

Multidimensional Assessment of Health Status in a Dependent Sample: An Exploratory Analysis for Adult Twins in China

Yan Ning,¹ Danan Gu,² Yonghua Hu,¹ Wenyan Ji,¹ Zengchang Pang,³ and Shaojie Wang³

¹ Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing 100191, China

² Center for the Study of Aging and Human Development, Duke University, Durham, NC 27705, USA

³ Qingdao Center for Disease Control and Prevention, Qingdao 266033, China

Health is a multidimensional and continual concept. Traditional latent analytic approaches have inherent deficits in capturing the complex nature of the concept; however, the Grade of Membership (GoM) model is well suited for this problem. We applied the GoM method to a set of 31 indicators to construct ideal profiles of health status based on physical, mental and social support items among 848 adult twins from Qingdao, China. Four profiles were identified: healthy individuals (pure type I), individuals with personality disorders (pure type II), individuals with mental impairments (pure type III) and individuals with physical impairments (pure type IV). The most frequently occurring combination in this population was profiles I, II, IV (14.74%), followed by profiles I, II, III, IV (13.44%), and then type I (11.08%). Only 13.56% of subjects fell completely into one single pure type, most individuals exhibited some of the characteristics of two or more pure types. Our results indicated that, compared to conventional statistical methods, the GoM model was more suited to capture the complex concept of health, reflecting its multidimensionality and continuity, while also exhibiting preferable reliability. This study also made an important contribution to research on GoM application in non-independent samples.

Keywords: twin study, health status, multidimensional assessment, Grade of Membership

Health is the most valuable resource that humans possess; an eternal theme to be pursued and explored. In 1948, the World Health Organization defined health as a 'state of complete physical, mental, and social wellbeing and not merely the absence of disease and infirmity'. Therefore, health can be viewed as a multidimensional concept that includes physical, mental and social dimensions, and that an extensive set of indicators is needed to capture this concept to its full extent. These indicators have distinct, albeit interrelated, contributions to health. Techniques to

reduce dimensionality are thus required as it is difficult to simultaneously handle all indicators. Numerous latent analytic approaches including cluster analysis, factor analysis, and discriminant analysis have been employed to handle the dimensionality reduction (Gold et al., 1990). These approaches, however, have inherent deficits in classification for individuals whose health conditions are complex and highly variable, with healthy and unhealthy elements coexisting (Xu, 1999). First, it is very likely in the real world that individuals may occupy more than one health states because the full spectrum of health ranges from wellbeing to death (Ware, 1987). Second, many health traits in analyses might be relatively rare and hence the distribution functions tend to be 'clumped' at the boundary, with summary measures generated from such rare traits also tending to be heavily concentrated near the boundary and with long, thin tails (Manton et al., 2004). Thus, it would be unwise to merely rely on those common latent statistical methods and models.

The Grade of Membership (GoM) method introduced by Woodbury and colleagues (Woodbury & Clive, 1974; Woodbury et al., 1978) is well suited to this problem. As a nonparametric classification methodology based on the theory of the fuzzy set, the GoM model is a general multivariate procedure for analyzing high dimensional discrete response data. The GoM technique was originally designed to generate medical classification systems (Woodbury & Manton, 1982) and has now been applied in studies for depressive symptoms and personality disorders (Cassidy et al., 2001; Nurnberg et al., 1999; Szádóczy et al., 2003), older adult health status (Deeg et al., 2002; Gu

Received 4 August, 2009; accepted 16 June, 2010.

Address for correspondence: Professor Yonghua Hu, Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, 38 Xue Yuan Road, Hai Dian District, Beijing 100191, China. E-mail: yhhu@bjmu.edu.cn

& Zeng, 2001, 2002; Jiang & Zhou, 1997; Lamb, 1996; Manton et al., 1997, 1998; McNamee, 2004; Portrait et al., 1999, 2001), genetic health studies (Corder et al., 2001; Manton et al., 2004) and other fields of medicine (e.g., Hughes et al., 1996; Maetzel et al., 2000; Woodbury & Fillenbaum, 1996). Szádóczy and colleagues (2003) employed the GoM analysis to describe lifetime patterns of depressive and anxiety symptoms reported by community respondents and primary care attenders. Six pure types provided the best description of the structure of the symptoms. McNamee (2004) used the GoM model to construct summary indicators of health and identified five main health types, which were then used to assess their association with costs of health and social care. Manton et al. (2004) found that the functional disability dimensions identified by GoM were more predictive of APOE polymorphism than specific diagnoses because they implicitly contained information on chronic conditions and severity of disease processes. Andreotti et al. (2009) analyzed World Health Survey data and identified three health profiles: Robust, Intermediate and Frail.

Yet, applications of the GoM method in dependent samples (e.g., twin data or other dyadic data) are still limited and in the exploration stage. So far, no widely acceptable method of dealing with sample dependence has been developed, although several options have been proposed (Gold et al., 1990; Manton & Land, 2000). In the present study, we apply the GoM method to a set of indicators from a rich database of the Chinese National Twin Registry (CNTR) to determine profiles of health status based upon physical, mental and social support items. Such an application is absent in China.

Methods

The GoM Model

Technical explanations of the GoM model have been described at length elsewhere (Berkman et al., 1989; Gold et al., 1990; Manton et al., 1992; Manton et al., 1994; Manton & Land, 2000); only a brief summary of the process is introduced here. The GoM model is developed to analyze data for multidimensional fuzzy states for I individuals with a set of J categorical variables, each of which is measured with L_j distinct outcomes or response levels (Manton & Land, 2000). Each response level (l) of a variable (x_{ij}) is recoded into a set of $J * L_j$ dichotomous variables denoted by y_{ijl} , where $y_{ijl} = 1$ if $x_{ij} = l$; and 0 otherwise. K latent profiles or states (known as ‘pure types’ in GoM terminology) of traits are identified in the GoM model from $J * L_j$ binary variables used in the analysis. This is similar to conventional latent class analyses, although the latent classes are determined by a set of variables either dichotomous or continuous.

