of plausible socio-demographic variables were significantly related to health. This framework however assumes that health is not affected by type of care receipt. Clearly the supply of help (informal or formal) is likely to be an important determinant in an individuals' capacity to carry out tasks. Therefore, as a large proportion of health is measured by activity of daily living items, this assumption is likely to be false in some cases.

To account for these factors, joint estimation may be a preferable strategy. Further work is planned to explore this approach, to compare qualitative and quantitative changes in parameter estimates with the models assessed in this paper.

References

Bond, J., Carstairs, V., 1982. Services for the elderly: a survey of the characteristics and needs of a population of 5,000 older people. Scottish Health Service Studies No.42, Scottish Home and Health Department, Edinburgh.

Borsch-Supan, A., Hajivassilou, V., Kotlikoff, L.J., Morris, J.N., 1992. Health, children, and elderly living arrangements: a multiperiod-multinomial model probit model with unobserved heterogeneity and autocorrelated errors, in: Wise, D. (Ed.), The Economics of Aging. Univ of Chicago Press, Chicago, pp. 79-104.

Borsch-Supan, A., Kotlikoff, L.J., Morris, J.N., 1989. The dyanamics of living arrangements of the elderly: health and family support. NBER Working Paper. National Bureau of Economic Research, Cambridge.

Clive, J.M., Woodbury, M.A., Siegler, I.C., 1983. Fuzzy and crisp set theoretic based classification of health and disease. J. Med. Systems. 7, 317-322.

Copeland, J., Dewey, M., Griffiths-Jones H., 1986. A computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. Psychol Med 16, 89-99.

Dwyer, D.S., Mitchell, O.S., 1999. Health problems as determinants of retirement: are self-rated measures endogenous? J. Health Econ. 18, 173-193.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975 Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J. Psych Research 12, 189-198.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 931-959.

Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W., 1963. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychological function. JAMA. 185, 914-919.

Kenkel, D.S., 1995. Should you eat breakfast? Estimates from health production functions. Health Econ. 4, 15-29.

Manton, K.G., Woodbury, M.A., Stallard, E., Corder, L.S., 1992. The use of Grade of Membership techniques to estimate regression relationships, in: Marsden, P.V. (Ed.), Sociological Methodology. American Sociological Association, Washington, pp.321-381.

Manton, K.G., Stallard, E., 1988. Chronic disease modelling: measurement and evaluation of the risks of chronic disease processes. Oxford University Press, New York.

MRC CFAS. 1998. Cognitive function and dementia in six areas of England and Wales: the distribution of MMSE and prevalence of GMS organicity level in the MRC CFA Study. Psychol Med. 28, 319-335.

Portrait, F., Lindeboom, M., Deeg, D., 1999. Health and mortality of the elderly: the Grade of Membership method, classification and determination. Health Econ. 8, 441-457.

Woodbury, M.A., Clive, J.M., Garson, A., 1978. Mathematical typology: a grade of membership technique for obtaining disease definition. Computers and Biomed. Research. 11, 277-298.

