

The Contribution of Body Mass Index in the Shared Etiology of Diabetes, Hypertension and Hyperlipidaemia: A Semi-Parametric Trivariate Probit Modeling Approach

Ibrahim Niankara and
Aminata Niankara^{1,2}

- 1 Department of Economics, Legal and Political Sciences, Université Aube Nouvelle (New Dawn University) Burkina Faso
- 2 Department of Cardiology, Office of the Workers' Health (OST), Ouagadougou, Burkina Faso

Abstract

Diabetes, hypertension, and hyperlipidaemia are three medical conditions usually linked to high body fat content. The Body Mass Index (BMI), calculated by dividing a person's weight in kilograms by the square of their height in meters, is the most commonly used measure for monitoring the prevalence of excess body fat content. Past studies linking BMI to the prevalence of diabetes, hypertension and hyperlipidaemia have looked at the effects in isolation, hence assuming independence in their occurrences.

This study takes a different approach, considering the potential interconnectedness of these three metabolic diseases, to model the effect of BMI on their joint likelihood for respondents in the 2008 Medical Expenditure Panel Survey (MEPS) dataset. For this, we specify and estimate a standard univariate probit model, then we move to a fully parametric trivariate probit specification to relax the independence assumption, followed by a semi-parametric trivariate probit specification to further relax the linearity assumption for the parametrically entering numerical risks factors (covariates) in each of the three equations for diabetes, hypertension, and hyperlipidaemia.

The results suggest that the semi-parametric trivariate probit specification is better at capturing the true effects of BMI on the likelihood of these three metabolic diseases in a population. In fact the statistically significant correlation coefficients 0.278, 0.362, and 0.356 between the diabetes, hypertension and hyperlipidaemia equations suggest their joint positive dependence. Furthermore, BMI contributes significantly more to hypertension (5%), followed by diabetes (4.4%), and hyperlipidaemia (2.7%).

Keywords: Body mass index; Metabolic diseases; Semi-parametric estimation; Trivariate probit

Corresponding author: Ibrahim Niankara

✉ brassbe1982@gmail.com

Department of Economics, Legal and Political Sciences, Université Aube Nouvelle (New Dawn University) 01 BP 234 Bobo-Dioulasso 01, Burkina Faso.

Tel: +226 64698205

Citation: Niankara I, Niankara A. The Contribution of Body Mass Index in the Shared Etiology of Diabetes, Hypertension and Hyperlipidaemia: A Semi-Parametric Trivariate Probit Modeling Approach. J Clin Med Ther. 2017, 2:1.

Received: December 03, 2016; **Accepted:** February 17, 2017; **Published:** February 23, 2017

Introduction

Global patterns of health risks show that more than one third of the world's deaths are attributable to a small number of risk factors, with the five leading global risks for mortality being high blood pressure, tobacco use, high blood glucose, physical inactivity, overweight and obesity [1]. These risks factors are responsible for raising the risk of chronic diseases, and affect countries across all income groups: High, middle and low [2].

However, as a country develops overtime, major risks to health shift from traditional risks which are associated with poverty (e.g. inadequate nutrition or unsafe water and sanitation) to modern risks (e.g., Overweight and obesity) [3].

In the US, excess adipose tissue has become an increasing public health concern [4], because of its deleterious effects on multiple body organ systems through thrombogenic, atherogenic, oncogenic, hemodynamic and neurohumoral mechanisms [5]. It has also been linked to multiple medical conditions, such as

diabetes, hypertension, dyslipidaemia and several types of cancer [6-10]. And was in fact identified as one of the leading global risks for mortality, and responsible for 5% of deaths worldwide [1]. For the last 30 years obesity has been primarily diagnosed by using the Body Mass Index (BMI), which is currently the main focus in obesity treatment recommendations, with different treatment cutoff points based upon the presence or absence of obesity related diseases. This simple index of body weight, calculated using a person's weight in kilograms divided by the square of their height in meters [11], has been consistently used in a myriad of epidemiological studies, and has been recommended for individual use in clinical practice to guide recommendations for weight loss and weight control [12].