However, unlike the conventional latent class analyses, the GoM model has a unique way of using the structure of the data in constructing dimensions or

states as ‘pure types’. Two types of parameters are estimated in the GoM model to define the characteristics of pure types and the probability (or degree) of each individual belonging to a given pure type. The first, λ_{kjl} , the GoM structural probability, which is subject to the constraints of $0 \leq \lambda_{kjl} \leq 1$ and

$$\sum_{l=1}^{L_j} \lambda_{kjl} = 1,$$

is the probability that the l th response ($l = 1, 2, 3, \dots, L_j$) of variable j ($j = 1, 2, 3, \dots, J$), or x_{ij} is observed for pure-type k ($k = 1, 2, 3, \dots, K$), or the probability of choosing response l of variable x_{ij} for a person completely from pure type k (not from other pure types, see g_{ik} below for more details). $\lambda_{kjl} = 1$ indicates that a person who is wholly from pure type k will 100 percent choose the l th response for a given variable x_{ij} ; $\lambda_{kjl} = 0$ indicates that a person who is completely from pure type k will definitely not choose the l th response for variable x_{ij} . In this regard, the main feature of a pure type is defined by λ_{kjl} .

The second is g_{ik} , which is the GoM score, relates the individual case to the latent profiles. g_{ik} represents the probability of individual i belongs to pure-type k or the degree to which individual i is a member of pure-type k . The g_{ik} score refers to individual’s grade of membership of pure type k . g_{ik} s are convexly constrained scores for individuals, that is, $0 \leq g_{ik} \leq 1$ and

$$\sum_{k=1}^K g_{ik} = 1.$$

$g_{ik} = 0$ indicates zero membership, that is, individual i does not belong to pure type k ; an individual with $g_{ik} = 1$ (complete membership) is exactly like the k th pure type or completely belongs to pure type k ; and partial membership ($0 < g_{ik} < 1$) is needed when the individual does not have all characteristics of a single pure type. When an individual has a complete membership of a pure type, he or she wholly belongs to that pure type only and cannot belong to other pure types. This is same as conventional classifications or other latent class analyses. However, for an individual whose g_{ik} is not equal to 1, he or she belongs to several pure types simultaneously with different probabilities or degrees. This is different from conventional classifications and traditional latent class analyses. The membership of multiple pure types better reflects the reality (Manton et al., 1994). Berkman et al. (1989) showed that an individual with $g_{ik} > 0.85$ still exhibits responses associated with a single profile of k .

Based on above definitions, the probability of $y_{ijl} = 1$ is defined by the following formula,

$$p_{ijl} = \text{Prob}(y_{ijl} = 1) = \sum_{k=1}^K g_{ik} \lambda_{kjl},$$

$$\text{where, } \sum_{k=1}^K g_{ik} = 1, \sum_{l=1}^{L_j} \lambda_{kjl} = 1.$$

For a given set of observations, the likelihood, L , is expressed as the product over i , j , and l of the set $\{\text{Prob}(y_{ijl} = 1)\}$

$$L = \prod_{i=1}^N \prod_{j=1}^J \prod_{l=1}^{L_j} \left(\sum_{k=1}^K g_{ik} \lambda_{kjl} \right)^{y_{ijl}}, \text{ or}$$

$$\ln L = \sum_{i=1}^N \sum_{j=1}^J \sum_{l=1}^{L_j} y_{ijl} \ln \left(\sum_{k=1}^K g_{ik} \lambda_{kjl} \right), \text{ where,}$$

$$\sum_{k=1}^K g_{ik} = 1, \quad \sum_{l=1}^{L_j} \lambda_{kjl} = 1.$$

λ_{kjl} s and g_{ik} s in the equation can be estimated by maximum likelihood following the procedure developed by Woodbury et al. (1994). The number of dimension (K) is determined by comparing the Akaike Information Criterion (AIC) for models with different K s, with the lowest value of the AIC designating the best model (Corder et al., 2001).

Variables in the GoM model are used either as internal or as external variables. Internal variables are those used to define pure types; λ_{kjl} for each internal variable can be compared to the corresponding marginal frequency (observed frequency of each response in the overall sample) to determine the attributes associated with each type. In the field of health studies, health measures are usually used as internal variables in the GoM model. External variables, such as demographic and personal characteristics, are not used in defining pure types; instead they are included to show their relationships to the pure types.

Several options for GoM analyses of non-independent data existed (Gold et al., 1990). First, one might analyze the dependent sets (e.g., twin 1 and twin 2, or a parent and a child, or different waves for a panel dataset) separately and compare the typological results. Alternatively, one might combine the two groups with no indication of which respondents are related. Finally, one might do a combined analysis with an additional internal variable indicating which respondents are related.

Several software packages for applying the GoM model are available; most are developed by the researchers themselves, and up to now no official release of commercial GoM software exists. DSIGOM beta version used in this paper is created by Decision Systems, Inc., and is available through its website (<http://www.dsisoft.com/index.html>) at no cost.

Subjects

Data for this study were collected as part of the Qingdao Twin Registry, which is a subset of the population-based twin registry, the Chinese National Twin Registry (CNTR) (Yang et al., 2002). All study protocols were approved by the Ethics Committee for Human Subjects Studies of Peking University Health Science Center and informed consent was obtained from subjects. Detailed data collection procedures can be found elsewhere (Huang et al., 2006; Ji et al.,

2006; Ji et al., 2008; Li et al., 2006). Thirty-one indicators used in the analysis covered three dimensions of health (i.e., the physical, psychological, and social dimensions). A total of 510 twin pairs were enrolled in Qingdao in 2002 for detailed phenotype studies, and 848 individuals participated in physical examinations (including anthropometrical measurement, biochemical assessment, etc.), psychological assessments, and social support assessments in 2004.

