Lifestyle factors, dietary quality and health: Econometric evidence from US micro data

*Tetyana DEMYDAS**

* **Corresponding author**: Justus Liebig University Giessen, Institute of Agricultural Policy and Market Research, D-35390 Giesse, Germany
e-mail: *Tetyana.Demydas@agrar.uni-giessen.de*

Summary – *The study applies household production theory to investigate how socioeconomic characteristics and behavioural choices affect health status of an individual (here related to cardiovascular diseases). Structural equation modelling (SEM) is used as an estimation approach, which allows to model health inputs as endogenous variables, to test the hypothesized linkages among them (e.g., between smoking and weight) and to investigate their contribution to a production of a particular health state. Moreover, direct effects of exogenous personal characteristics on health can be disentangled from their indirect impact via lifestyle choices. To improve its measurement properties, health status is modelled via SEM as a latent variable. The results indicate a multidimensionality of cardiovascular health, which is more adequately represented by a two-factor latent variable model rather than by a single-factor model. Moreover, the specified two health dimensions (hypertension and lipids) are (partly) differently explained by the model variables. While higher weight negatively and significantly impacts both health dimensions, a diet rich in F&V (Fruit & Vegetable) has stronger contribution to a lower hypertension risk. At the same time, higher educational level showed to be rather related to lower measurements of blood fat. A special focus in the model is given to the (in)direct impact of education and income. Both variables showed to be related to less frequent smoking and engaging in sedentary leisure time activities. However, their impact on diet was rather indirect through their positive contribution to nutrition knowledge.*

Keywords: *Cardiovascular diseases, Health production, Indirect effect, Latent variable, Structural equation modelling.*

Mode de vie, qualité des aliments et santé : une analyse économétrique à partir de données individuelles américaines

Résumé – Cette étude s'appuie sur la théorie de la « production domestique » afin de déterminer comment les caractéristiques socio-économiques et les choix comportementaux affectent l'état de santé d'un individu (ici relatif aux maladies cardio-vasculaires). Le Modèle d'équation structurelle (SEM), utilisé afin d'établir une estimation, permet de modéliser les données médicales en variables endogènes dans le but d'évaluer les liens hypothétiques entre ces variables (par exemple entre le poids et l'usage de la cigarette) et d'étudier leur contribution au développement d' un état de santé particulier. De plus, les effets directs des caractéristiques personnelles exogènes sur la santé peuvent être différentiés de leur impact indirect provenant des choix de styles de vie. Afin d'améliorer les caractéristiques de mesures, l'état de santé est modélisé par une variable latente dans le SEM.

Les résultats indiquent une multi-dimensionnalité de l'état de santé cardio-vasculaire, qui est représenté de manière plus adéquate par un modèle de variable latente à deux facteurs plutôt qu'à un modèle à facteur unique. De plus, les deux dimensions spécifiques de santé (hypertension et taux de lipides) sont (partiellement) différemment expliquées par les variables du modèle. Alors qu'un poids élevé a un impact négatif important sur les dimensions de santé, un régime riche en FV (fruits & légumes) diminue le risque d'hypertension. D'autre part, un niveau d'études supérieur s'est révélé être en corrélation avec des taux de graisses dans le sang faibles. Une attention particulière a été portée dans le modèle sur l'impact (in)direct de l'éducation et du revenu. Les deux variables se sont révélées être liées à un tabagisme moins fréquent et à un temps consacré à des loisirs « sédentaires » plus réduit. Cependant, il n'y a pas d'impact direct sur le régime alimentaire. Celui-ci est plutôt lié à l'effet positif de ces deux variables sur les connaissances en nutrition.

Mots-clés : maladies cardio-vasculaires, santé, effet indirect, variable latente, modélisation structurelle équation

JEL Classification: I12, C39 and D13

1. Introduction

According to the World Health Organisation (WHO) estimates, Cardiovascular diseases (CVD) referring to stroke, heart failure, hypertension, diseases of the arteries and other related disorders, accounted for 30% of all causes of deaths in 2005 and will remain the leading cause of disability and premature death globally (WHO, 2007). CVD is the main cause of death in the USA, where one out of three adults has some form of these diseases. CVD are connected with high personal losses and a significant economic burden on the health care system. The estimated total cost of CVD in the United States was $403.1 billion in 2006. By comparison, the costs of all cancers accounted to 190 billion in 2004 and costs related to HIV infections was about $28.9 billion in 1999 (American Heart Association, 2006).

A high incidence of these negative health outcomes in the USA and worldwide are attributable to unhealthy behaviour flourishing among the populations, such as physical inactivity, smoking, and poor diets, along with the growing overweight and obesity rates (Lloyd-Jones *et al.*, 2010). In spite of a role of heredity, the modifiable behaviour is believed to be responsible for the majority of CVD cases (Public Health Agency of Canada, 2009). The existing informational and educational campaigns aimed to promote healthier lifestyles and thus to lower the health risks seem to have limited effects. Although smoking rates have been reduced considerably in the last decades, about 21% of the US population over 18 years are still characterized as current smokers. Only one third of the adults engage in a regular leisure-time physical activity (American Heart Association, 2010). Furthermore, the majority of Americans fail to choose a healthy diet. According to the USDA's Healthy Eating Index (HEI), the diet of about 90% of the population is poor or needs improvement (Basiotis *et al.*, 2002). An apparent problem is overweight and obesity, which rates reached 34% and 68% among the Americans of above 20 years, respectively (Flegal *et al.*, 2010).

These negative trends and their high societal costs induced an active research in the field. Health outcomes have been modelled in their relation to a number of factors, such as medical services utilisation (Auster *et al.*, 1969), socioeconomic status (Lairson *et al.*, 1984), economic determinants, e.g., taxes (Chou *et al.*, 2004; Schroeter and Lusk, 2008). Some authors have focused on a particular individual behaviour such as smoking and exercising (Blaylock and Blisard, 1992; Kenkel, 1995). In this respect, an important problem discussed in the economic literature is the endogeneity of health and behavioural variables and its consequences (Rosenzweig and Schultz, 1983; Behrmann and Deolalikar, 1988). It is argued that this aspect has not received a necessary attention in the empirical studies (in particular in epidemiological research) that might have led to biased estimates of the effects (Contoyannis and Jones, 2004; Chen *et al.*, 2002; Briscoe *et al.*, 1990).

The present study employs a health production framework to investigate the relationship between health state (related to CVD) and its determinants.

It contributes to the existing literature in several dimensions. First, an interdisciplinary approach allows an analysis of diverse health factors and their complex interrelations. Benefits of such an approach have been emphasised in the literature (Berman *et al*., 1994; Nayga, 2008). Second, the study treats health-related behaviours as endogenous in the model. The empirical analysis employs the Structural Equation Modelling (SEM) technique, that allows to specify a comprehensive model and to test the relationships among its components simultaneously, i.e., health inputs choices and health production technology, providing estimates of direct and indirect linkages. SEM is seen as a result of a merger of multi-equation regression models from econometrics and measurement models from psychology (Hair, 1995). Third, a special attention during model specification is given to the measurement of health state. In contrast to the previous studies, which usually use a single health indicator (most often a self-assessed health status measure), the present study employs a latent variable approach that permits to represent an unobservable health in terms of a number of observed indicators. This feature of SEM offers an advantage in terms of the improved measurement properties that come from multiple measures of the same construct (Kline, 1998). Moreover, the relationships between several unobserved theoretical constructs can be estimated. Finally, the representative US National Health and Nutrition Examination Survey (NHANES) 2005 to 2006 provided the data for this analysis. This survey delivers diverse information on socioeconomic characteristics of the respondents, detailed 2-day dietary records as well as medical information obtained from blood examination that facilitates the analysis of the various factors affecting a person's health and their interrelations.

The main results of the study can be summarised as follows. Cardiovascular health showed to be a multidimensional concept, being more adequately represented by a two-factor latent variable model compared to a single-factor model. Moreover, the specified two dimensions (hypertension and lipids) are (partly) differently explained by the model variables, e.g., better diet is related to a lower blood pressure but not to the lipid blood measurements. On the other hand, smoking and less physical activity showed to be stronger related to lipids risk dimension. Further, higher education and income are directly related to such healthier behaviours as less smoking and more active leisure time activities, but not to a better dietary quality. However, the findings suggest an indirect positive impact of these variables on the diet through their contribution to nutrition knowledge of an individual. Higher income is indirectly associated with lower weight, but this was due to less sedentary activities during leisure time rather than to a healthier diet.

In the following section, a review of the relevant literature on the linkages between health state and its major determinants is given with a focus on the difficulties connected with empirical estimation. The third section derives the theoretical model for health production. The fourth section describes the data set, introduces the variables of the model and main hypotheses and gives a short overview of the estimation strategy. Empirical results are presented in the fifth section. The final section concludes with a discussion of study limitations and extensions for future researches.