Despite the numerous studies that looked at the associations between BMI and the risks of metabolic diseases in the past [13-16], no study to the best of our knowledge has investigated these effects jointly. In fact, they all looked at the BMI relationship with each individual disease separately. Given the potential interconnectedness of metabolic diseases [17-19], and the fact that high BMI has its roots in a complex chain of events over time, consisting of socioeconomic factors, environmental and community conditions, and individual behavior, modeling its effects on diabetes, hypertension, and hyperlipidaemia in a joint fashion is very crucial to understanding how their likely incidences relate to one another. By quantifying this joint impact of high BMI, evidence based choices can be made about the most effective interventions [20-23] to jointly target them and improve global health.

Therefore, the objective of this research is simply to model the effects of BMI on the joint likelihood of diabetes, hypertension, and hyperlipidaemia, so as to see not only the differential effects of BMI on each disease condition, but also to understand the correlation between the unobserved risk factors causing them. In doing so, we seek to answer two basic questions: Q01: Are the effects of BMI on each of these metabolic diseases the same? And Q02: Are the unobserved risk factors affecting these three diseases conditions linked? With the following maintained null hypothesis:

- H01: The effects of BMI on these three disease conditions are different.
- H02: The unobserved risk factors affecting these three disease conditions are linked (Presence of shared etiology).

In our quest to test the above hypothesis, the rest of the study is organized as follows: Section 2 provides a background on BMI, diabetes, hypertension, and hyperlipidaemia in our studied population. Section 3 presents the data and the variables used for the econometric estimations. Section 4 illustrates the trivariate probit model linking BMI to the three metabolic diseases (diabetes, hypertension, and hyperlipidaemia). Section 5 discusses the results from the estimations, while section 6 concludes the analysis.

Background on BMI, Diabetes, Hypertension and Hyperlipidaemia

BMI is a summary measure of an individual's height and weight, calculated by dividing a person's weight in kilograms by the square of their height in meters [11]. Using a measure such as BMI allows for a person's weight to be standardized for their height, thus enabling individuals of different heights to be compared. BMI is the most commonly used measure for monitoring the prevalence of overweight and obesity at population level. However, it is only a proxy measure of the underlying problem of excess body fat. As a person's body fat increases, both their BMI and their future risk of obesity-related illness also rise [24]. The MEPS uses the following classification [25]:

- Underweight: If BMI is less than 18.5.
- Normal weight: If BMI is between 18.5-24.9 inclusive.
- Overweight: If BMI is between 25.0-29.9 inclusive.
- Obesity: If BMI is between 30.0-39.9 inclusive.
- Morbidly obese: If BMI greater than or equal to 40.0.

Diabetes is a group of diseases characterized by high levels of blood glucose resulting from defects in insulin production, insulin action, or both [26,27]. Diabetes can lead to serious vascular deterioration and premature death. For example, damage to large blood vessels causes accelerated atherosclerosis and puts diabetics at a 2-to-4 fold higher risk of dying from heart attack or stroke than individuals of the same age without diabetes, while damage to small blood vessels results in end-organ diseases that significantly erode quality and length of life [28]. A combination of blood vessel and nerve damage contributes to poorly healing foot ulcers resulting in over 80,000 lower limb amputations per year. People with diabetes are also at increased risk for other conditions, including higher rates of cancer and infections [29].

The prevalence of diabetes continues to grow in the U.S., with the number of people with diagnosed diabetes now reaching 17.5 million.¹ The rising impact of diabetes is recognized by most specialties and disciplines, usually as a co-morbidity that has to be considered within the broader context of the patient's overall management plan. Diabetes and its complications are potentially preventable yet they are taking an increasing slice of U.S. health dollar. In 2007, one out of every five health care dollars was spent in caring for someone with diagnosed diabetes, while one in ten health care dollars was attributed to diabetes. The estimated cost of diabetes in 2007 was \$174 billion, which includes \$58 billion for diabetes related chronic complications and \$58 billion in indirect cost, in the form of reduced national productivity. For the same year, diabetes caused 445,000 cases of unemployment disability, and accounted for 120 million work days absent, and 6 million reduced productivity days for those not in the workforce.