Physical Health Indicators

Physical health was jointly reflected by body mass index (BMI), waist circumference (WC), blood pressure, total cholesterol (TC), triglyceride (TG), high density lipoprotein (HDL), low density lipoprotein (LDL), fast plasma glucose (FPG) and fasting insulin. Except BMI, physical health indicators were divided into two discrete categories indicating normal and abnormal. A waist circumference of 85 cm or over for men and 80 cm or over for women was considered to represent central obesity. High blood pressure was defined by a systolic blood pressure of 140 mmHg or over and/or diastolic blood pressure of 90 mmHg or over. High TC, TG, LDL, FPG and low HDL were defined by cut-off points of 5.72, 1.70, 3.64, 6.10 and 0.91 mmol/L, respectively. People in the highest quartile of fasting insulin were regarded as high insulin level. BMI was divided into four categories indicating low weight (< 18.5), normal weight (18.5~23.9), overweight (24.0~27.9) and obesity (≥ 28.0). The cut-off values defining each category for each biological indicator were determined on the basis of clinically defined criteria in the Chinese population (Cooperative Meta-analysis Group of China Obesity Task Force, 2002; Dyslipidemia study group, 1997; Expert Panel on Metabolic Syndrome of Chinese Diabetes Society, 2004; Guidelines Revising Committee, 2006).

Mental Health Indicators

Mental health was indicated by psychiatric symptoms and personality disorders. Symptom Checklist 90 (SCL-90) was used to measure psychiatric symptoms. The SCL-90 is a self-report instrument comprised of 90 items. The SCL-90 items presumably cover nine different factors of psychological distress: somatization, obsession-compulsion, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Each of the nine subscale scores measures the respondents' severity of symptom distress. Subjects with a subscale score above the 75th percentile were considered suspected positive. To test the sensitivity, we explored its robustness by using alternative cut-point values at the 80th and 90th percentiles. The results were only altered slightly.

The Personality Diagnostic Questionnaire-4th edition (PDQ-4) was used to assess ten personality disorders (PD): paranoid, schizoid, schizotypal, histrionic, narcissistic, antisocial, borderline, avoidant, dependent, and obsessive-compulsive. The PDQ-4 is a self-administered questionnaire that yields personality diagnoses

consistent with the DSM-IV diagnostic criteria for axis II disorders. The Chinese version of the PDQ-4 is translated and culturally adapted from international PDQ-4, and it has good reliability and validity for assessing PDs in Chinese normal subjects (Huang et al., 2002). The subscale score is indicative of the presence of each PD respectively. The PDQ-4 integrated scoring key is followed to indicate clinically relevant PDs (Huang et al., 2002).

The SCL-90 measures psychological distress over a given period of time, usually between 7 and 14 days. As a reflection of respondents' self-perceived psychological status during the time reference, it is likely to be affected by many factors, and therefore only using the SCL-90 to assess mental health is inappropriate (Shan, 1998). While the PDQ-4 focuses on lifelong stable traits rather than on states during a short period of time, using the SCL-90 and PDQ-4 together could help evaluate mental health more accurately.

Social Function Measures

Social function measures were derived from three dimensions of social support with a 10-item instrument: (1) objective support (three items), referring to supports received from direct material aids and social network; (2) subjective support (four items), reflecting emotional and perceived supports; and (3) utilization of support (three items), referring to one's use of social network (Ji et al., 2008). The score for each dimension is obtained by summing the scores of the corresponding items. The validity of scale has been verified (for details of social support scale, see Ji et al., 2008). A score above the 75th percentile indicated a high level of social support. Sensitivity analyses using alternative cut-points at the 80th and 90th percentiles yielded a minor difference.

Statistical Analyses

Variables described above measure different domains of health; together, they likely encompass most aspects of health. However, inclusion of all these items in conventional statistical analyses is impractical. We therefore applied the GoM model, which is specifically designed to characterize the complex concept of health, to these 31 indicators to construct ideal profiles of health conditions of adult twins from Qingdao, China.

The sample is a dependent sample in that the observations of one twin are dependent of those of its co-twin due to genetic and environmental factors. Unlike previous studies that either overlooked interdependence between subjects in the study or mainly focused on simple approach to adjust the interdependence, we tried a new method, λ_{kjl} regression, to adjust the dependence between subjects. That is, g_{ik} s were estimated for subsample 2 conditional on the λ_{kjl} solution of subsample 1 for the same set of variables. In other words, we assumed that λ_{kjl} in subsample 1 was equal to λ_{kjl} in sub-sample 2. Specially, one member was randomly selected from each twin pair to generate subsample 1 ($n = 419$) using 'sample' command in

STATA statistical package (StataCorp, 2007), and the remaining members constituted subsample 2 ($n = 429$). As the sampling procedure is completely at random and only one member of each twin is selected each time, it is reasonable to consider that the two subsamples are chosen independently.

Two alternative designs can be selected in performing λ_{kjl} regression after randomly splitting each twin pair. First, GoM analysis is performed on sub-sample 1 and then g_{ik} s for sub-sample 2 are estimated using λ_{kjl} regression. The second one is just the reversal of the first, that is, GoM analysis is conducted on subsample 2 and then g_{ik} s are estimate for subsample 1. Because the two designs produced very similar results (data not shown), only results of the first one were reported here.

Results

Subjects consisted of 419 pairs and 10 individuals, ranging in age from 23 to 70 years (mean = 40.09, $SD = 9.09$); 37.38% were male. The demographic characteristics of subjects are summarized in Table 1. No statistically significant difference was found between subsample 1 and subsample 2.

GoM analyses were completed for three, four and five pure types, and AICs were calculated for each. AIC was lower for the model specifying 4 types (AIC = -2035.64) than either 3 (AIC = -1914.82) or 5 types (AIC = -1772.08) indicating that four pure types provide a model with the best fit to these data, that is, $K = 4$.