2. Empirical considerations about health and its determinants

In economic literature health state and health-related behavioural variables are both seen as "endogenous", i.e., determined by the forces within the model. The problem of endogeneity of health factors can be outlined by the words of Rosenzweig and Schulz from their seminal paper in the health production literature (1983, p. 723): "Estimates of health technology must be obtained from a behavioural model in which health inputs are themselves choices". These choices are dependent on many factors among which are prices, income, individual characteristics and preferences. Further complexity arises due to a) potential interrelations among health inputs and b) because health outcome might not only be affected by health choices but, conversely, a choice of a particular behaviour might be a response to the actual (perceived) health status (reverse causality). Accordingly, modelling only simple unidirectional causation may lead to biases (Rosenzweig and Schultz, 1983; Behrmann and Deolalikar, 1988). A consistent estimation of model parameters needs a more complex modelling technique.

A commonly used empirical approach aimed to account for endogeneity of health inputs is Instrumental Variable (IV) method and its special case Two-stage-least square (2SLS) technique (Rashad, 2006; Variyam *et al.*, 1998). 2SLS involves an estimation of reduced-form equations for the possibly endogenous causal variables by regressing them on a number of exogenous variables called instruments. In the next stage, the predicted values from these regressions are substituted for the "problematic" input variables in the main regression equation.

Chen *et al.* (2002) used IV method and showed empirically that accounting for endogeneity of health inputs can lead to changes in the direction and intensity of postulated relations. Blood pressure is modelled in their study to be dependent on person's nutrients intake, physical activity and medication usage. When endogeneity of these inputs is controlled (the choices are modelled as dependent on prices, wages and income), the effect of sodium intake on blood pressure turned negative, which, according to the authors, is in line with biomedical evidence on this relation. Similarly, Contoyannis and Jones (2004) applied the framework of household production to investigate the impact of the selected lifestyles (sleeping well, exercising, and not smoking) on the health self-evaluation on the basis of the British panel data and demonstrated an increase of the effects of these variables on health when accounting for the inputs' endogeneity. IV method shows to be helpful in obtaining consistent estimates in health economic studies, but it is largely dependent on the availability of appropriate instruments and their quality. Thus, Kenkel (1995), who aimed to estimate the impact of a number of behaviours on adults' health, evaluates the results from his two-stage model as implausible. He attributes the failure to account for the endogeneity in this study partly to the lack of explanatory power of the instruments such as money prices and argues that they might be irrelevant for many behavioural choices.

SEM presents another method of simultaneous equation modelling that allow estimation of complex interrelations. It is a full information method offering a simultaneous estimation of all model parameters (Kennedy, 2003). SEM has a number of strength, e.g., a convenient way to specify and test direct and indirect effects, feedback relations and to employ latent variables. These features are especially important for this study. Empirical examples of SEM can be found in, e.g., Wolfe and Behrman (1984), Behrman and Wolfe (1987), Erbsland *et al*. (1995), Giuffrida *et al*. (2005), Mazzocchi and Traill (2008), Van der Gaag and Wolfe (1991), Häkkinen (1991) and Wagstaff (1993).

Behrman and Wolfe (1987) estimate health production functions for maternal and child health in Nicaragua. Health status and inputs such as nutrition, medical care and community endowments are specified as latent constructs. The authors present several alternative models and come to a conclusion that a strong positive effect of mother's education on health and nutrition may be overestimated when health endowments (e.g., health knowledge, abilities, habits, heath state in childhood) are not accounted for.

Connell *et al*. (2001) specify a comprehensive model aimed to investigate the linkages among personal socio-demographic and economic characteristics, food insufficiency, diet quality, and other health behaviours and the CVD outcomes in a sample of adults from the NHANES III. However, the authors indicate a low fit of the hypothesized model. They stress an importance of choosing right indicators for latent variables (measurement model of CVD risk is one of them) and suggest a need to analyse the direct and indirect effects of CVD determinants.

Of a special interest is the study of Mazzocchi and Traill (2008) and Mazzocchi *et al*. (2009). They employ household production theory as a theoretical background and SEM as an estimation approach to test relationships among wealth, nutrition, weight and health based on the data of the UK National Diet and Nutrition Survey from 2000-2001. Among other findings, the authors discuss a direct positive effect of better diet, higher exercises and lower weight on health and indirect effect of wealth on health due to healthier diet. Further numerous examples of modelling of health and health related behaviours with SEM approach can be also found in medical, epidemiological, sociological and psychological research (e.g., Cobas *et al*., 1996; Oh and Seo, 2001; Shen and Takeuchi, 2001, McAlister *et al*., 1984; Hays *et al*., 2005).

An important challenge of consistent estimation of health production models is the absence of omitted variables and unobserved characteristics.

Thus, presence of unobserved heterogeneity in a model (e.g., individual's time preferences, mental abilities) may cause biases in the estimated relations. A special attention in the literature is given to the linkages between education and health and a potential bias in this relation due to unobservables (Grossman, 2004; Kawachi *et al.*, 2010; Fuchs, 1982 and 2004; Leigh, 1983). Further, a role of nutritional knowledge in the relationship between education and health behaviours and outcomes has been in a focus of the research (Nayga, 2000, Variyam *et al.*, 1998). *E.g.*, Variyam *et al.* (1998) argue that a positive effect of education (and income) on dietary choices may be explained by a positive contribution of these variables to person's nutrition knowledge and stress a need to model the latter as endogenous variable.

Another challenge is due to potential reverse causality among variables in health production model. Thus, a person, who is already in a poor health, may tend to less healthy dietary and lifestyle choices. Rashad (2006, p. 278), who tested an impact of caloric intake on BMI (Body Mass Index), stresses that "[...] caloric intake not only influences BMI but is also likely to be influenced by BMI, especially if caloric intake is habituating". Grossman (2004) discusses that the relation between health and education can also have a reverse direction with better health state causing more schooling. Reverse causality can be tested in frames of SEM. However, such interrelations are very complex and empirical estimation is usually confronted with identification problem and a need in incorporation of adequate instruments for each variable in reciprocal relation. Moreover, estimation of reciprocal causality should be performed using rather longitudinal data.

Finally, an important aspect to be accounted for is an adequate representation of health status in the model. Due to the fact that health itself is not directly observable, a number of indicators has been developed and employed that propose its rough description. They range from subjective health assessment (self-reports) (Denton and Walters, 1999; Blaylock and Blisard, 1992) to more objective measures such as country mortality rates (Or, 2000), and clinical health measures, e.g., blood pressure (Kenkel, 1995; Chen *et al.*, 2002). A person's BMI is also employed as an indicator of health state (Loureiro and Nayga, 2005; Rashad, 2006; Behrman *et al.*, 1988). However, these measures can represent an individual's health only approximately and measurement error may be high due to, e.g., errors in self-reports as these are believed to correlate with respondent's education, culture and socioeconomic status (Strauss, 1999). As discussed, the SEM approach allows specification of the theoretical construct "health" as a latent variable measured by a number of observed indicators, and, thus, to profit from multiple measurements of the same construct (Kline, 1998).

Taking into account the discussed challenges and as considering findings from the previous research, the main goal of this study is to propose and test a model that depicts multiple pathways (direct and indirect) through which individual and behavioural characteristics contribute to a particular health state (here related to the CVD risk). In the scope of this contribution, a special focus is given to the measurement model of health as well as to the (in)direct effects of education and income on other model variables.

3. The health production model

The study works within the household production model of consumer behaviour (Becker, 1965; Grossman, 1972) that is widely applied to study health behaviour and outcomes (Rosenzweig and Schulz, 1983; Chen *et al.*, 2002). According to this framework, households combine purchased goods, own time and capacities within some type of a household production process to produce the final commodities (e.g., health), which are the actual sources of their satisfaction. In this respect, Grossman (1972) makes a distinction between a) the demand relationships related to the individual's choice of health inputs, which is constrained by input prices, available resources and individuals' personal characteristics (the reduced-form input demand functions) and b) the relationship between the chosen inputs and health outcome in a form of health production function. A model incorporating simultaneously the production technology and inputs choices is considered to be more appropriate than single-equation estimations (Schultz, 1984; Berman *et al.*, 1994; DaVanzo and Gertler, 1990).

The present study aims to specify a full structural model of health production including demand for health inputs and the health production function. However, some simplifying assumptions are necessary due to data limitations. First, in the absence of longitudinal data we specify a one-period model and, therefore, do not consider potential long-term effects of some health behaviours. Second, exogenous market prices and wage rates[1] that should enter the demand functions were not available and are not part of the model. However, prices are often assumed to be fixed in cross-section analysis (Blaylock *et al.*, 1999).