These 2007 costs figures did not improve as of 2012, American Diabetes Association shows that the total estimated cost of diagnosed diabetes is \$245 billion, including \$176 billion in direct medical costs and \$69 billion in reduced productivity [30]. For

¹<http://diabetes.niddk.nih.gov/dm/pubs/statistics>

the year 2012, the largest components of medical expenditures are hospital inpatient care (43% of the total medical cost), prescription medications to treat the complications of diabetes (18%), antidiabetic agents and diabetes supplies (12%), physician office visits (9%), and nursing/residential facility stays (8%).

The value of lost productivity due to diabetes related premature death was \$26.9 billion in 2007 [31]. The total annual cost of \$174 billion is an increase of \$42 billion since 2002, and suggests that the dollar amount has risen over \$8 billion more each year [32]. This rising cost of diabetes imposes a burden on all sectors of society, in the form of higher insurance premiums paid by employees and employers, reduced earnings through productivity loss, and reduced overall quality of life for people with diabetes. People with diagnosed diabetes incur average medical expenditures of about \$13,700 per year, of which about \$7,900 is attributed to diabetes, and have average medical expenditures approximately 2.3 times higher than what expenditures would be in the absence of diabetes American Diabetes Association et al. [30].

For hypertension, commonly referred to as high blood pressure [33], it is defined as a Systolic² Blood Pressure (SBP) of 140 mm Hg or more, or a diastolic blood pressure (DBP) of 90 mm Hg or more, or taking antihypertensive medication [34]. The standard classification of blood pressure, and also adopted by MEPS [35] is as follows:

- Normal BP: SBP=less than 120 mmHg, and DBP=less than 80mmHg.
- At risk (prehypertension): SBP=120–139 mmHg, and DBP=80–89 mmHg.
- High BP: SBP=140 mmHg or higher, and DBP=90 mmHg or higher.

Hypertension is a serious medical condition which, if not controlled, can lead to more serious cardiovascular conditions [36,37]. Control of hypertension has become a key national priority in the US as part of the Million Hearts Initiative from the Department of Health and Human Services, which aims to prevent 1 million heart attacks and strokes in the US by 2017 [38]. Lifestyle factors, such as salt intake [39] exercise, weight control, and stress reduction [40], can affect the risk and impact of hypertension [41,42].

In the US, an estimated 19.4 percent of adults with hypertension are unaware they have the condition [43], and 25.8 percent or 59.4 million adults were reported to have been told at two or more different health care visits that they have hypertension [25]. Between 2006 and 2011, there was a 25% increase in the number of people visiting US emergency rooms for essential hypertension. Emergency department visits for hypertension with complications and secondary hypertension also rose, from 71.2 per 100,000 population in 2006 to 84.7 per 100,000 population in 2011 [44], costing the US economy almost \$46 billion annually in direct medical expenses and \$3.6 billion in lost productivity [45].

²The minimum blood pressure level [35].

In regards to dyslipidemia, it is defined as an abnormal elevation of plasma lipids such as triglycerides, cholesterol and/or fat phospholipids [46]. In the US the most common form of dyslipidemias are hyperlipidemias; that is, an elevation of lipids in the blood, often due to diet and lifestyle [47]. Hyperlipidaemia itself usually causes no symptoms but can lead to symptomatic vascular disease, including coronary artery disease (CAD), stroke, and peripheral arterial disease [46]. High levels of Triglycerides (TGs) (>1000 mg/dL [>11.3 mmol/L]) can cause acute pancreatitis. Severe hypertriglyceridemia (>2000 mg/dL [>22.6 mmol/L]) can give retinal arteries and veins a creamy white appearance and may contribute to the development of atherosclerosis [47]. High levels of Low-Density Lipoprotein (LDL) can cause arcus corneae and tendinous xanthomas over joints [48]. Early screening of young adults with asymptomatic hyperlipidaemia allow them to benefit from lipid-lowering therapies [49], and avoid complications.