Subsets of characteristics that distinguish one pure type from another, and thus form the basis for verbal descriptions of the pure types, are identified using criteria developed by Berkman and colleagues (Berkman et al., 1989), which compares each λ_{kjl} to the corresponding marginal frequency. Specifically, a particular response will be defined as a distinguishing characteristic for a pure type if its estimated probability is at least twice the marginal frequency. For relatively prevalent conditions (assumed here to be responses with marginal probabilities greater than 0.4), the response will be considered to be distinguishing if the estimated probability is at least 35% greater than the marginal frequency. The profile probabilities (λ_{kjl} s) for each of the four pure types are presented in Table 2 together with the marginal frequency associated with each deleterious condition. The GoM classification of the subjects' health status can be described as follows.

Pure Type I: Healthy Individuals. This profile can be interpreted as healthy, i.e., individuals have no indications of physical impairments, mental impairments, or social ill-adaptation.

Pure Type II: Individuals With Personality Disorders. This type has no physical impairments or psychiatric symptoms, but he/she displays significant PDs. 100% probabilities of seven PDs (paranoid, histrionic, obsessive-compulsive, schizoid, narcissistic, avoidant, and

Table 1
Demographic Profile of Subjects

| Demographic variables | Subsample 1 | | Subsample 2 | | Total | |
|--------------------------|-------------------------------|-------|-------------|-------|----------|-------|
| | <i>N</i> | % | <i>N</i> | % | <i>N</i> | % |
| Sex | | | | | | |
| Male | 163 | 38.00 | 154 | 36.75 | 317 | 37.38 |
| Female | 266 | 62.00 | 265 | 63.25 | 531 | 62.62 |
| | <i>P</i> = 0.709 ^a | | | | | |
| Residence | | | | | | |
| City | 265 | 62.21 | 270 | 64.59 | 535 | 63.39 |
| Countryside | 161 | 37.79 | 148 | 35.41 | 309 | 36.61 |
| | <i>P</i> = 0.472 ^a | | | | | |
| Marital status | | | | | | |
| Single | 35 | 8.20 | 34 | 8.15 | 69 | 8.18 |
| Married | 392 | 91.80 | 383 | 91.85 | 775 | 91.82 |
| | <i>P</i> = 0.982 ^a | | | | | |
| Education | | | | | | |
| Illiterate or elementary | 32 | 7.51 | 26 | 6.22 | 58 | 6.87 |
| Junior high | 155 | 36.38 | 166 | 39.71 | 321 | 38.03 |
| Senior high | 141 | 33.10 | 138 | 33.01 | 279 | 33.06 |
| College or higher | 98 | 23.00 | 88 | 21.05 | 186 | 22.04 |
| | <i>P</i> = 0.684 ^a | | | | | |

Note: ^a*P*-value of the Chi-square test for heterogeneity of proportions between subsample 1 and subsample 2.

schizotypal) are found for pure type II and the probability of dependent PD is also very high (63.04%).

Pure Type III: Individuals With Mental Impairments.

This type is physically healthy but comprises nine psychiatric symptoms and some PDs. The probabilities of six PDs (paranoid, antisocial, obsessive-compulsive, avoidant, schizotypal, and borderline) are 75.73%, 5.59%, 100%, 100%, 59.78%, and 29.72% respectively for pure type III.

Pure Type IV: Individuals With Physical Impairments.

This type is characterized by the presence of physical impairments including being overweight, obesity, central obesity, high blood pressure, high TC, high TG, low HDL, high LDL, high FPG, and high INS.

Our GoM analysis also revealed that HDL, FPG, antisocial, objective support, subjective support and utilization of support had negligible contributions to defining the pure types given that their *H* values were small with values of 0.01, 0.04, 0.01, 0.01, 0.04 and 0.02, respectively, where *H* value measures how well the variable corresponds to the final pure type definition.

Table 3 reports the g_{ik} distribution across four pure types. Seven hundred and four individuals (83.02%) are found to have some characteristics of pure type I ($g_1 \neq 0$), of which 81 individuals (9.55%) are complete members ($g_1 = 1$). There are 416 individuals (49.06%) whose g_{ik} s for pure type III are estimated to be zero, whereas there are 61.20% of the subjects showing symptoms associated with pure type II and 58.73% for pure type IV.

Table 4 shows the distribution of GoM profiles by sex. Consistent with Berkman and colleagues

(Berkman et al., 1989), individuals with a GoM score of 0.9 or higher are defined as belonging to a single pure type, and individuals are defined to hold partial membership in two (or more) pure types if the corresponding two (or more) GoM scores sum to unity. Given this definition, most subjects (73.00%) share the characteristics of two or three pure types, and only 115 individuals (13.56%) are complete members of a single type, of which few are represented solely by pure type II (0.47%), III (0.94%) and IV (1.06%). The largest proportion of subjects (14.74%) are described by a combination of pure types I, II and IV, followed by a mix of pure types I, II, III and IV (13.44%), and pure type I as the third most prevalent class (11.08%).

Discussion

Multidimensional assessment of health status would facilitate the targeting of health resources based on health needs from a multidimensional perspective, and would also enable international and national comparisons, monitoring trends over time, and evaluation of the effect of interventions. The uniqueness of the current study lies in its application of the GoM model to construct profiles of health status based on a set of 31 indicators from physical, mental, and social health domains in adult twins from Qingdao, China. In the field of health studies, most traditional classification methods rely on the 'crisp set' perspective, which requires that individuals belong to one and only one category. In contrast, the GoM model generates nosological types and simultaneously quantifies the degree