Based on household production theory of Becker (1965) and taking into account its empirical applications (e.g., in Mazzocchi *et al.*, 2009 and Chen *et al.*, 2002), the stock of health is assumed in this study to be produced according to the following health production function:

$$H = H\,(F, Q, S, A, P, M, W, x_i), \qquad i = 1, \ldots, 7 \tag{1}$$

Health state of an individual *H* is related to the total amount of food consumed *F* and to its nutritional quality *Q*. Further inputs, which are believed to have a

[1] Becker (1965) discusses a "full-income" constraint that combines the income and time constraints. The income constraint presents a money income equal to the sum of earnings from wages and non-labour income, while time constraint is a sum of all time inputs into production of commodities including time for work and leisure. Due to data constraints, the model proposed here incorporates only the information on household money income that is treated here as exogenous variable.

direct health impact, are cigarettes *S,* alcohol intake *A,* physical activity *P*, and medical care *M*. Weight *W* is considered to be related to the health outcome, while overweight and obesity have been showed to be associated with a number of adverse health conditions including CVD risk. x_i is a vector of individual exogenous characteristics such as age, age^2, gender, race, education, household size and income. Thereby, the personal exogenous characteristics are believed to impact both, the efficiency of health production and the demand for inputs (Chen *et al*., 2002; Variyam, 2003).

The health production function is assumed to be unique for each individual, while the chosen inputs may improve health (e.g., nutritious food and exercising) as well as reduce it (*e.g.*, smoking and alcohol). As discussed, health inputs are subjects of individual choices and therefore, are endogenous in the model.

Weight of an individual is seen in the model as in intermediate input into health production that is related to the amount of food consumed *F*, quality of the diet *Q,* physical activity *P*, smoking *S* (the latter is due to empirical evidence of such relation, e.g., Chou *et al*., 2004), and personal exogenous characteristics x_i:

$$W = W(F, Q, P, S, x_i) \tag{2}$$

Another endogenous input, diet quality *Q*, is hypothesized to be affected by the amount of foods consumed *F*, by a person's knowledge of nutrition-related aspects *K* as well as personal characteristics x_i:

$$Q = Q(F, K, x_i) \tag{3}$$

Nutritional knowledge *K* is seen as endogenous in the model. Its intermediate effect in the relation between education and diet is tested. Thus, education is believed to have an indirect effect on dietary quality inducing better knowledge of nutrition-related aspects that in its turn leads to a more balanced diet.

The health production function (1) is estimated together with equations of weight (2) and dietary quality (3) as well as the demand functions of the health inputs *K, F, S, A, P,* and *M*, which are modelled to be dependent on individual's observed personal characteristics (x_i). Moreover, nutrition knowledge *K* is also believed to have an impact on the overall energy intake *F* of an individual:

$$K = K(x_i),\ F = F(K, x_i),\ S = S(x_i);\ A = A(x_i);\ P = P(x_i);\ M = M(x_i) \tag{4}$$

No reciprocal relation was tested by this model due to a) their high complexity and b) a need in a longitudinal data to analyse such linkages (Wong and Law, 1999). The hypothesized relations in the model are estimated simultaneously with consideration of all available information in the data set (full information estimation procedure).

4. Empirical methods

4.1. Data

The data are from the US National Health and Nutrition Examination Survey (NHANES) 2005 to 2006, which is a nationwide probability sample of the population designed to assess the health and nutritional status of adults and children in the United States. Details on the design of the NHANES survey are given elsewhere (Dubowitz *et al.*, 2008). Briefly, the NHANES collects data through home-based interviews, standardised physical exams, and laboratory tests conducted by trained specialists. Dietary information is based on two 24-hour dietary recalls. Data on physical activity, smoking status and other behavioural characteristics were obtained by standardised questionnaires. Blood tests were conducted in laboratory conditions. The sample was restricted to 2,505 adults of 20 to 59 years of age, who were non-pregnant and non-lactating and had reliable two-day dietary information (90% of the NHANES 2005-2006 adult sample).

Missing data on the incorporated variables were estimated and imputed by a two-step iterative procedure of the Expectation Maximisation (EM) procedure in SPSS Missing Values 17.0. The data examination in respect to univariate normality showed the skewed distribution of some variables. However, no data transformation was performed because this distortion seems to be reasonable in the population (e.g., smoking or exercising). Two multivariate outliers were detected based on the Mahalanobis distance available in the software Analysis of Moment Structure (AMOS) 18.0 and were excluded from the further analysis (Byrne, 2001). Therefore, the final sample is comprised of 2,503 persons.

4.2. Variables in the analysis and main hypotheses

Table 1 provides a descriptive statistics of the model variables.

Health status of an individual is specified via a latent variable model. An important advantage of the NHANES data set is that it contains a number of objective (rather than self-reported) health state measurements collected by trained health technicians. It is hypothesized that health status related to CVD can be adequately presented in terms of three indicators. These are diastolic blood pressure, systolic blood pressure, and total cholesterol level. All of them are important markers of diverse health risks and in particular of CVD. As discussed in the following section, at a later stage a respecification of the measurement model for health with introduction of an additional indicator (level of triglyceride) was performed.

Table 1. **Definition, means and standard deviations of variables in the model (N = 2503)**

Variables	Description	Mean (SD)
Endogeneous		
Cardiovascular health (latent), H	*Indicators:*	
	Systolic blood pressure (1st measurement)[2], mmHg, SBP	120.30 (15.51)
	Diastolic blood pressure (1st measurement), mmHg, DBP	71.69 (12.19)
	Total cholesterol, mg/dL, TC	195.65 (39.63)
	Triglyceride, mg/dL, TR	131.16 (82.58)
Dietary quality, Q	F&V consumed (excluding juice), g	260.82 (207.90)
Caloric intake, F	Average energy intake based on two 24-h recalls, kcal	2213.36 (278.63)
Weight, W	Waist circumference, cm	97.01 (15.97)
Smoking, S	Number of days on which smoked cigarettes during past 30 days, range from 0 to 30.	6.75 (12.16)
Alcohol, A	Average number of alcohol drinks per day during past year; a range from 0 to 4	1.78 (1.42)
Medical care, M	Number of times received professional health care over past year, range from 1 (1 time) to 5 (13 times and more)	1.89 (1.48)
Nutritional knowledge, K	Number of positive responses regarding respondent's awareness and usage of nutrition information, range from 0 to 7	3.68 (2.22)
Physical activity, P	Television viewing as an indicator of sedentary leisure time activity. Hours per day of TV or videos watching over past 30 days, 0 h to 5 h and more	2.16 (1.54)
Exogenous, x_i		
Age	Age of respondent in a range from of 20 to 59	39.09 (11.17)
Age^2	Age of respondent squared in a range from 400 to 3,481	1652.81 (877.16)
Male	Dichotomous variable that equals 1 if respondent is male, 0 otherwise	0.50 (0.50)
Non-Hispanic white	Dichotomous variable that equals 1 if respondent is Non-Hispanic white, 0 otherwise	0.47 (0.50)
Household size	Number of people living in household; a range from 1 to 7 and more persons	3.41 (1.63)
Education level	Level of education reached by respondent; a range from 1 (less than 9th grade) to 5 (college education and above)	3.48 (1.20)
Household income	Total annual household income; a range from 1 (lowest income group of up to $5,000) to 11 (highest income of over $ 75,000)	7.69 (2.87)

Standard deviation is reported in parentheses. NHANES sampling weights are not applied in calculation. Missing values are imputed by EM method.

[2] Four blood measurements were performed during NHANES 2005-06. Due to a large share of missing data in the 2nd, 3rd and 4th measurements, the estimates from the 1st tests are used.

Fruit and Vegetable (F&V) intake is employed as an indicator of *dietary quality.* This is due to rich scientific evidence on its preventive effect against CVD (WHO/FAO, 2005; Van Duyn and Pivonka, 2000). Higher F&V consumption is hypothesized in the model to be related to a better health status, i.e., to the lower blood pressure and cholesterol level. *Caloric intake* is the variable reflecting average energy obtained from foods and drinks reported by respondents during two nonconsecutive 24-h recalls.

The majority of studies suggest that Waist Circumference (WC) is a better predictor of cardiovascular disease than the Body Mass Index (BMI) or waist-to-hip ratio (Dobbelsteyn *et al*., 2001; Buchholz and Bugaresti, 2005). Therefore, it serves in the model as a *weight* indicator. Weight is assumed to be an intermediate health determinant, which is affected by personal exogenous characteristics and lifestyles (e.g., diet, exercising), and at the same time contributes to a particular health state.

Smoking is believed to be associated with an increased risk of heart attack and stroke (Lloyd-Jones *et al*., 2010). However, the predicted effects of smoking on health are ambiguous in the empirical studies (Contoyannis and Jones, 2004; Mazzocchi and Traill, 2008). Number of days on which the respondent smoked cigarettes during the past 30 days is used as an indicator of smoking in the model. It is hypothesized that smoking is negatively related to health. Furthermore, to verify the result from previous studies (*e.g.*, Chou *et al*., 2004), an impact of smoking on weight is tested (assumed to be negative).