Data and Variables Description

The study uses data from the 2008 MEPS public use file, MEPS HC-121: 2008 Full Year Consolidated Data File. The MEPS collects nationally representative data on health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS is cosponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS)³. For a detailed description of the MEPS survey design [50].

Dependents variables

Diabetes: Adults were classified as having diagnosed diabetes if there was a response of "yes" to a survey question asking whether the adult had been told they had diabetes by a health care professional at two or more different medical visits.

Hypertension: Respondents were classified as having diagnosed hypertension if there was a response of "yes" to a survey question asking whether the adult had been told they had hypertension by a health care professional at two or more different medical visits.

Hyperlipidaemia: Individuals were classified as having diagnosed hyperlipidaemia if there was a response of "yes" to a survey question asking whether the adult had been told they had hyperlipidaemia by a health care professional at two or more different medical visits.

Independents variables

The main independent variable of interest in this study is BMI followed by contextual variables (private, visits.hosp), predisposing conditions (health, limitation), geographical variable (region) and finally socio-demographic control variables (age, Income, education, gender, race). The **Table 1** below provides further description and summary statistics for the variables used in this analysis.

³<http://www.meps.ahrq.gov/>

Trivariate Probit Model of BMI and Diabetes, Hypertension and Hyperlipidaemia

Given that the incidence of each disease is captured by a binary variable (taking the value $D=1$ if the disease occurs, and $D=0$ otherwise), we can model the shared etiology of the three metabolic diseases (D_i for $i=1, 2, 3$) using a trivariate probit model. The general specification (with the individual subscript suppressed for simplicity) for our multivariate probit model with three dependent variables and BMI fixed effect is

$$D_i^* = \beta_{0i} + \beta_{1i}BMI_i + \beta_i'X_i + \epsilon_i, \quad i=1, 2, 3 \quad (1)$$

Where D_i^* is an unobserved variable representing the latent utility (well-being) under disease i , and BMI_i captures the fixed contribution of BMI to the well-being under disease i , while X_i is a vector of observed risk factors (or characteristics) believed to be relevant under disease i , β_{0i} is the intercept coefficient, that is the minimum level of well-being not accounting for any risk factor. β_i' is a vector of unknown coefficients to be estimated. The last term ϵ_i represents the impact of unobserved risk factors on the well-being under disease i . ϵ_i is assumed normally distributed with mean μ_i and variance σ_i and a variance-covariance matrix of:

$$\Sigma = \begin{pmatrix} \sigma_1 & \theta_{12} & \theta_{13} \\ & \sigma_2 & \theta_{23} \\ & & \sigma_3 \end{pmatrix} \quad (2)$$

Therefore, the stochastic component of the general multivariate probit specification in equation (1) with the three latent (unobserved) continuous variables D_i^* follow the trivariate normal distribution:

$$\begin{pmatrix} D_1^* \\ D_2^* \\ D_3^* \end{pmatrix} \sim N_3 \left[\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1 & \theta_{12} & \theta_{13} \\ & \sigma_2 & \theta_{23} \\ & & \sigma_3 \end{pmatrix} \right] \quad (3)$$