Table 1 PURE TYPE PROBABILITIES

	Freq	I	II	III	IV	V
Agecat score						
0-2	68.82	100.00	0.00	100.00	100.00	53.87
3-5	31.18	0.00	100.00	0.00	0.00	46.13
MMSE score						
00-10	2.41	0.00	3.74	0.00	0.00	6.60
11-22	47.31	0.00	86.69	100.00	0.00	93.40
23-30	50.28	100.00	9.57	0.00	100.00	0.00
Leg pain						
no	48.98	6.42	87.25	64.04	48.30	64.09
yes	44.67	93.58	12.75	35.96	51.70	7.01
immobile	6.35	0.00	0.00	0.00	0.00	28.90
High blood pressure						
no	68.58	31.47	82.30	0.00	100.00	100.00
yes	31.42	68.53	17.70	100.00	0.00	0.00
Stroke						
no	78.05	100.00	100.00	0.00	100.00	100.00
yes	21.95	0.00	0.00	100.00	0.00	0.00
Diabetes						
no	89.45	100.00	100.00	23.74	100.00	100.00
yes	10.55	0.00	0.00	76.26	0.00	0.00
Epilepsy						
no	96.63	100.00	100.00	67.90	100.00	100.00
yes	3.37	0.00	0.00	32.10	0.00	0.00
Hearing problems						
no	68.03	100.00	87.74	100.00	22.61	42.01
yes	31.97	0.00	12.26	0.00	77.39	57.99
Poor eyesight						
no	66.64	56.08	90.94	100.00	62.31	53.08
yes	33.36	43.92	9.06	0.00	37.69	46.92
Asthma						
no	89.51	76.42	100.00	100.00	89.57	92.82
yes	10.49	23.58	0.00	0.00	10.43	7.18
Arthritis						
no	37.25	0.00	80.72	100.00	0.00	41.13
yes	62.75	100.00	19.28	0.00	100.00	58.87
Bronchitis						
no	80.74	65.88	95.07	75.07	86.52	85.88
yes	19.26	34.12	4.93	24.93	13.48	14.12
Thyroid problems						
no	92.22	80.02	100.00	80.52	100.00	100.00
yes	7.78	19.98	0.00	19.48	0.00	0.00
Peptic ulcer disease						
no	88.64	76.94	95.14	82.03	90.07	100.00
yes	11.36	23.06	4.86	17.97	9.93	0.00
Anaemia						
no	95.20	86.65	97.17	95.23	100.00	100.00
yes	4.80	13.35	2.83	4.77	0.00	0.00
Parkinson's disease						
no	96.16	100.00	100.00	66.49	100.00	100.00
yes	3.84	0.00	0.00	33.51	0.00	0.00

Table 1 continued		**PURE TYPE PROBABILITIES**				
	Freq	I	II	III	IV	V
Difficulties cutting toenails						
needs help	75.31	69.83	0.00	100.00	100.00	100.00
some difficulty	13.40	30.17	29.79	0.00	0.00	0.00
no problems	11.29	0.00	70.21	0.00	0.00	0.00
Difficulties washing self						
needs help	47.02	38.55	0.00	100.00	0.00	100.00
some difficulty	30.88	61.45	0.00	0.00	100.00	0.00
no problems	22.10	0.00	100.00	0.00	0.00	0.00
Difficulties getting on a bus						
needs help	68.81	100.00	0.00	100.00	0.00	100.00
some difficulty	18.10	0.00	0.00	0.00	100.00	0.00
no problems	13.09	0.00	100.00	0.00	0.00	0.00
Difficulties going up and down stairs						
needs help	49.14	100.00	0.00	100.00	0.00	100.00
some difficulty	36.76	0.00	0.00	0.00	100.00	0.00
no problems	14.11	0.00	100.00	0.00	0.00	0.00
Difficulties with heavy housework						
needs help	81.43	100.00	0.00	100.00	75.51	100.00
some difficulty	8.93	0.00	31.73	0.00	24.49	0.00
no problems	9.64	0.00	68.27	0.00	0.00	0.00
Difficulties with shopping						
needs help	85.03	100.00	0.00	100.00	85.91	100.00
some difficulty	8.07	0.00	44.51	0.00	14.09	0.00
no problems	6.90	0.00	55.49	0.00	0.00	0.00
Difficulties cooking a hot meal						
needs help	47.34	0.00	0.00	100.00	0.00	100.00
some difficulty	23.98	100.00	0.00	0.00	33.14	0.00
no problems	28.68	0.00	100.00	0.00	66.86	0.00
Difficulties reaching a shelf						
needs help	58.70	100.00	0.00	100.00	0.00	100.00
some difficulty	22.18	0.00	0.00	0.00	100.00	0.00
no problems	19.12	0.00	100.00	0.00	0.00	0.00
Difficulties tying a knot						
needs help	18.89	0.00	0.00	100.00	0.00	0.00
some difficulty	19.59	54.06	0.00	0.00	0.00	22.34
no problems	61.52	45.94	100.00	0.00	100.00	77.66
Difficulties putting on shoes and socks						
needs help	18.67	0.00	0.00	0.00	0.00	100.00
some difficulty	36.94	100.00	0.00	100.00	0.00	0.00
no problems	44.39	0.00	100.00	0.00	100.00	0.00
Mobility						
ambulatory	41.35	0.00	100.00	0.00	100.00	0.00
housebound	48.57	100.00	0.00	96.77	0.00	59.39
chairbound	9.76	0.00	0.00	0.00	0.00	39.90
bedfast	0.32	0.00	0.00	3.23	0.00	0.71