Heavy drinking showed to be related to chronic diseases, including CVD and cancer. At the same time, moderate *alcohol* consumption might have a protective effect against CVD (US DHHS and US DA, 2005). To test this relation, an average number of alcohol drinks consumed per day during the past year is incorporated in the model.

US Department of Health and Human Services (US DHHS, 2008) emphasises benefits of *leisure time physical activity* in relation to obesity and heart diseases. Empirical studies showed its association with lower weight and better self-reported health (Contoyannis and Jones, 2004; Kenkel, 1995). Average duration of daily TV or video viewing is used as indicator of person's physical activity. A positive association between this measure and overweight and obesity has been demonstrated previously (Eisenmann *et al*., 2002; Jeffery and French, 1998).

An impact of *medical care* utilisation to the observed health differences is tested. Number of times a person used a professional medical care at a doctor's office, hospital or at home during the last 12 months is used in the model as an indicator for medical services usage. It can be also considered as an intermediate factor in health production. Thus, income may contribute to more access to health care that in its turn might have a positive health effect.

Participants' exogenous health endowments are important potential predictors of health behaviour and outcomes. Especially the impact of education and income in these relationships is of an interest for this contribution.

The study tests direct and indirect effect of these variables in the model. Specifically, it is hypothesized that education has an indirect positive impact on diet quality via *nutrition knowledge*. Kenkel (1991) showed that effect of schooling on cigarettes consumption, drinking and exercising decreases by 5% to 20% when health knowledge of a person is accounted for. In the actual model, nutritional knowledge is an endogenous variable affected by a number of personal characteristics (e.g., age, gender). It is a composite variable derived from answers to seven questions related to nutritional behaviour and knowledge. The respondents were asked whether they: heard about 5-a-day program, heard of dietary guidelines, heard of food guide pyramid, use nutrition facts panel on food label, use ingredients list on food label, use serving size information on food label and health claims on food packages.

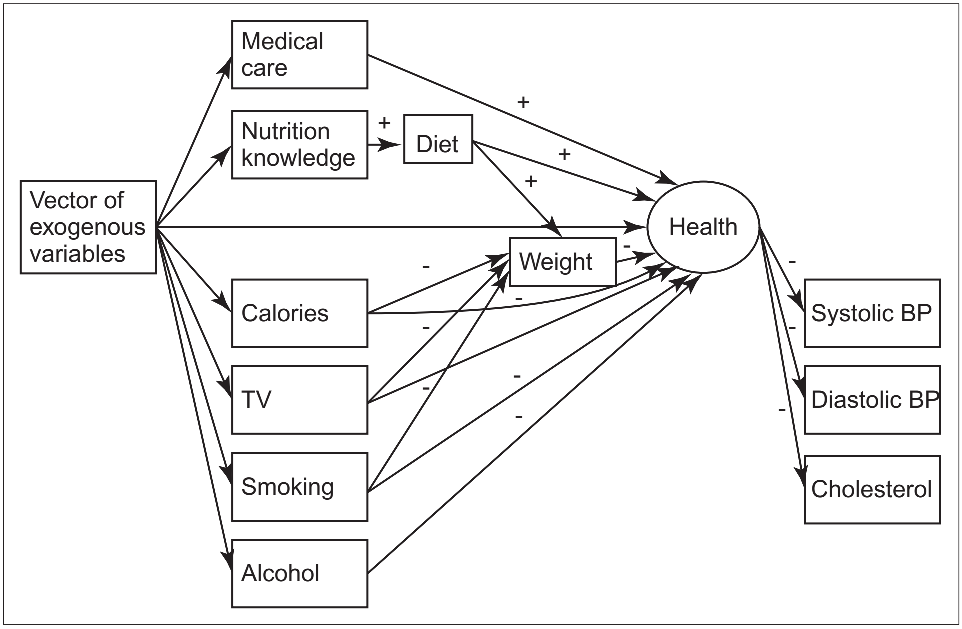
Similarly to the relationship between education and health, there might be various mechanisms by which income affects health status (Smith, 1999). Although it has been a focus of research, there is no agreement reached about its direction (Adelaja *et al*., 1997; Lin *et al*., 2004; Fuchs, 2004). Higher income may be associated with affordability of healthier dietary choices. On the other hand, more affluent individuals may tend to eat out frequently, while restaurant meals showed to be associated with higher total energy intake (Jeffery and French, 1998). They might either join a gym or buy a car and eventually be less physically active. A reverse causation in the relationship between income and health as well as omitted variables could also bias the results (Kawachi, 2010). This study contributes to this research by testing direct and indirect relationships among income, health-related behaviours, and health state.

4.3. Analytical technique: SEM

Briefly, SEM allows to investigate whether a hypothesized model is consistent with the actual patterns in the sample data. The estimation incorporates fitting data to a model or practically, solving a set of equations. The main aspects assessed are the adequacy of parameter estimates and the model fit as a whole. The SEM is comprised of two main parts: the measurement and structural models. A measurement model specifies how the unobserved (latent) variables are measured in terms of observed indicators, whereas a structural model presents the postulated causal relations between all endogenous and exogenous variables in the model. The measurement model (latent variable model) is tested in the SEM methodology through factor analysis, which has a primary goal to explain the covariances or correlations between a number of observed variables by means of some underlying factors. Latent variable is seen as a single factor presented as a condensed statement of the relationships between a set of variables. Factor loadings show relationship between the variables and the factors. They are interpreted as regression coefficients. All relations in SEM are usually expressed graphically by path diagrams or can be formulated by linear regression equations. A full discussion of the SEM approach is available elsewhere (Byrne, 2001; Bollen, 1989).

The structural model showed in Figure 1 gives a schematic representation of the relationships assumed in the model and postulated in the equations (1-4). The latent health construct is depicted as an ellipse with three rectangles for directly measured variables. For the ease of presentation, the left part of the model represents the vector of exogenous variables (socioeconomic characteristics). Single-headed arrows are causal paths.

Figure 1. **Structural model of health**



Note: +/- show the hypothesized positive/negative effects. BP is referred to blood pressure. This is a simplified presentation of full structural model, not showing error terms.

The software AMOS 18.0. is used to verify the fit of the hypothetical model to the sample data. Raw data was prepared in SPSS 18.0 and used as the input for estimation. The correlation matrix of the model variables is showed in the Annex. As AMOS software does not have the option to take into account the *complex sampling design* used in NHANES survey (*i.e.*, by utilization of sampling weights), the analysis is performed on unweighted sample[3], which limits the possibilities of generalisations for the whole population. However, bootstrapping procedure was performed with 2000 iterations to ensure the stability of estimates. Therefore, the results are the means of the parameter estimates from 2000 bootstrap samples.

[3] SEM software packages Mplus and LISREL have an option to use sampling weights in the model fitting procedure. For further discussion of analysis of complex sample data in SEM see, e.g., Muthen and Satorra (1995).

Due to data non-normality and the categorical character of some variables, the estimation employed the Asymptotically Distribution Free (ADF) procedure that does not make the assumption of normality and takes into account the degree of data kurtosis by adjusting the estimation results (Kline, 1998). The final sample of 2,503 persons satisfies the required for this approach a large sample size (over 1,000 cases). To evaluate the overall fit of a model, traditionally the chi-square statistics based on a comparison of the predicted and observed covariance matrices has been used, where the non-significant value indicates a good model fit. However, chi-square is often criticised as unrealistic, due to its dependence on sample size (tendency to a significant chi-square when large sample is used) and assumption of normal distribution (Maruyama, 1998). Therefore, it is suggested to employ additional measures of fit. The main goodness-of-fit indices evaluated in this study are the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) with values of > 0.95 and < 0.05, respectively, indicating a close model fit. Other fit measures presented are the normed fit index (NFI) (ideally > 0.90), the goodness of fit index (GFI) and adjusted goodness of fit index (AGFI), where the values close to 1.00 indicate an adequate model fit (Maruyama, 1998; Byrne, 2001).

After the first run of the full model a minor modification was performed based on the modification indexes in AMOS. This included adding an error correlation between alcohol consumption and smoking. This did not change the magnitude of parameters in the model and their signs, but slightly improved the goodness-of-fit measures. Such modifications are acceptable if they are in line with the theoretical knowledge (Kline, 1998).

5. Results and discussion

Estimation of measurement and structural parts of the model was performed simultaneously. For the sake of simplification, the results from these two parts are presented and discussed separately.

5.1. Health measurement model

The latent health construct is formed by three observed variables, higher values of which suggest a higher risk of CVD. For the convenience of interpretation and in order to give the scale to the latent variable, the coefficient for systolic blood pressure is fixed to -1[4].