Table 2
Probabilities of the Presence of Each Condition for Four GoM Profiles (%)

| <i>H</i> | Condition | Marginal frequency | Pure types | | | |
|------------------------------|-----------------------------|--------------------|------------|---------------|---------------|---------------|
| | | | I | II | III | IV |
| Physical impairments | | | | | | |
| 0.51 | Low weight | 4.44 | 6.81 | 0.00 | 14.51 | 0.00 |
| | Overweight | 30.61 | 0.00 | 43.81 | 0.00 | 73.79 |
| | Obesity | 8.41 | 0.00 | 0.00 | 0.00 | 26.21 |
| 0.51 | Central obesity | 29.98 | 0.00 | 0.00 | 0.00 | 100.00 |
| 0.47 | High blood pressure | 16.78 | 0.00 | 0.00 | 0.00 | 100.00 |
| 0.47 | High TC | 13.29 | 0.00 | 0.00 | 0.00 | 100.00 |
| 0.24 | High TG | 11.42 | 0.00 | 0.00 | 0.00 | 65.35 |
| 0.01 | Low HDL | 0.47 | 0.00 | 0.00 | 0.00 | 2.70 |
| 0.47 | High LDL | 13.99 | 0.00 | 0.00 | 0.00 | 99.65 |
| 0.04 | High FPG | 3.03 | 0.00 | 3.64 | 0.00 | 13.64 |
| 0.48 | High insulin | 23.29 | 0.00 | 0.00 | 0.00 | 100.00 |
| Psychiatric symptoms | | | | | | |
| 0.41 | Somatization | 20.98 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.40 | Obsession-compulsion | 18.88 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.40 | Interpersonal sensitivity | 18.18 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.42 | Depression | 22.38 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.41 | Anxiety | 19.81 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.41 | Hostility | 20.51 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.39 | Phobic anxiety | 16.32 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.39 | Paranoid ideation | 16.32 | 0.00 | 0.00 | 100.00 | 0.00 |
| 0.40 | Psychoticism | 18.65 | 0.00 | 0.00 | 100.00 | 0.00 |
| Personality disorders | | | | | | |
| 0.54 | Paranoid | 36.83 | 0.00 | 100.00 | 75.73 | 0.00 |
| 0.51 | Histrionic | 28.44 | 0.00 | 100.00 | 0.00 | 0.00 |
| 0.01 | Antisocial | 1.18 | 0.00 | 2.24 | 5.59 | 0.00 |
| 0.32 | Obsessive-Compulsive | 59.21 | 0.00 | 100.00 | 100.00 | 64.64 |
| 0.42 | Schizoid | 24.01 | 0.00 | 100.00 | 26.24 | 9.42 |
| 0.46 | Narcissistic | 19.35 | 0.00 | 100.00 | 10.70 | 0.00 |
| 0.63 | Avoidant | 36.60 | 0.00 | 100.00 | 100.00 | 0.00 |
| 0.48 | Schizotypal | 16.55 | 0.00 | 100.00 | 59.78 | 0.00 |
| 0.09 | Borderline | 6.53 | 0.00 | 13.72 | 29.72 | 0.00 |
| 0.23 | Dependent | 11.89 | 0.00 | 63.04 | 15.36 | 0.00 |
| High social support | | | | | | |
| 0.01 | High objective support | 18.65 | 21.72 | 11.74 | 21.02 | 15.36 |
| 0.04 | High subjective support | 16.32 | 20.56 | 26.51 | 0 | 6.72 |
| 0.02 | High utilization of support | 21.21 | 29.05 | 17.12 | 9.36 | 12.02 |

Note: Numbers in bold type indicate the conditions defining a profile.

to which an individual's features fit each type, allowing individuals to hold complete membership in a single category or partial membership in multiple categories (Manton et al., 1992). Importantly, in latent class model, as the number of variables increases, the posterior probabilities tend toward 1 for the correct class and 0 for all other classes; whereas in the GoM model, as more information is used in the analysis, the GoM scores are better estimated without convergence to the boundary values, 0 or 1 (Stallard, 2005). Therefore, the multidimensionality and continuity of health can be well embodied. The advantage of the GoM model lies in its ability to identify latent profiles of health status using information on health indicators from all dimensions; furthermore, based on the 'fuzzy

set' paradigm, it generalizes the traditional discrete taxonomy and treats health as a continuous variable, so it is a superior alternative to the conventional classification methodologies used to capture the complex nature of health in its full extent. However, similar to other 'crisp' conventional latent class methods, the GoM model is somewhat subjective: naming of pure types can be likened to artwork, requiring not only complicated objective calculations but also critical thoughts concerning specialty and research topic (Manton et al., 1994).

In our study four profiles were identified and characterized as healthy individuals (pure type I), individuals with personality disorders (pure type II), individuals with mental impairments (pure type III),

Table 3
Distribution of g_{ik} s Across Four Pure Types

| Range | Pure types | | | |
|-----------|--------------|--------------|--------------|--------------|
| | I | II | III | IV |
| 0.00 | 144 (16.98) | 329 (38.80) | 416 (49.06) | 350 (41.27) |
| 0.01-0.10 | 33 (3.89) | 68 (8.02) | 120 (14.15) | 32 (3.77) |
| 0.11-0.20 | 49 (5.78) | 148 (17.45) | 89 (10.50) | 122 (14.39) |
| 0.21-0.30 | 65 (7.67) | 74 (8.73) | 54 (6.37) | 97 (11.44) |
| 0.31-0.40 | 79 (9.32) | 90 (10.61) | 39 (4.60) | 102 (12.03) |
| 0.41-0.50 | 72 (8.49) | 61 (7.19) | 39 (4.60) | 49 (5.78) |
| 0.51-0.60 | 71 (8.37) | 30 (3.54) | 29 (3.42) | 52 (6.13) |
| 0.61-0.70 | 98 (11.56) | 21 (2.48) | 27 (3.18) | 14 (1.65) |
| 0.71-0.80 | 67 (7.90) | 15 (1.77) | 16 (1.89) | 12 (1.42) |
| 0.81-0.90 | 76 (8.96) | 8 (0.94) | 11 (1.30) | 9 (1.06) |
| 0.91-0.99 | 13 (1.53) | 2 (0.24) | 3 (0.35) | 4 (0.47) |
| 1.00 | 81 (9.55) | 2 (0.24) | 5 (0.59) | 5 (0.59) |
| Total | 848 (100.00) | 848 (100.00) | 848 (100.00) | 848 (100.00) |

Note: Data are shown as $N(\%)$.

and individuals with physical impairments (pure type IV). Less than 14% of the sample were classified as complete members of a single type and majority of them (11%) were subject to pure type I (i.e., healthy). The rest of the sample (more than 86%) exhibited some of the characteristics of two or more pure types

rather than being extreme representatives of one type, indicating an advantage of the GoM model over the traditional latent analytic methods.