Table 2 PURE TYPE PROBABILITIES

	Freq	I	II	III	IV	V
Sex						
male	27.90	9.68	43.67	68.92	29.85	20.19
female	72.10	90.32	56.33	31.08	70.15	79.81
Marital status						
married	32.42	27.07	35.59	76.56	28.50	22.08
cohabit	0.31	0.00	0.45	0.64	0.00	0.60
single	10.37	8.49	15.05	0.96	8.92	14.07
widow	54.40	59.26	47.19	17.66	61.65	62.84
divorce	2.49	5.18	1.71	4.18	0.93	0.41
Accommodation type						
own	69.89	81.52	78.57	76.49	79.59	36.83
warden	14.35	17.69	14.44	0.37	17.47	12.83
crhome	4.99	0.79	3.20	8.92	0.00	14.67
prhome	3.98	0.00	2.14	6.86	2.94	10.34
pnhome	6.01	0.00	1.65	5.26	0.00	22.78
hospital	0.78	0.00	0.00	2.09	0.00	2.54
Household structure						
alone	48.71	65.50	46.91	0.00	60.49	39.33
with	51.29	34.50	53.09	100.00	39.51	60.67
Education						
no tertiary	75.56	64.93	82.18	91.36	78.03	75.08
tertiary	24.44	35.07	17.82	8.64	21.97	24.92
Frequency of most frequent contact						
daily	18.81	15.21	14.09	20.73	22.03	22.69
2-3pw	25.36	32.72	23.85	24.53	17.46	24.42
weekly	33.33	35.49	32.75	37.26	31.00	31.40
monthly	11.96	8.45	15.42	17.48	11.46	11.55
little	10.53	8.13	13.89	0.00	18.07	9.94
Self assessed health						
excellent	7.04	0.00	24.02	0.00	4.41	8.59
good	32.64	7.54	49.10	16.27	44.21	48.24
fair	38.80	47.79	23.99	37.11	47.34	32.53
poor	21.52	44.66	2.89	46.62	4.05	10.65
Current smoker						
no	84.75	87.72	81.39	86.87	74.60	90.56
yes	15.25	12.28	18.61	13.13	25.40	9.44
Age						
65-69	9.04	15.26	5.36	21.95	5.21	0.83
70-74	10.91	14.18	17.30	23.94	6.15	0.00
75-79	20.97	25.90	27.00	37.25	13.03	9.15
80-84	27.59	28.35	31.26	16.86	33.73	25.23
85+	31.49	16.32	19.08	0.00	41.89	64.79
Service type						
no formal	7.75	1.95	22.59	9.26	7.17	4.27
formal1	45.51	58.63	42.63	22.10	79.04	12.47
formal2	26.43	34.07	19.66	35.18	12.79	29.55
formal3	20.31	5.35	15.12	33.45	1.00	53.70