Where μ_i and σ_i are respectively the mean and variance for D_i^*

and θ_{ij} are scalar correlation parameters. In this formulation each triplet of metabolic diseases (diabetes, hypertension, hyperlipidaemia) (Y_{i1}, Y_{i2}, Y_{i3}) has $2 \times 2 \times 2 = 8$ potential outcomes, ($Y_{i1}=1, Y_{i2}=1, Y_{i3}=1$), ($Y_{i1}=1, Y_{i2}=1, Y_{i3}=0$), and ($Y_{i1}=1, Y_{i2}=0, Y_{i3}=1$), and ($Y_{i1}=0, Y_{i2}=1, Y_{i3}=2$), and ($Y_{i1}=0, Y_{i2}=1, Y_{i3}=0$), and ($Y_{i1}=0, Y_{i2}=0, Y_{i3}=1$), and ($Y_{i1}=1, Y_{i2}=0, Y_{i3}=0$), ($Y_{i1}=0, Y_{i2}=0, Y_{i3}=0$). The joint probability for each of these eight outcomes is modeled with six systematic components: The marginal probabilities $Pr(Y_{i1}=1)$, $Pr(Y_{i2}=1)$, and $Pr(Y_{i3}=1)$ and the correlation parameters ϑ_{12} , ϑ_{13} , and ϑ_{23} for the three marginal distributions. For identification purposes, the standard probit model restricts the diagonal elements (variances) σ_i , $i=1, 2, 3$ in equation (3) to 1. Since the correlation parameters do not correspond to one of the metabolic disease outcomes, the model estimates ϑ_{12} , ϑ_{13} , and ϑ_{23} as constants by default. Hence, only the three means equations (average well-being μ_1 under the first disease "diabetes"; and the average well-being μ_2 under second disease "hypertension"; and the average well-being μ_3 under third disease "hyperlipidaemia") are required. Each of these systematic components are modeled as functions of the sets of risk factors or explanatory variables. The following observation mechanism links the observed disease status, D_i , with the latent variables (well-being) D_i^* .

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Thus the joint probability of a triplet of disease outcomes $\{D_i=d_i, i=1, 2, 3\}$, conditioned on parameters β_{0i} , BMI, Σ and a set of risks factors (explanatory variables) X , can be written as:

$$Pr[D_i=d_i, i=1, 2, 3 | BMI, \beta, \Sigma] = \int_{A_1} \int_{A_2} \int_{A_3} \phi(z_1, z_2, z_3, \theta_{12}, \theta_{13}, \theta_{23}) dz_1 dz_2 dz_3 \quad (5)$$

Where ϕ is the standard multivariate normal density function with mean 0 and variance covariance matrix Σ , and A_i is the interval $(-\infty, \beta_{0i} + \beta_{1i}BMI_i + \beta_i'X_i)$ if $d_i=1$ and $(\beta_{0i} + \beta_{1i}BMI_i + \beta_i'X_i, \infty)$ if $d_i=0$ [51]. The parameters β_{0i} , β_{1i} , β_i' and the three correlations of the error terms are estimated via maximum likelihood methods. However, In addition to the fully parametric trivariate probit specification in equation (1), we consider a more flexible specification in the form of a semi-parametric trivariate probit as

Table 1: Summary description of the variables used in the econometric modeling. Source: 2008 medical expenditure panel survey (MEPS) data set.

	18592	Mean	SD
Bmi	Body mass index	27.86	6.195
Diabetes	Equal to 1 if diabetic	0.077	0.267
Hypertension	Equal to 1 if hypertensive	0.249	0.432
Hyperlipidaemia	Equal to 1 if hyperlipidemic	0.241	0.428
Age	Age in years	39.89	13.459
Income	Income in 1000 of \$	62499	53732.8
Education	Years of education	12.66	2.991
Gender	Equal to 1 if male	0.47	0.499
Limitation	Equal to 1 if health limits physical activity	5.92	0.271
Private	Equal to 1 if individual has private health insurance	0.635	0.481
Region	Levels: 2 northeast, 3 mid-west, 4 south, 5 west	3.768	1.015
Race	Levels: 2 white, 3 black, 4 native American, 5 others	2.477	0.895
Health	Levels: 5 excellent, 6 very good, 7 good, 8 fair, 9 poor	6.287	1.082
Visits.hosp	Equal to 1 if at least one visit to hospital outpatient departments	0.136	0.342

shown in equation (6).

$$D_i^* = \beta_{0i} + \beta_{1i}BMI_i + g(\beta_i'X_i) + \epsilon_i, \quad i = 1, 2, 3 \quad (6)$$