[4] A metric is assigned to each unobserved (latent) variable in SEM by constraining one of the indicators' paths a value of 1.0 (a reference item). Given the measurement range of this item, the other paths can be estimated.

Table 2. **Estimation results of the one-factor health measurement model (N = 2,503)**

	Health		
Indicators	**Unstandardised coefficient**	**Standardised coefficient**	**Variance explained (R^2)**
Systolic BP	-1.00[a]	-0.83***	0.69
Diastolic BP	-0.67 (0.03)***	-0.73***	0.54
Total cholesterol	-0.77 (0.08)***	-0.26***	0.07

BP is referred to blood pressure. [a]Set to be a reference indicator. Standard errors are in parentheses.

Parameter estimates, standard errors and significance levels are estimated by bootstrapping with 2,000 iterations.

*** indicate statistical significance at the 99.9% level.

Own computations with the data from NHANES 2005-06.

Since the goodness of fit tests cannot be produced for the three-indicator model (the model is just-identified with df = 0), the CFA solution presented in the Table 2 is evaluated on the basis of the direction, size, and statistical significance of the parameter estimates (Brown, 2006).

Indicators are negatively related to the latent construct "Health" meaning that better health state (in terms of lower risk of CVD) is associated with lower values of these blood measurements. Factor loadings indicate factor's impact on each indicator and in the following are interpreted as standardised regression coefficients[5]. The magnitude of the coefficients of systolic and diastolic BP (-0.83 and -0.73 respectively) and their high significance (p < 0.001) confirm validity of the indicators, while the proportion of explained variance (69% and 54% respectively) indicates their adequate reliability. More specifically, improvement in a health state by 1 standardised score is associated with a decrease in systolic blood pressure by 0.83. About 69% of the variance in systolic blood pressure is explained by its underlying factor "cardiovascular health".

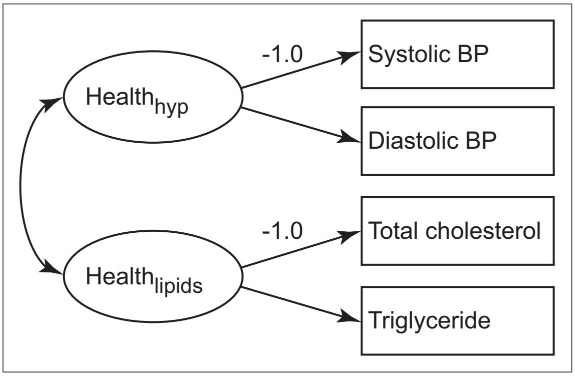
Factor loading of total cholesterol indicator is also highly significant, but lower in magnitude (-0.26). The factor explains only about 7% of the variance in cholesterol level. This finding is in line with the study of Mazzocchi and Traill (2008), who specified a health measurement model in a similar way and showed that the loading of cholesterol variable on health factor was much lower compared to the blood pressure variables. This indicates that total cholesterol level may not reflect the actual latent construct of cardiovascular health well enough, which could be due to a more complex character of the cardiovascular health construct. Thus, it might have several (latent) dimensions, one of which is related to blood pressure measures (hypertension dimension) and another to cholesterol level (lipid risk dimension).

To test this hypothesis, the initial measurement model for health can be respecified in a two-factor latent variable model with one latent variable to

[5] Unstandardised coefficients are more applicable for model comparisons across different samples. In the following only standardised coefficients are presented and discussed.

be presented by two blood pressure indicators and another one by cholesterol measurement. However, such model would be unidentified (df = -1)[6]. To reach the identifiability of the model, one more observed variable is added to the health measurement model, which is the serum triglyceride level. Total cholesterol and triglyceride are believed to represent a risk dimension of cardiovascular health related to elevated level of blood fats (lipids), while blood pressure measurements reflect the hypertension risk dimension. A new modified two-factor measurement model is depicted in Figure 2. Two indicators are specified per each latent variable and the two latent constructs are assumed to co-vary (the double-sided arrow in figure 2). The loading of systolic blood pressure and total cholesterol on their respective factors are fixed to -1.0 to give the scale to the constructs.

Figure 2. **Respecified two-factor measurement model of cardiovascular health (N = 2,503)**



Note: Health$_{hyp}$ shows a dimension of CVD risk, which is related to hypertension, while Healthl$_{lipids}$ reflects a dimension of elevated lipids level in the blood. This is a simplified presentation of the measurement part of the full structural model not showing error measurement of the indicators.

The results from the alternative measurement model in the context of the full structural model are shown in Table 3 including values of selected fit indexes available for this overidentified model (df = 1).

The two-factor latent variable model fit the data well in terms of chi-square test ($\chi^2 = 0.90$, $p > 0.10$) as well and fit indexes (e.g., CFI = 1.0, RMSEA = 0.00). Indicators are significantly related to their corresponding latent constructs with factor loadings ranging from -0.77 for diastolic blood pressure to -0.38 for triglyceride. Notably, the loading of total cholesterol on the "lipids" dimension of cardiovascular health (-0.65) is much higher

[6] Two-factor measurement model with three indicators is empirically unidentified (df = -1). It has 3 observed variables and 6 observations (3(4)/2=6), but 7 parameters to be estimated (including 2 variances of the factors and 3 measurement errors, 1 unanalysed association between factors and 1 factor loading).

Table 3. **Estimation results of the two-factor health measurement model (N = 2,503)**

Indicators	Unstandardised coefficient	Standardised coefficient	Variance explained (R^2)
	$Health_{hyp}$		
Systolic BP	-1.00[a]	-0.66***	0.44
Diastolic BP	-0.92 (0.12)***	-0.77***	0.60
	$Health_{lipids}$		
Total cholesterol	-1.00[a]	-0.65***	0.43
Triglyceride	-1.19 (0.22)***	-0.38***	0.15
χ^2	0.90	CFI	1.000
df	1	RMSEA	0.000
N	2003	NFI	0.999
p-value	0.342	AGF	0.999

BP is referred to blood pressure. [a]Set to be a reference indicator. Standard errors are in parentheses.

Parameter estimates, standard errors and significance levels are estimated by bootstrapping with 2,000 iterations.

***indicate statistical significance at the 99.9% level.

Own computations with the data from NHANES 2005-06.

in comparison to its loading in the initial one-factor measurement model (-0.26). Correlation coefficient between the factors equals to 0.32, which is not high, thus suggesting a discriminant validity among them. Therefore, the two-factor measurement model fits the data well confirming the hypothesis about multidimensionality of cardiovascular health.

5.2. Full structural model with two-factor measurement model of cardiovascular health

Results of the parameter estimations from the full model are reported in Table 4. Chi-square test ($\chi^2 = 505.61$, df = 59, N = 2503, $p < 0.001$) indicates a difference between the estimated and observed covariance matrices, which is likely to be affected by a large sample size. Fit indices employed as additional measures of model fit suggest an acceptable fit of the hypothesized model (CFI = 0.940, RMSAE = 0.055, NFI = 0.934, GFI = 1.0 and AGFI = 1.0). About 47% of variation in cardiovascular health was explained by the model compared to 26% in the initial structural model with one-factor health construct. The associations among variables showed to be logical and mainly statistically significant. However, several coefficients did not reach statistical significance or indicated only small effects, e.g., calorie intake and weight. This might be due to the cross-sectional nature of the data as the effect of the behavioural variables represented over a brief period of time (e.g., variable of total energy consumed is derived based on the information of two days) is lower than when it is accumulated over longer time.

Table 4. **Estimation results of the full structural model of health production (N = 2,503)**

	Health_{hyp}	Health_{lipids}	Weight	Diet	Calorie	Know	Smok	Alc	Med	TV	SBP	DBP	TC	TR
Health_{hyp}											-0.85 ***	-0.75 ***		
Health_{lipids}													-0.59 ***	-0.44 **
Weight	-0.22 ***	-0.23 **												
Diet	0.06 **	$0.06^{\$}$	-0.05*											
Calorie	-0.10 ***	-0.04	0.03	0.27 ***										
Know				0.13 ***	$-0.05^{\$}$									
Smok	0.02	-0.09*	-0.08 ***											
Alc	-0.06 **	-0.09 **												
Med	0.01	0.01												
TV	-0.04	-0.08*	0.10 ***											
Educ	-0.01	0.14 **	-0.06*	0.01	0.05*	0.35 ***	-0.17 ***	-0.11 ***	0.09 **	$-0.05^{\$}$				
HH inc	0.01	-0.04	-0.06*	0.04	0.03	0.06 **	-0.13 ***	0.05*	-0.01	-0.08 **				
Age	-0.31	-0.79 **	0.80 ***	0.19	0.20	0.19	$0.37^{\$}$	0.08	$-0.31^{\$}$	-0.80 ***				
Male	-0.12 ***	-0.02	0.11 ***	-0.03	0.48 ***	-0.30 ***	0.06 **	0.24 ***	-0.24 ***	$0.04^{\$}$				
White	0.07 **	$-0.05^{\$}$	0.06*	0.01	0.08 ***	0.11 ***	0.20 ***	0.12 ***	0.06*	-0.09 ***				
HH size	0.05*	-0.06	0.01	0.03	-0.01	-0.04	-0.06 **	-0.06*	-0.06*	-0.10 ***				
Age^2	-0.01	$0.49^{\$}$	-0.54 ***	-0.01	-0.31*	-0.12	$-0.36^{\$}$	-0.32	0.46 **	0.83 ***				
R^2	0.23	0.24	0.10	0.10	0.26	0.30	0.08	0.13	0.11	0.05	0.72	0.56	0.34	0.20
χ^2	505.61	CFA	0.940											
df	59	RMSEA	0.055											
N	2503	NFI	0.934											
p-value	<0.001	AGF	1.00											
AIC	767.61	GFI	1.00											

The latent health construct is specified as a two-factor model with two indicators per factor.