Table 3 HAUSMAN SPECIFICATION TESTS

	Type I		Type II		Type III		Type IV		Type V	
Formal1*	-0.21	(1.11)	-0.81	(3.73)	0.17	(1.62)	-0.32	(2.22)	1.25	(6.51)
Formal2*	-0.54	(2.06)	-0.36	(1.20)	-0.06	(0.41)	-0.02	(0.08)	1.05	(3.99)
Formal3*	-0.01	(0.16)	-0.40	(4.03)	-0.07	(1.41)	-0.16	(2.41)	0.67	(7.63)
Rformal1	0.07	(2.72)	-0.17	(6.36)	0.01	(1.06)	0.05	(2.60)	0.04	(1.71)
Rformal2	0.02	(0.59)	-0.23	(7.59)	0.06	(4.08)	-0.05	(2.39)	0.22	(8.04)
Rformal3	-0.08	(3.79)	-0.13	(5.67)	0.05	(4.00)	-0.07	(4.58)	0.24	(11.65)
F	6.32		14.47		11.51		13.26		43.94	
Sig F	0.00		0.00		0.00		0.00		0.00	

From equation 4, tests were constructed for $\xi_{1\ldots n}$, with regressions on the vector $Z_{1,}$ followed by regression of η^* on the estimated values and residuals obtained in the first step.

Table 4 REDUCED FORM EQUATIONS - HEALTH

	Type I		Type II		Type III		Type IV		Type V	
Sex	0.21	(1.18)	-0.52	(3.40)	-0.37	(2.41)	0.05	(0.33)	0.10	(0.68)
Mar1	-1.78	(1.71)	0.05	(0.10)	-0.49	(1.13)	-0.16	(0.34)	0.01	(0.03)
Mar2	-2.64	(2.51)	1.01	(2.12)	-0.68	(1.45)	-0.22	(0.43)	-0.24	(0.51)
Mar3	-1.89	(1.84)	0.26	(0.60)	-0.06	(0.14)	-0.13	(0.29)	-0.15	(0.37)
School	-0.53	(2.73)	0.07	(0.49)	0.09	(0.60)	0.02	(0.11)	-0.12	(0.75)
Age1	-0.48	(1.35)	0.61	(2.19)	-0.14	(0.53)	0.00	(0.01)	-0.22	(0.82)
Age2	-0.70	(2.17)	0.51	(1.99)	-0.24	(0.99)	0.05	(0.18)	0.32	(1.30)
Age3	-0.84	(2.64)	0.52	(2.08)	-0.42	(1.74)	-0.03	(0.10)	0.67	(2.76)
Age4	-0.74	(2.28)	0.56	(2.18)	-0.71	(2.85)	0.08	(0.32)	1.14	(4.54)
Contact1	-0.20	(0.92)	0.44	(2.36)	0.50	(2.62)	0.12	(0.59)	0.36	(1.93)
Contact2	-0.10	(0.45)	0.29	(1.58)	0.45	(2.38)	-0.01	(0.06)	-0.11	(0.58)
Structure	-0.63	(2.56)	0.21	(1.01)	0.82	(3.81)	-0.26	(1.18)	0.29	(1.37)
Smoke	-0.08	(0.40)	-0.06	(0.34)	-0.10	(0.55)	0.16	(0.85)	-0.02	(0.12)
Constant	4.52	(4.09)	-1.00	(1.92)	-0.12	(0.25)	1.03	(1.87)	-0.41	(0.82)
Pseudo R^2	0.04		0.02		0.04		0.01		0.04	
p	0.0002		0.001		0.0001		0.87		0.0001	