Where everything is as previously defined for equation (1), with the only added exception being $g(\cdot)$ an unknown function to be estimated, along with the parameters of the model. This specification allows us to relax the linearity assumption of the former specification in regards to the numerical variables (risks factors), such as “age”, “education” and “income”. More specifically the unknown function $g(\cdot)$ will be represented as $g(\text{age}, \text{education}, \text{income})$. This choice is motivated by the fact that variables such as age, education, income are likely to have non-linear relationships with disease outcomes [52], because they embody productivity and life-cycle effects that are likely to have non-linear influences on disease prevalence. Imposing a priori linear relationship (or non-linear by simply using for example quadratic polynomials) could mean failing to capture the true and more complex relationships. Both the fully parametric model and semi-parametric model are estimated using the library in the R Statistical Software [52,53].

Results

Given the aim of the analysis, to evaluate the impact of BMI on the joint likelihood/incidence of diabetes, hypertension, hyperlipidaemia, we’ve specified and estimated three regression

models. The first model as a univariate regression assumes independence of the three disease conditions, and captures the effects of BMI individually on each disease. The second specification acknowledges the potential interactions between the three metabolic diseases, and models them jointly, while assuming complete linearity of the effects of all parametrically entering co-variables. The third and last specification goes one step further to relax the linearity assumption in the previous two models, by specifying a general form for the parametrically entering numerical covariates “age”, “education”, and “income”. In doing so, the three models correspond respectively to “(1) the binary univariate probit model”, “(2) the fully parametric trivariate probit model”, “(3) the semi-parametric trivariate probit model”, with the results presented in **Tables 2-4** respectively for diabetes, hypertension, and hyperlipidaemia. The value of these three specifications is in allowing us to test the robustness of the relationship between BMI and each of these three health conditions, to potential misspecification problems.

Diabetes equation results

Starting with the results from the diabetes equation, as shown in **Table 2**, the second column contains the results of the univariate probit model, the third column those of the fully parametric trivariate probit model, and the fourth column presents the results of the semi-parametric trivariate probit model. The

Table 2: Fully-parametric and semi-parametric trivariate probit estimates for the diabetes (†Standard deviation of the parameters in parentheses; ††The 95% confidence intervals for the theta correlations; ***0.01% level significance; **1% level significance; *5% level significance).

Diabetes	Binary Univariate Probit (1)	Fully-Parametric Trivariate Probit (2)	Semi-Parametric Trivariate Probit (3)
Const	-4.484*** (0.149)†	-4.572*** (0.147)	-3.513*** (0.111)
BMI	0.043*** (0.002)	0.044*** (0.002)	0.044*** (0.002)
Age	0.034*** (0.001)	0.034*** (0.001)	p-val <2e-16*** (edf=5.985)
Education	-0.024*** (0.005)	-0.022*** (0.005)	p-val=13e-5*** (edf=1.662)
Income	0.000 (0.000)	-0.000 (0.000)	p-val=0.322 (edf=1.569)
Male	0.087** (0.032)	0.090*** (0.032)	0.090** (0.032)
Black	0.168*** (0.040)	0.170*** (0.040)	0.169*** (0.040)
Native Amer	0.415** (0.130)	0.426*** (0.127)	0.434*** (0.127)
Others	0.223*** (0.060)	0.224*** (0.060)	0.220*** (0.060)
Very Good	0.244*** (0.059)	0.249*** (0.058)	0.251*** (0.058)
Good	0.685*** (0.056)	0.678*** (0.055)	0.682*** (0.055)
Fair	0.971*** (0.062)	0.966*** (0.062)	0.971*** (0.062)
Poor	1.130*** (0.081)	1.124*** (0.081)	1.132*** (0.081)
Limitation	-0.076 (0.048)	-0.078 (0.048)	-0.083 (0.048)
Midwest	-0.135* (0.055)	-0.137* (0.048)	-0.138* (0.055)
South	-0.032 (0.047)	-0.034 (0.047)	-0.034 (0.047)
West	-0.006 (0.052)	-0.010 (0.051)	-0.013 (0.051)
Private	0.058 (0.038)	0.071 (0.038)	-0.063 (0.038)
Visithosp	0.179*** (0.041)	0.171*** (0.041)	0.171*** (0.040)
θ^{12}	-	0.276 (0.248, 0.321)††	0.278 (0.24, 0.315)
θ^{13}	-	0.361 (0.327, 0.394)	0.362 (0.328, 0.396)
θ^{23}	-	0.353 (0.327, 0.383)	0.356 (0.334, 0.387)
AIC	-	38384.34	38349.59