Parameter estimates (standardised coefficients) and significance levels are bootstrapped estimated based on 2,000 replications of the analysis sampled with replacement.

***, **, *, $^{\$}$ indicate statistical significance at the 99.9%, 99%, 95%, 90% level.

Know: nutritional knowledge, Smok-smoking, Alc: alcohol, Med: medical care, TV: television viewing, Educ: education, HH inc: household income, SBP: systolic blood pressure, DPB: diastolic blood pressure, TC: total cholesterol, TR: triglyceride.

Own computations with the data from NHANES 2005-06.

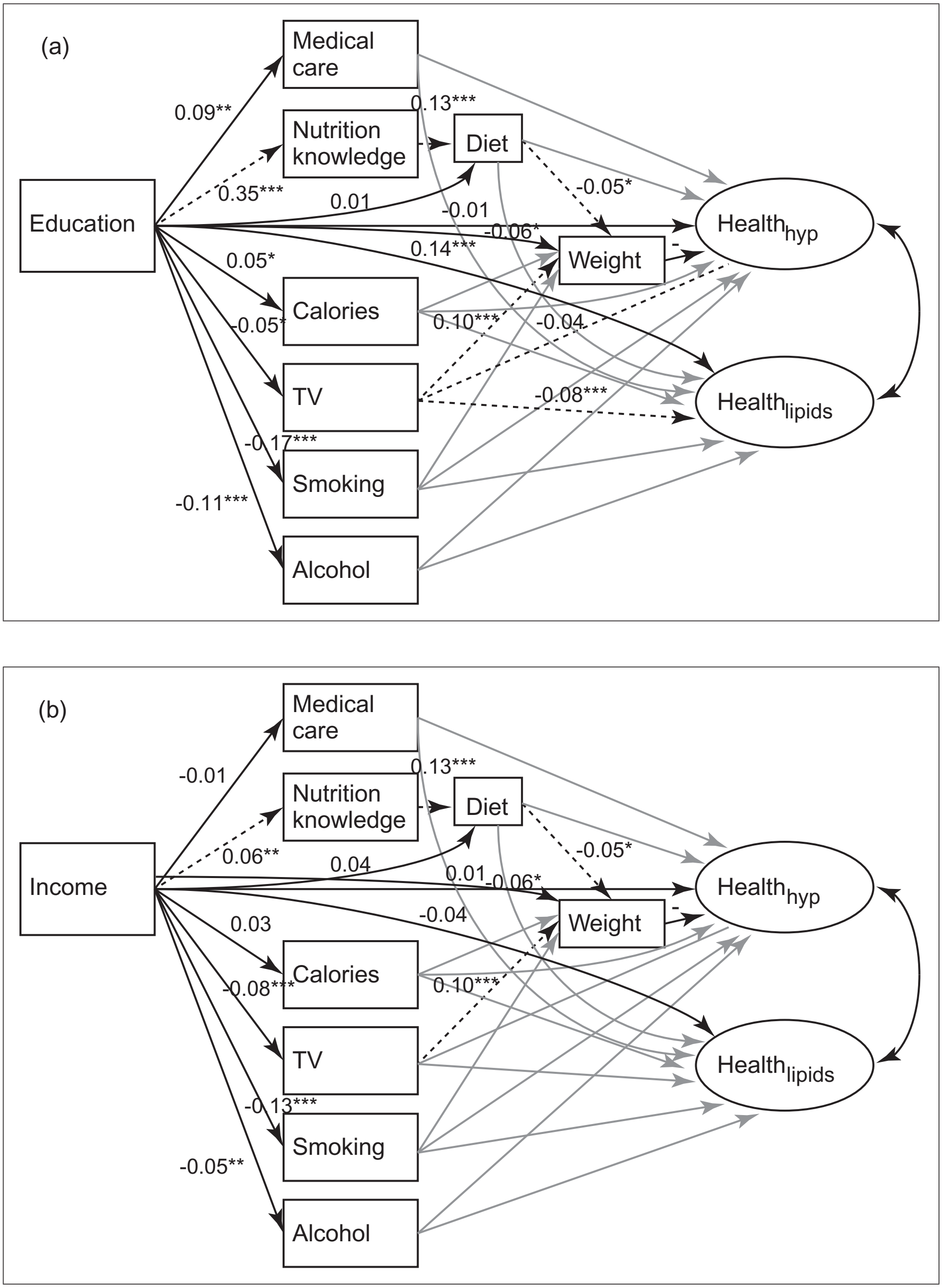
As expected, weight had a strong negative impact on cardiovascular health, while F&V intake and smoking showed to be related to lower weight. Further, hypertension and lipid dimensions of the CVD are explained somewhat differently by the model variables. While higher weight negatively and significantly impacts both health dimensions, a diet rich in F&V contributes more to lowering the hypertension risk. At the same time, smoking, alcohol intake and sedentary leisure time activities are stronger related to the lipid dimension of cardiovascular health (i.e., increase the disease risk). Further, higher education level is strongly associated with better health in terms of low fat blood measurements, but not with lower hypertension risk. These additional insights confirm an importance of accounting for complexity of health status measure.

In the following figures, an impact of education and income on other model variables is presented. For the sake of simplification, Figure 3 focuses on the specific parts of the full model and depicts the (in)direct effects of (a) education and (b) income. The numbers above the arrows represent direct effects that show the expected amount of change in the variable at the end of the arrow produced by a one-unit change in the variable at the beginning of the arrow. The indirect effects of a variable may be mediated by more than one intervening variable, while the total effects can be estimated as a sum of direct and indirect effects (Bollen, 1989).

Higher educational level contributes to a lower risk of cardiovascular health related to its lipids dimension, but not to the dimension of hypertension. This may be due to a higher awareness among higher educated groups of the risks related with cholesterol level in comparison to risks of high blood pressure. Further, higher educational level showed to be associated with healthier lifestyles such as less frequent alcohol intake, cigarette smoking, television viewing and more frequent usage of professional medical care. As hypothesized, there is no significant direct effect of education on diet quality, while nutritional knowledge is the important intervening variable in this relation (dashed line). Higher education contributes to a better awareness of nutrition-related aspects that in its turn leads to a healthier diet. This supports the finding of Variyam *et al*. (1998). In the study of Kenkel (1991), a part of schooling effect on health behaviour still remained after differences in knowledge were controlled for. Based on the actual findings, the indirect effect of education on weight via physical activity can be calculated as follows: -0.05ξ0.10 = -0.01; and via better diet as: 0.01ξ(-0.05) = -0.001. Therefore, although the magnitudes of the effects are rather low, the general trend is that a higher education positively contributes to a lower weight rather via less sedentary activities rather than due to a better diet.

The impact of income on the parameters of the model is showed in Figure 3 (b). No direct health effect was found, but a number of indirect relations via choices of health inputs. Thus, a higher income is associated with less smoking and time spent on television viewing, but with more frequent alcohol intake. In line with previous studies (e.g., Rashad, 2006), higher

Figure 3. **(In)direct impact of education (a) and income (b) on dietary quality and other variables in the model**



Note: ***, **, *, $ indicate statistical significance at the 99.9%, 99%, 95%, 90% level. Own computations with the data from NHANES 2005-06.

earnings are related to lower weight. Higher income contributes to healthier dietary choices rather indirectly via better awareness of nutrition-related aspects, while its direct effect on diet is insignificant. Thus, accounting for complexity of health status and considering (in)direct interrelations provides additional insights in the hypothesized complex relationships.

6. Conclusion

CVD are the main health risks faced by Americans today. This study provides empirical estimates of the relationships between person's socio-demographic characteristics, health-related behaviour, and health state referred here to the CVD. The hypothesized model is constructed and tested based on the household production theory. The person's demand for health inputs as well as decisions related to their combination to produce the optimal health state are viewed and modelled as simultaneously determined using the SEM approach.