Table 5 REDUCED FORM EQUATIONS - TYPE OF CARE

	Formal1		Formal2		Formal3	
Sex	0.22	(1.45)	-0.14	(0.86)	-0.10	(0.58)
Mar1	-0.28	(0.64)	0.21	(0.43)	-0.95	(2.12)
Mar2	-0.95	(2.03)	0.24	(0.48)	0.75	(1.63)
Mar3	-0.42	(1.01)	0.22	(0.48)	0.28	(0.65)
School	-0.06	(0.37)	-0.07	(0.45)	-0.09	(0.63)
Age1	-0.25	(0.92)	0.46	(1.55)	0.05	(0.15)
Age2	0.09	(0.38)	0.12	(0.44)	0.05	(0.15)
Age3	-0.17	(0.72)	0.42	(1.55)	0.32	(1.07)
Age4	-0.36	(1.45)	0.50	(1.81)	0.53	(1.78)
Contact1	0.14	(0.77)	-0.21	(1.10)	-	-
Contact2	-0.04	(0.25)	-0.02	(0.10)	-	-
Structure	0.34	(1.66)	-0.15	(0.67)	-	-
Smoke	0.08	(0.46)	-0.01	(0.05)	-0.60	(2.84)
Constant	0.31	(0.62)	-1.01	(1.83)	-0.92	(1.78)
Pseudo R^2	0.02		0.01		0.08	
p	0.003		0.61		0.0001	

Coding of variables was as follows: Sex ($X_{1,}$ =1 female, =0 male), Mar1 ($X_{2,}$ =1 married/cohabiting, =0 otherwise), Mar2 ($X_{3,}$ =1 single, =0 otherwise), Mar3 ($X_{4,}$ =1 widowed, =0 otherwise), School ($X_{5,}$ =1 if no higher education, =0 otherwise), Age1 ($X_{6,}$ =1 if 70-74 years, =0 otherwise), Age2 ($X_{7,}$ =1 if 75-79 years, =0 otherwise), Age3 ($X_{8,}$ =1 if 80-84 years, =0 otherwise), Age4 ($X_{9,}$ =1 if 85 years or over, =0 otherwise), Contact1 ($X_{10,}$ =1 if daily or 2 to 3 times per week, =0 otherwise), Contact2 ($X_{11,}$ =1 if weekly or monthly contact, =0 otherwise), Structure ($X_{12,}$ =1 if lives with other/s, =0 if lives alone), Smoke ($X_{13,}$ =1 if current smoker, =0 otherwise). The relevant missing categories for marital status, age and contact were divorced, age 65-69 years and little contact respectively.

Table 6 STRUCTURAL HEALTH EQUATIONS (OLS)

	Type I		Type II		Type III		Type IV		Type V	
Formal1*	-3.52	(4.99)	-5.54	(6.72)	1.10	(2.73)	-2.21	(3.94)	10.46	(15.25)
Formal2*	-2.87	(4.23)	-3.51	(4.44)	0.98	(2.52)	-1.63	(3.04)	7.39	(11.22)
Formal3*	-0.73	(5.19)	-0.77	(4.67)	0.33	(4.14)	-0.55	(4.87)	1.82	(13.22)
Sex	0.16	(6.01)	0.09	(2.97)	-0.05	(2.95)	0.05	(2.44)	-0.26	(10.09)
Mar1	-	-	-	-	-	-	-	-	-	-
Mar2	-0.36	(4.75)	-0.50	(5.59)	0.07	(1.62)	-0.20	(3.34)	1.02	(13.63)
Mar3	-	-	-	-	-	-	-	-	-	-
School	-0.15	(5.26)	-0.13	(3.74)	0.05	(2.75)	0.07	(2.84)	0.32	(11.11)
Age1	-	-	-	-	-	-	-	-	-	-
Age2	0.05	(1.51)	0.20	(5.40)	-0.05	(2.93)	0.06	(2.58)	-0.26	(8.57)
Age3	0.02	(0.50)	0.06	(1.82)	-0.08	(4.32)	0.06	(2.46)	-0.08	(2.59)
Age4	-0.10	(2.80)	-0.13	(3.10)	-0.08	(4.08)	0.01	(0.29)	0.28	(8.44)
Contact1	0.02	(0.92)	0.07	(2.15)	0.02	(1.15)	0.00	(0.21)	-0.11	(4.26)
Contact2	0.07	(2.46)	0.08	(2.22)	-0.02	(1.46)	0.05	(2.06)	-0.18	(6.42)
Structure	0.06	(1.22)	0.30	(5.43)	0.03	(0.98)	0.03	(0.88)	-0.41	(8.89)
Smoke	-	-	-	-	-	-	-	-	-	-
Constant	3.24	(5.61)	4.26	(6.33)	-0.84	(2.54)	1.99	(4.34)	-7.94	(14.17)
Adjusted R^2	0.08		0.07		0.08		0.03		0.29	
F	8.52		7.13		8.17		3.54		35.12	
Sig F	0.00		0.00		0.00		0.00		0.00	

* denotes IV estimates obtained from the reduced form equations in Table 5.