We further found that indicators from the social domain had very small contributions to the identification of health profiles, suggesting that patterns of social support are similar among this sample. This may indicate that these variables might not be adequately defined or might not be very useful in this sample to capture heterogeneity within social domain of health. Literature has consistently shown that high or good social support may increase our sense of control over the environment, dampen physiological arousal, strengthen immune responses, promote healthy behavior, and buffer the negative effects of life events and chronic stressors (Fuhrer & Stansfeld, 2002). However, associations between social wellbeing and other health domains are normally less stronger than those across other domains (Sugisa et al., 1994; Unger et al., 1999; Zunzunegui et al., 2004), Berkman and colleagues (2000) presented a conceptual model which postulated a cascading causal process from social networks to emotional support through factors more proximate to individual health, including behavioral, psychological and physiological pathways, and to sense of wellbeing. Thus, it's no wonder that social indicators lose their significance when numerous indicators from physical or mental dimensions are present.

Table 4
Distribution of Pure Types

| Distribution class | Males | | Females | | Total | |
|----------------------------|----------|--------------|----------|--------------|----------|--------------|
| | <i>N</i> | % | <i>N</i> | % | <i>N</i> | % |
| Single ($g_k \geq 0.90$) | | | | | | |
| Pure type I | 26 | 8.20 | 68 | 12.81 | 94 | 11.08 |
| Pure type II | 3 | 0.95 | 1 | 0.19 | 4 | 0.47 |
| Pure type III | 2 | 0.63 | 6 | 1.13 | 8 | 0.94 |
| Pure type IV | 4 | 1.26 | 5 | 0.94 | 9 | 1.06 |
| Subtotal | 35 | 11.04 | 80 | 15.07 | 115 | 13.56 |
| Paired | | | | | | |
| I and II | 33 | 10.41 | 58 | 10.92 | 91 | 10.73 |
| I and III | 15 | 4.73 | 37 | 6.97 | 52 | 6.13 |
| I and IV | 35 | 11.04 | 53 | 9.98 | 88 | 10.38 |
| II and III | 10 | 3.15 | 10 | 1.88 | 20 | 2.36 |
| II and IV | 11 | 3.47 | 12 | 2.26 | 23 | 2.71 |
| III and IV | 10 | 3.15 | 9 | 1.69 | 19 | 2.24 |
| Subtotal | 114 | 35.96 | 179 | 33.71 | 293 | 34.55 |
| Triple | | | | | | |
| I, II and III | 31 | 9.78 | 51 | 9.60 | 82 | 9.67 |
| I, II and IV | 44 | 13.88 | 81 | 15.25 | 125 | 14.74 |
| I, III and IV | 18 | 5.68 | 40 | 7.53 | 58 | 6.84 |
| II, III and IV | 33 | 10.41 | 28 | 5.27 | 61 | 7.19 |
| Subtotal | 126 | 39.75 | 200 | 37.66 | 326 | 38.44 |
| Mixed | | | | | | |
| I, II, III and IV | 42 | 13.25 | 72 | 13.56 | 114 | 13.44 |
| Total | 317 | 100.00 | 531 | 100.00 | 848 | 100.00 |

Note: Numbers in bold type show the three most frequent classes.

Further research is clearly warranted to shed more light on this topic.

The uniqueness of the present study lies in our attempt to use λ_{kjl} regression aiming to correct for dependence between twin clusters. As aforementioned, to address inter-twin correlation one can use either separate sample analysis, whole sample analysis without dependence adjustment, or whole sample analysis considering inter-twin dependence by adding one additional variable in the analysis to indicate relating respondents. The separate sample approach normally produces two different sets of 'pure types' for two sub-samples and their probabilities will be also different. Accordingly, their results cannot be compared or combined. The whole sample analysis simply overlooked inter-twin correlations. We ever considered using the third method. Yet, because every variable (either indicator variable, external variable, or internal variable) in the GoM model must be categorical and the maximum number of categories should not exceed twenty in the current version of the DSIGoM package, it is impossible for us to use a variable (i.e., 'twin id') to indicate relating subjects given that the number of twin pairs is 419 in our dataset. Other packages also likely suffer from this constraint. In this regard, the GoM model is unable to deal with large pair data using the third approach. This is the major reason why most previous studies using the GoM model did not adjust inter-subject dependence. For example, Manton and Land (2000) combined four waves of longitudinal datasets and only included a time variable in the GoM model and assumed a fixed λ_{kjl} over time. Their approach is indeed a whole sample analysis without taking intersubject dependence into consideration as the time variable could not capture the intersubject correlation. In other words, λ_{kjl} regression, which assumes a fixed λ_{kjl} within each twin pair, might be a practical approach in the GoM model for better addressing inter-twin dependence.

We randomly divided twin pairs into two samples of unrelated individuals to satisfy the assumption of independent observations, and the two alternative designs produced very close results. The results based on the whole sample that did not adjust inter-twin correlation are somewhat different from these results. Additional analytical results based on the conventional factor analysis further confirmed the importance of physical and mental domains and unimportance of social domain indicators in our sample. All together, we are confident about the stability and reliability of the new method. Compared to previous methods, our new method extends previous simple approaches for adjustment of intersubject dependence by removing the structure of dependency using λ_{kjl} regression. Indeed, several previous approaches did not fully take interdependence into consideration (Gold et al., 1990; Manton & Land, 2000). In this regard, the present study makes an important contribution to research on the GoM application in

non-independent samples. Nevertheless, more efforts are certainly needed to further address this issue in the GoM model.