The study provides further evidence of the complex links and extends the previous work (Mazzocchi and Traill, 2008; Chen *et al*., 2002; Contoyannis and Jones, 2004) by incorporating additional variables and reporting the (in)direct effects in the health production model. A special focus of this contribution is given to the construction of the measurement model for cardiovascular health. The cardiovascular health was found to be multidimensional and could be adequately presented by two subfactors describing different dimensions of the same health state. Furthermore, the (in)direct effects of income and education on the model variables are discussed. The structural model corresponds to interrelations in the data reasonably well. However, specific aspects of empirical analysis could be a subject of critical discussion.

Although such variables as wages, prices and non-labour income are suggested as good candidates for instruments of endogenous health inputs, they had to be omitted in the analysis due to data limitations. Thus, there is a risk of bias due to omission of important economic variables as well as due to a potential presence of unobserved individual characteristics (e.g., time preferences, individual cognitive abilities, and genetic predisposition). This challenge is especially actively discussed in economic literature with regard to the links between education and health/health-related behaviours. Another limitation concerns potential reverse causality in the model (e.g., from health state to its inputs), which was discussed but not tested in frames of this contribution. Such analysis should be rather done based on longitudinal data. Further, similarly to any other statistical approach, establishing causality from SEM needs satisfying of specific requirements (e.g., temporal precedence of cause and effect, isolation of independents, directionality of effect) (Lei and Wu, 2007). Although these conditions are rather difficult to meet with one period cross-sectional data, a theoretical-driven modelling and usage of accumulated scientific knowledge when making conclusions, may offer important insights into relationships under research (Pearl, 2012).

To summarise, health production presents a complex relationship that can be hardly fully presented and estimated using a single-period cross-sectional data. Future research should focus on testing alternative models to deliver more conviction on the effects of particular variables on health behaviours and CVD risks. Additionally, the longitudinal SEM design might help to investigate the direction of causality of health behaviour and outcomes.

References

Adelaja A., Nayga R. and Lauderbach T. (1997) Income and racial differentials in selected nutrient intakes, *American Journal of Agricultural Economics 79* (5), 1452-1460.

American Heart Association (2006) *Heart Disease and Stroke Statistics - 2006 Update*, American Heart Association, Dallas, USA, 40 p.

American Heart Association (2010) *Heart Disease and Stroke Statistics - 2010 Update*, Dallas, American Heart Association, 36 p.

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

Basiotis P., Carlson A., Gerrior S., Juan W. and Lino M. (2002) *The Healthy Eating Index: 1999-2000, CNPP-12,*, US Department of Agriculture, Center for Nutrition Policy and Promotion, Washington, DC, USA, 20 p.

Becker G. (1965) A theory of the allocation of time, *Economic Journal 75* (299), 493-517.

Behrman J., Deolalikar A. and Wolfe B. (1988) Nutrients: impact and determinants, *The World Bank Economic Review 2* (3), 299-320.

Behrman J., Wolfe B. (1987) How does mother's schooling affect family health, nutrition, medical care usage, and household sanitation?, *Journal of Econometrics 36* (1–2), 185-204.

Behrmann J. and Deolalikar A. (1988) *Health and nutrition*, *in: Handbook of Development Economics*, Chenery H. and Srinivasan T. (eds), *vol. I*, Elsevier Science Publishers, 631-711.

Berman P., Kendall C. and Bhattacharyya K. (1994) The household production of health: integration of social science perspectives on micro-level health determinants, *Social Science and Medicine 38* (2), 205-215.

Blaylock J. and Blisard W. (1992) Self-evaluated health status and smoking behaviour, *Applied Economics 24* (4), 429-435.

Blaylock J., Variyam J. and Lin B. (1999) *Maternal Nutrition Knowledge and Children's Diet Quality and Nutrient Intakes*, Food Assistance and

Nutrition Research Report No. 1, US Department of Agriculture, Economic Research Service, Economic Research Service Report, Washington, DC, USA, 30 p.

Bollen K. (1989) *Structural Equations with Latent Variables*, John Wiley Series in Probability and mathematical statistics, New York, USA, 514 p.

Briscoe J., Akin J. and Guilkev D. (1990) People are not passive acceptors of threats to health: endogeneity and its consequences, *International Journal of Epidemiology 19* (1), 147-153.

Brown T. (2006) *Confirmatory Factor Analysis for Applied Research*, Timothy A. Brown (ed), The Guilford Press, New York, USA, 475 p.

Buchholz A., Bugaresti J. (2005) A review of body mass index and waist circumference as markers of obesity and coronary heart disease risk in persons with chronic spinal cord injury, *Spinal Cord 43* (9), 513-518.

Byrne B. (2001) *Structural Equation Modeling with AMOS: basic concepts, applications, and programming*, Lawrence Erlbaum Associates, New Jersey, USA, 338 p.

Chen S., Shogren J., Orazem P. and Crocker T. (2002) Prices and health: identifying the effects of nutrition, exercise, and medication choices on blood pressure, *American Journal of Agricultural Economics 84* (4), 990-1002.

Chou S., Grossman M. and Saffer H. (2004) An economic analysis of adult obesity: results from the behavioral risk factor surveillance system, *Journal of Health Economics 23* (3), 565-587.

Cobas J., Balcazar H., Benin M., Keith V. and Chong Y. (1996) Acculturation and low-birthweight infants among Latino women: a reanalysis of HHANES data with structural equation models, *American Journal of Public Health* 86 (3), 394-396.

Connell C., Yadrick K., Johnson J. and Su L. (2001) *Structural Analysis of the Relationship of Food Insufficiency to Cardiovascular Disease Risk and Outcomes among Adults from the Southern Region of NHANES III*, Food Assistance and Nutrition Programs: RIDGE Project Summary, USDA, 35 p.

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

DaVanzo J. and Gertler P. (1990) *Household Production of Health: A microeconomic perspective on health transitions*, The Rand Publication Series, The Rand Corporation, Santa Monica, USA, 37 p.

Denton M. and Walters V. (1999) Gender differences in structural and behavioural determinants of health: an analysis of the social production of health, *Social Science and Medicine 48*, 1221-1235.

Dobbelsteyn C., Joffres M., MacLean D. and Flowerdew G. (2001) A comparative evaluation of waist circumference, waist-to-hip ratio and body mass index as indicators of cardiovascular risk factors, The Canadian heart health surveys, *International Journal of Obesity 25* (5), 652-661.

Dubowitz T., Heron M., Bird C., Lurie N., Finch B., Basurto-Davila R., Hale L. and Escarce J. (2008) Neighbourhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican Americans in the United States, *American Journal of Clinical Nutrition 87* (6), 1883-1891.

Eisenmann C., Bartee R. and Wang M. (2002) Physical activity, TV viewing, and weight in US youth: 1999 youth risk behaviour survey, *Obesity Research 10* (5), 379-385.

Erbsland M., Ried W., and Ulrich V. (1995) Health, health care, and the environment. Econometric evidence from German micro data, *Health Economics 4* (3), 169-182.

Flegal K., Carroll M., Ogden C. and Curtin L. (2010) Prevalence and trends in obesity among US adults, 1999-2008, *The Journal of the American Medical Association 303* (3), 235-241.

Fuchs V. (1982) *Time preference and health: an exploratory study*, *in: Economic Aspects of Health*, Fuchs V. (ed), National Bureau of Economic Research, University Chicago Press, Chicago, USA, 93-120.

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

Giuffrida A., Iunes R. and Savedoff W. (2005) *Health and Poverty in Brazil: Estimation by Structural Equation Model with Latent Variables*, Technical Note on Health 1/2005, IDB Publications n°53778, Inter-American Development Bank, New York, Washington, USA, 27 p.

Grossman M. (1972) On the concept of health capital and the demand for health, *Journal of Political Economy 80* (2), 223-255.

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

Hair J. (1995) *Multivariate Data Analysis: with Readings*, 4th ed., Prentice-Hall, New Jersey, USA, 757 p.

Häkkinen U. (1991) The production of health and the demand for health care in Finland, *Social Science and Medicine 33*(3), 225-237.

Hays R., Revicki D. and Coyne K. (2005) Application of structural equation modeling to health outcomes research, *Evaluation and the Health Professions 28* (3), 295-309.

Jeffery R., French S. (1998) Epidemic obesity in the United States: are fast foods and television viewing contributing?, *American Journal of Public Health 88* (2), 277-280.

Kawachi I., Adler N. and Dow W. (2010) Money, schooling, and health: Mechanisms and causal evidence, *Annals of the New York Academy of Sciences 1186*, 56-68.

Kenkel D. (1991) Health behavior, health knowledge, and schooling, *The Journal of Political Economy 99* (2), 287-305.

Kenkel D. (1995) Should you eat breakfast? Estimates from health production functions, *Health Economics 4* (1), 15-29.