Table 3: Fully-parametric and semi-parametric trivariate probit estimates for the hypertension equation (†standard deviation of the parameters in parentheses; ††The 95% confidence intervals for the theta correlations; ***0.01% level significance; **1% level significance; *5% level significance).

Hypertension	Binary Univariate Probit (1)	Fully-Parametric Trivariate Probit (2)	Semi-Parametric Trivariate Probit (3)
Const	-4.466*** (0.109)†	-4.482*** (0.109)	-2.766*** (0.081)
BMI	0.050*** (0.002)	0.050*** (0.002)	0.050*** (0.002)
Age	0.043*** (0.001)	0.043*** (0.001)	p-val <2e-16*** (edf=2.585)
Education	-0.030 (0.004)	-0.004 (0.004)	p-val=0.423 (edf=1.706)
Income	0.000* (0.000)	-0.000* (0.000)	p-val=0.038* (edf=1.116)
Male	0.161*** (0.023)	0.161*** (0.023)	0.161** (0.023)
Black	-0.300*** (0.030)	-0.300*** (0.030)	-0.298*** (0.030)
Native Amer	0.260* (0.111)	0.253*	0.254* (0.109)
Others	0.116** (0.044)	0.116** (0.044)	0.119** (0.044)
Very Good	0.317*** (0.033)	0.318*** (0.033)	0.318*** (0.033)
Good	0.529*** (0.034)	0.527*** (0.034)	0.528*** (0.034)
Fair	0.888*** (0.043)	0.886*** (0.043)	0.888*** (0.043)
Poor	0.119*** (0.067)	0.120*** (0.066)	0.124*** (0.067)
Limitation	-0.126** (0.042)	-0.130** (0.042)	-0.123** (0.042)
Midwest	-0.093* (0.040)	-0.089* (0.039)	-0.090* (0.039)
South	0.031 (0.035)	0.034 (0.035)	0.033 (0.034)
West	-0.085* (0.038)	-0.085* (0.038)	-0.083* (0.038)
Private	0.095*** (0.028)	0.0945*** (0.028)	0.096*** (0.028)
Visithosp	0.168*** (0.032)	0.164*** (0.032)	0.163*** (0.032)
θ^{12}	-	0.276	0.278
	-	(0.248, 0.321)††	(0.24, 0.315)
θ^{13}	-	0.361	0.362
	-	(0.327, 0.394)	(0.328, 0.396)
θ^{23}	-	0.353	0.356
	-	(0.327, 0.383)	(0.334, 0.387)
AIC	-	38384.34	38349.59

correlation coefficients $\hat{\theta}_{12}=0.276$, $\hat{\theta}_{13}=0.361$, $\hat{\theta}_{23}=0.353$ in the third column suggest that the processes leading to the incidence of diabetes, hypertension and hyperlipidaemia are significantly correlated as shown by their respective 95% confidence intervals (0.248, 0.321), (0.327, 0.394), (0.327, 0.383). These results suggest that the unobserved risk factors affecting the three diseases are positively correlated. As such, the independence assumption through the univariate probit formulation (as formerly adopted in the literature) is not as appropriate as the trivariate probit representation. Comparing now the fully parametric trivariate probit representation to the semi-parametric trivariate probit, we note that the latter is a better model based on the AIC criteria ($AI C_{full}=38384.34 > AI C_{semi}=38349.59$). These results suggest that relaxing the independence assumption, along with the linearity assumption for the parametrically entering numerical risk factors (covariates) yield a better model for the description of diabetes incidence.