Several limitations deserve additional attention. First, although the GoM scores of each individual on health profiles are superior to outcomes based on conventional methods, the results of the present study are not readily comparable with other GoM analyses of health status given that the typology obtained strongly relies on the choice of indicators and measurement instruments, and that not all studies use the same instruments, nor do these instruments capture similar aspects of the health concept. Second, some indicator values in this analysis were divided into two categories designed to distinguish 'high' values from those within the 'normal' range based on the 75th percentile. The choice of these cut-off points may not be accurate, but it might be reasonable in the absence of clinically defined or otherwise substantively meaningful criteria and can approximately reflect the distributions of these indicators in the sample. Importantly, other cut-off points produce similar results. Third, due to space limitations, we were unable to explore the associations between socioeconomic characteristics (external variables) and pure types. We also did not apply the approach used by previous studies (McNamee, 2004; Portrait et al., 1999, 2001; Seplaki et al., 2004) where the individual degrees of involvement in the different health dimensions (g_{ik} s) obtained from the GoM model were used in other multivariate analyses. We will leave analyses on such associations for future studies. Finally, we used λ_{kjl} regression to address inter-twin dependence. Although it is novel, its practical value still needs further confirmation.

Acknowledgments

This study was supported in part by the China Medical Board (01-746). We sincerely thank Prof. Yueqin Huang from Institute of Mental Health, Peking University for her enthusiastic help on mental health assessment, and Prof. Liming Li, Prof. Weihua Cao, Prof. Jun Lv and Prof. Ying Qin from Peking University Health Science Center for their helpful commitments on project planning and data collecting. We gratefully acknowledge the valuable comments and suggestions from anonymous reviewers and from Prof. Dafang Chen at Peking University Health Science Center. Special thanks to all the participants for their participation in this study.

References

- Andreotti, A., Minicuci, N., Kowal, P., & Chatterji, S. (2009). Multidimensional Profiles of Health Status: An Application of the Grade of Membership Model to the World Health Survey. *PLoS ONE*, 4, e4426. doi: 10.1371/journal.pone.0004426
- Berkman, L., Singer, B., & Manton, K. (1989). Black/white differences in health status and mortality among the elderly. *Demography*, 26, 661–678.

- Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, *51*, 843–857.
- Cassidy, F., Pieper, C. F., & Carroll, B. J. (2001). Subtypes of Mania Determined by Grade of Membership Analysis. *Neuropsychopharmacology*, *25*, 373–383.
- Cooperative Meta-analysis Group of China Obesity Task Force. (2002). Predictive values of body mass index and waist circumference to risk factors of related diseases in Chinese adult population. *Chinese Journal of Epidemiology*, *23*, 5–10 [in Chinese].
- Corder, E. H., Woodbury, M. A., Manton, K. G., & Field, L. L. (2001). Grade-of-membership sibpair linkage analysis maps IDDM11 to chromosome 14q24.3-q31. *Annals of Human Genetics*, *65*, 387–394.
- Deeg, D. J., Portrait, F., & Lindeboom, M. (2002). Health profiles and profile-specific health expectancies of older women and men: The Netherlands. *Journal of Women & Aging*, *14*, 27–46.
- Dyslipidemia study group. (1997). Recommendations on prevention and treatment of dyslipidemia in China. *Chinese Journal of Cardiology*, *25*, 169–173 [in Chinese].
- Expert Panel on Metabolic Syndrome of Chinese Diabetes Society. (2004). Recommendations on Metabolic Syndrome of Chinese Diabetes Society. *Chinese Journal of Diabetes*, *12*, 156–161 [in Chinese].
- Fuhrer, R., & Stansfeld, S. A. (2002). How gender affects patterns of social relations and their impact on health: A comparison of one or multiple sources of support from 'close persons'. *Social Science & Medicine*, *54*, 811–825.
- Gold, D. T., Woodbury, M. A., & George, L. K. (1990). Relationship classification using grade of membership analysis: A typology of sibling relationships in later life. *Journal of Gerontology*, *45*, S43–51.
- Gu, D., & Zeng, Y. (2001). Analysis of the Grade of Membership of health status of the Chinese oldest old. *Chinese Journal of Population Science, Suppl.*, 8–14. [in Chinese]
- Gu, D., & Zeng, Y. (2002). Healthy life expectancy of the oldest old in China. *Population & Economics*, *2*, 9–15. [in Chinese]
- Guidelines Revising Committee. (2006). Guidelines for the prevention and treatment of hypertension in China. *Chinese Journal of Stroke*, *1*, 575–582 [in Chinese].
- Huang, A. Q., Hu, Y. H., Zhan, S. Y., Xu, B., Pang, Z. C., Cao, W. H., Lu, J., Qin, Y., & Lee, L. M. (2006). Lipoprotein lipase gene S447X polymorphism modulates the relation between central obesity and serum lipids, a twin study. *International Journal of Obesity*, *30*, 1693–1701.
- Huang, Y. Q., Liu, B. H., Liu, Z. R., Zhang, G. Z., & Zhang, H. M. (2002). A cross-sectional study of personality dysfunction among students of first-grade senior high schools in Beijing. *Chinese Journal of Epidemiology*, *23*, 338–340 [in Chinese].
- Hughes, D. C., Fillenbaum, G. G., Woodbury, M. A., & Siegler, I. C. (1996). Psychometric characteristics of the mini-mental state examination in a community population — A grade of membership analysis: Part I. *International Journal of Geriatric Psychiatry*, *11*, 439–455.
- Ji, W. Y., Hu, Y. H., Huang, Y. Q., Cao, W. H., Lu, J., Qin, Y., Pang, Z. C., Wang, S. J., & Li, L. M. (2006). A twin study of personality disorder heritability. *Chinese Journal of Epidemiology*, *27*, 137–141 [in Chinese].
- Ji, W. Y., Hu, Y. H., Huang, Y. Q., Cao, W. H., Lu, J., Qin, Y., Peng, Z. C., Wang, S. J., & Lee, L. M. (2008). A genetic epidemiologic study of social support in a Chinese sample. *Twin Research and Human Genetics*, *11*, 55–62.
- Jiang, J. M., & Zhou, Y. S. (1997). Sectional estimate of the active life expectancy. *Chinese Journal of Health Statistics*, *14*, 21–25 [in Chinese].
- Lamb, V. L. (1996). A cross-national study of quality of life factors associated with patterns of elderly disablement. *Social Science & Medicine*, *42*, 363–377.
- Li, L., Gao, W., Lv, J., Cao, W., Zhan, S., Yang, H., & Hu, Y. (2006). Current status of the Chinese National Twin Registry. *Twin Research and Human Genetics*, *9*, 747–752.
- Maetzel, A., Johnson, S. H., Woodbury, M., & Bombardier, C. (2000). Use of grade membership analysis to profile the practice styles of individual physicians in the management of acute low back pain. *Journal of Clinical Epidemiology*, *53*, 195–205.
- Manton, K. G., Gu, X., Huang, H., & Kovtun, M. (2004). Fuzzy set analyses of genetic determinants of health and disability status. *Statistical Methods in Medical Research*, *13*, 395–408.
- Manton, K. G., & Land, K. C. (2000). Multidimensional disability/mortality trajectories at ages 65 and over: The impact of state dependence. *Social Indicator Research*, *51*, 193–221.
- Manton, K. G., Stallard, E., & Corder, L. S. (1997). Changes in the age dependence of mortality and disability: cohort and other determinants. *Demography*, *34*, 135–157.
- Manton, K. G., Stallard, E., & Corder, L. S. (1998). The dynamics of dimensions of age-related disability 1982 to 1994 in the U.S. elderly population. *The Journals of Gerontology: Biological Sciences and Medical Sciences*, *53*, B59–70.
- Manton, K. G., Woodbury, M. A., Stallard, E., & Corder, L. S. (1992). The use of Grade-of-Membership techniques to estimate regression relationships. *Sociological Methodology*, *22*, 321–381.
- Manton, K. G., Woodbury, M. A., & Tolley, H. D. (1994). Statistical applications using fuzzy sets. New York: John Wiley & Sons, Inc.