Kennedy P. (2003) *A guide to econometrics*, 5th ed., MIT Press Cambridge, USA, 623 p.

Kline R. (1998) *Principles and practice of structural equation modeling, Methodology in social science*, Kenny D. (series editor), The Guilford Press, New York, USA, 354 p.

Lairson D., Lorimor R. and Slater C. (1984) Estimates of the demand for health: males in the pre-retirement years, *Social Science and Medicine 19* (7), 741-747.

Lei P.-W. and Wu W. (2007) Introduction to structural equation modeling: issues and practical considerations, *Educational Measurement: Issues and Practice 26* (3), 33-43.

Leigh J. (1983) Direct and indirect effects of education on health, *Social Science and Medicine 17* (4), 227-234.

Lin B., Huang C. and French S. (2004) Factors associated with women's and children's body mass indices by income status, *International Journal of Obesity 28*, 536-542.

Lloyd-Jones D., Adams R., Brown T., Carnethon M., Dai S., De Simone G. *et al*. (2010) Heart disease and stroke statistics-2010 update. *A report from the American Heart Association, Circulation 121*, e46–e215.

Loureiro M. and Nayga R. (2005) International dimensions of obesity and overweight related problems: an economic perspective, *American Journal of Agricultural Economics 87* (5), 1147-1153.

Maruyama G. (1998) *Basics of structural equation modelling*, Sage Publications, London, 328 p.

Mazzocchi M. and Traill W. (2008) A structural model of wealth, obesity and health in the UK, Selected Paper Presented at the 12th Congress of European Association of Agricultural Economists (EAAE), August 26-29, Ghent, Belgium, electronic copy available at: http://ageconsearch.umn.edu/bitstream/43968/2/298.pdf

Mazzocchi M., Traill W. and Shorgen J. (2009) *Fat Economics: Nutrition, Health, and Economic Policy*, Oxford University Press, UK, 208 p.

McAlister A., Krosnick J. and Milburn M. (1984) Causes of adolescent cigarette smoking: tests of a structural equation model, *Social Psychology Quarterly 47* (1), 24-36.

Muthen B. O., Satorra A. (1995) Complex sample data, *Sociological Methodology 25*, 267-316.

Nayga R.M. (2000) Schooling, health knowledge and obesity, *Applied Economics 32* (7), 815-822.

Nayga R. (2008) Nutrition, obesity and health: policies and economic research challenges, *European Review of Agricultural Economics 35* (3), 281-302.

Oh H., Seo W. (2001) Development of a structural equation model for causal relationships among arteriosclerosis risk factors, *Public Health Nursing 18* (6), 409-417.

Or Z. (2000) Determinants of health outcomes in industrialized countries: a pooled, cross-country, time-series analysis, *OECD Economic Studies 30*, 53-77.

Pearl J. (2012) *The causal foundations of structural equation modelling*, *in: Handbook of Structural Equation Modeling*, Hoyle R. H. (Ed.), Guilford Press, New York, USA, 68-91.

Public Health Agency of Canada (2009) Minimizing the risks of cardiovascular disease, available online at: http://www.phac-aspc.gc.ca/cd-mc/cvd-mcv/risk-risques-eng.php

Rashad I. (2006) Structural estimation of caloric intake, exercise, smoking, and obesity, *The Quarterly Review of Economics and Finance 46* (2), 268-283.

Rosenzweig M. and Schultz T. (1983) Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight, *Journal of Political Economy 91* (5), 723-746.

Schroeter C. and Lusk J. (2008) Economic factors and body weight: an empirical analysis, *Journal of Agricultural and Applied Economics 40* (2), 523-538.

Schultz T. (1984) Studying the impact of household economic and community variables on child mortality, *Population and Development Review 10*, supplement Child survival: strategies for research, 215-235.

Shen B., Takeuchi D. (2001) A structural model of acculturation and mental health status among Chinese Americans, *American Journal of Community Psychology 29* (3), 387-418.

Smith J. (1999) Healthy bodies and thick wallets: the dual relation between health and economic status, *The Journal of Economic Perspectives 13* (2), 145-166.

Strauss R. (1999) Comparison of measured and self-reported weight and height in a cross-sectional sample of young adolescents, *International Journal of Obesity and related Metabolic Disorders*, *23* (8), 904-908.

US DHHS (US Department of Health and Human Services) (2008) *Physical activity guidelines for Americans: Be active, healthy, and happy!*, Washington, DC, US Government Printing Office n°U0036, 63 p.

US DHHS and US DA (US Department of Health and Human Services and US Department of Agriculture) (2005) *Dietary Guidelines for Americans, 2005*, Washington, DC, US Government Printing Office, 71 p.

Van der Gaag J. and Wolfe B. (1991) *Estimating demand for medical care: health as a critical factor for adults and children*, *in: Econometrics of health care, Advanced studies in theoretical and applied econometrics*, Duru G., Paelinck J. (eds), Kluwer, Amsterdam, The Netherlands, 31-58.

Van Duyn M. and Pivonka E. (2000) Overview of the health benefits of fruit and vegetable consumption for the dietetics professional: selected literature, *Journal of American Dietetic Association 100* (12), 1511-1521.

Variyam J. (2003) *Factors affecting the macronutrient intake of US adults*, Bulletin No. 1901, Electronic report from the Economic Research Service, Washington, US DA, 45 p.

Variyam J., Blaylock J. and Smallwood D. (1998) *USDA's Healthy Eating Index and Nutrition Information*, Technical Bulletin No.1866, Economic Research Service and Center for Nutrition Policy and Promotion, Washington, USDA, 26 p.

Wagstaff A. (1993) The demand for health: an empirical reformulation of the Grossman model, *Health Economics 2* (2), 189-198.

WHO (World Health Organization) (2007) *Prevention of cardiovascular disease: Guidelines for assessment and management of cardiovascular risk*, WHO Press, Geneva, Switzerland, 86 p.

WHO/FAO (World Health Organization/Food and Agriculture Organization of the United Nations) (2005) *Fruit and Vegetables for Health*, Report of a Joint FAO/WHO Workshop (1-3 September 2004 in Kobe, Japan), WHO, Geneva, Switzerland, 46 p.

Wolfe B. and Behrman J. (1984) Determinants of women's health status and health-care utilization in a developing country: a latent variable approach, *Review of Economics and Statistics* 66 (4), 696-703.

Wong C-S. and Law K.S. (1999) Testing reciprocal relations by nonrecursive structural equation models using cross-sectional data, *Organizational Research Methods 2* (1), 69-87.

Annex Correlation matrix of model variables

	Age	Age^2	Male	White	HH size	HH inc	Educ	Know	Alc	TV	Weight	Smok	SBP	DBP	TC	Calorie	Diet	TR	Med
Age	1																		
Age^2	.991	1																	
Male	-.012	-.014	1																
White	.109	.113	.036	1															
HH size	-.181	-.198	.005	-.223	1														
HH inc	.111	.097	.034	.217	.064	1													
Educ	.065	.061	-.065	.296	-.308	.367	1												
Know	.114	.112	-.301	.219	-.186	.195	.432	1											
Alc	-.208	-.210	.243	.084	.003	.024	-.054	-.127	1										
TV	.031	.044	.041	-.093	-.074	-.121	-.096	-.093	.035	1									
Weight	.223	.213	.157	.017	-.008	-.037	-.059	-.023	-.061	.133	1								
Smok	.003	.002	.082	.132	-.046	-.135	-.146	-.124	.211	.153	-.029	1							
SBP	.294	.299	.182	-.046	-.097	-.026	-.036	-.026	.008	.072	.257	.001	1						
DBP	.284	.264	.101	.056	-.073	.076	.053	.011	-.020	.050	.219	-.015	.517	1					
TC	.247	.235	.015	.073	-.009	.040	-.027	-.018	.007	.045	.126	.033	.134	.167	1				
Calorie	-.099	-.104	.472	.084	.009	.044	.026	-.166	.217	.023	048	.217	.086	.075	.020	1			
Diet	.167	.162	.042	.083	-0.46	.082	.100	.128	-.013	-.075	-.017	-.130	.000	.025	.057	.203	1		
TR	.076	.071	.112	.020	.014	.002	-.047	-.047	.039	.091	.151	.079	.087	.090	.249	.038	.004	1	
Med	.148	.152	-.221	.117	-.132	.029	.137	.211	-.146	.083	.085	-.004	-0.005	.056	.016	-.116	-.007	.019	1

In bold are correlations significant at least at the 0.05 level (2-tailed).

Know: nutritional knowledge, Smok: smoking, Alc: alcohol, Med: medical care, TV: television viewing, Educ: education, HH inc: household income, SBP: systolic blood pressure, DPB: diastolic blood pressure, TC: total cholesterol, TRI: triglyceride.