Table 7 STRUCTURAL DEMAND FOR CARE EQUATIONS

	Formal1		Formal2		Formal3	
Type I*	0.16	(0.07)	-1.64	(0.72)	13.74	(4.36)
Type II*	-3.98	(0.39)	-16.89	(1.59)	9.36	(4.31)
Type III*	2.80	(0.38)	11.33	(1.48)	-3.36	(1.83)
Type V*	2.81	(1.01)	2.26	(0.77)	-0.44	(0.27)
Sex	-0.12	(0.17)	-1.26	(1.70)	0.34	(0.99)
Mar1	0.07	(0.09)	1.38	(1.49)	1.52	(1.89)
Mar2	0.60	(0.18)	5.58	(1.60)	1.53	(1.25)
Mar3	-0.03	(0.04)	1.18	(1.47)	1.53	(2.09)
School	-	-	-	-	1.22	(3.82)
Age1	0.56	(0.32)	3.26	(1.78)	-	-
Age2	0.50	(0.38)	2.38	(1.71)	-	-
Age3	0.12	(0.08)	3.01	(1.94)	0.56	(1.58)
Age4	-0.15	(0.08)	3.77	(1.95)	0.12	(0.22)
Contact1	-	-	-	-	-	-
Contact2	-	-	-	-	-	-
Structure	-0.17	(0.14)	-1.77	(1.44)	-	-
Smoke	-	-	-	-	-0.22	(0.78)
Constant	-1.16	(0.50)	-.958	(0.39)	-17.63	(4.03)
Pseudo R^2	0.02		0.01		0.10	
p	0.003		0.43		0.0001	

* denotes IV estimates obtained from the reduced form equations in Table 4.

Table 8 RECURSIVE MODEL

	Formal1		Formal2		Formal3	
Type I	-1.09	(2.59)	1.25	(2.76)	-0.17	(0.36)
Type II	-1.86	(4.92)	0.86	(2.05)	0.27	(0.62)
Type III	-3.05	(5.10)	2.92	(4.73)	1.31	(2.22)
Type V	-2.87	(7.31)	2.49	(6.08)	2.27	(5.88)
Sex	0.14	(0.90)	-0.14	(0.85)	-0.10	(0.57)
Mar1	-0.41	(0.90)	0.36	(0.74)	-1.12	(2.42)
Mar2	-1.05	(2.16)	0.41	(0.80)	0.75	(1.57)
Mar3	-0.52	(1.21)	0.33	(0.70)	0.30	(0.67)
School	0.00	(0.01)	-0.09	(0.52)	-0.09	(0.61)
Age1	-0.28	(0.99)	0.61	(1.95)	0.18	(0.50)
Age2	0.11	(0.42)	0.23	(0.81)	0.05	(0.15)
Age3	-0.16	(0.63)	0.52	(1.83)	0.29	(0.91)
Age4	-0.29	(1.10)	0.54	(1.83)	0.27	(0.86)
Inform1	0.31	(1.63)	-0.39	(1.91)	-	-
Inform2	0.02	(0.10)	-0.09	(0.44)	-	-
Structure	0.65	(2.98)	-0.43	(1.84)	-	-
Smoke	0.05	(0.30)	0.03	(0.17)	-0.59	(2.65)
Constant	1.73	(2.89)	2.37	(3.64)	-1.57	(2.50)
Pseudo R^2	0.08		0.06		0.14	
p	0.0001		0.0001		0.0001	