- McNamee, P. (2004). A comparison of the grade of membership measure with alternative health indicators in explaining costs for older people. *Health Economics*, 13, 379–395.
- Nurnberg, H. G., Woodbury, M. A., & Bogenschutz, M. P. (1999). A mathematical typology analysis of DSM-III-R Personality Disorder Classification: Grade of membership technique. *Comprehensive Psychiatry*, 40, 61–71.
- Portrait, F., Lindeboom, M., & Deeg, D. (1999). Health and mortality of the elderly: The grade of membership method, classification and determination. *Health Economics*, 8, 441–457.
- Portrait, F., Lindeboom, M., & Deeg, D. (2001). Life expectancies in specific health states: Results from a joint model of health status and mortality of older persons. *Demography*, 38, 525–536.
- Seplaki, C. L., Goldman, N., Weinstein, M., & Lin, Y. H. (2004). How are biomarkers related to physical and mental well-being? *The Journals of Gerontology. Series A: Biological Sciences and Medical Sciences*, 59, 201–217.
- Shan, M. H. (1998). Using SCL-90 and 16PF to assess mental health status correctly. *Chinese Mental Health Journal*, 12, 81–82 [in Chinese].
- Stallard, E. (2005). Trajectories of disability and mortality among the U.S. elderly population: Evidence from the 1984–1999 NLTC. *Living to 100 and Beyond Symposium*. Orlando, FL: the Society of Actuaries.
- StataCorp (2007). STATA 10.0. College Station, TX [in Chinese].
- Sugisawa, H., Liang, J., & Liu, X. (1994). Social networks, social support, and mortality among older people in Japan. *Journal of Gerontology: Social Sciences*, 49, S3–S13.
- Szádóczky, E., Rózsa, S., Pattern, S., Arató, M., & Füredi, J. (2003). Lifetime patterns of depressive symptoms in the community and among primary care attenders: An application of grade of membership analysis. *Journal of Affective Disorders*, 77, 31–39.
- Unger, J. B. McAvay, G., Bruce, M. L., Berman, L., & Seeman, T. (1999). Variation in the impact of social network characteristics on physical functioning in elderly persons: MacArthur Studies of Successful Aging. *Journal of Gerontology: Social Sciences*, 54B, S245–S251.
- Ware, J. E. Jr. (1987). Standards for validating health measures: Definition and content. *Journal of Chronic Diseases*, 40, 473–480.
- Woodbury, M. A., & Clive, J. (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics*, 4, 111–121.
- Woodbury, M. A., Clive, J., & Garson, A. (1978). Mathematical typology: A grade of membership technique for obtaining disease definition. *Computers and Biomedical Research*, 11, 277–298.
- Woodbury, M. A., & Fillenbaum, G. G. (1996). Psychometric characteristics of the mini-mental state examination in patients with Alzheimer's disease — A grade of membership analysis of cerad data: Part II. *International Journal of Geriatric Psychiatry*, 11, 543–553.
- Woodbury, M. A., & Manton, K. G. (1982). A new procedure for analysis of medical classification. *Methods of Information in Medicine*, 21, 210–220.
- Woodbury, M. A., Manton, K. G., & Tolley, H. D. (1994). A general model for statistical analysis using fuzzy sets: Sufficient conditions for identifiability and statistical properties. *Information Sciences*, 1, 149–180.
- Xu, J. (1999). Health status assessment. *Foreign Medical Sciences (Section of Social Medicine)*, 16, 1–3.
- Yang, H., Li, X., Cao, W., Lu, J., Wang, T., Zhan, S., Hu, Y., & Li, L. (2002). Chinese National Twin Registry as a resource for genetic epidemiologic studies of common and complex diseases in China. *Twin Research*, 5, 347–351.
- Zunzunegui, M. V., Koné, A., Johri, M., Béland, F., Wolfson, C., & Bergman, H. (2004). Social networks and self-rated health in two French-speaking Canadian community dwelling populations over 65. *Social Science & Medicine*, 58, 2069–2081.