Focusing on our primary independent variable of interest, BMI, its effect on the incidence of diabetes is fairly stable and consistent across all three specifications. The statistically significant coefficient value of 0.044 on our preferred model in the fourth column of **Table 2** suggests that a one unit increase in

BMI raises the likelihood of the respondent developing diabetes by 4.4%. With regards to the other risk factors (variables) entering the diabetes equation parametrically, the results in the fourth column of **Table 2** suggests that: Male respondents have 9.0% more chances of being diagnosed with diabetes compared to females. Blacks, native Americans, and other races have respectively 16.9%, 43.4%, and 22% more chances of developing diabetes than whites. Compared to respondents with excellent health conditions, those with very good, good, fair, and poor health conditions have respectively 25.1%, 68.2%, 97.1% and 113.2% more chances of being diagnosed with diabetes. The coefficient value of -0.083 suggests that respondents for which health limits physical activities have relatively 8.3% less chances of being diagnosed with diabetes. The regional dummy variables suggest that compared to respondents from the northeast of the US, those living in the midwest, south, and west have respectively 13.8%, 3.4%, and 1.3% less chances of being diagnosed with diabetes. The coefficient value of -0.063 suggests that respondents with private health insurance coverage have 6.3% less chances of being diagnosed with diabetes compared to those who do not have private coverage. Finally, respondents with at least one visit to hospital outpatient departments have

Table 4: Fully-parametric and semi-parametric trivariate probit estimates for the hyperlipidaemia equation (†Standard deviation of the parameters in parentheses; ††The 95% confidence intervals for the theta correlations; ***0.01% level significance; **1% level significance; *5% level significance).

Hyperlipidaemia	Binary Univariate Probit (1)	Fully-Parametric Trivariate Probit (2)	Semi-Parametric Trivariate Probit (3)
Const	-3.989*** (0.106)†	-3.985*** (0.106)	-2.061*** (0.080)
BMI	0.028*** (0.002)	0.028*** (0.002)	0.027*** (0.002)
Age	0.043*** (0.001)	0.043*** (0.001)	p-val <2e-16*** (edf=3.787)
Education	0.012** (0.004)	-0.013** (0.004)	p-val=0.003** (edf=2.560)
Income	0.000** (0.000)	-0.000** (0.000)	p-val=0.044 (edf=1.400)
Male	0.189*** (0.023)	0.185*** (0.023)	0.186** (0.023)
Black	-0.130*** (0.031)	-0.131*** (0.031)	-0.125*** (0.031)
Native Amer	0.110 (0.112)	0.113 (0.111)	0.113 (0.112)
Others	0.131** (0.041)	0.131** (0.041)	0.125** (0.041)
Very Good	0.271*** (0.032)	0.270*** (0.032)	0.269*** (0.032)
Good	0.428*** (0.033)	0.422*** (0.033)	0.420*** (0.033)
Fair	0.765*** (0.043)	0.763*** (0.043)	0.760*** (0.043)
Poor	0.806*** (0.065)	0.806*** (0.065)	0.797*** (0.065)
Limitation	-0.098* (0.041)	-0.102* (0.041)	-0.114** (0.041)
Midwest	-0.104** (0.039)	-0.102** (0.039)	-0.101** (0.041)
South	0.019 (0.034)	0.019 (0.034)	0.018 (0.034)
West	-0.055 (0.037)	-0.056 (0.036)	-0.060 (0.037)
Private	0.164*** (0.028)	0.165*** (0.028)	0.161*** (0.028)
Visithosp	0.288*** (0.031)	0.287*** (0.031)	0.290*** (0.031)
θ^{12}	-	0.276	0.278
	-	(0.248, 0.321)††	(0.24, 0.315)
θ^{13}	-	0.361	0.362
	-	(0.327, 0.394)	(0.328, 0.396)
θ^{23}	-	0.353	0.356
	-	(0.327, 0.383)	(0.334, 0.387)
AIC	-	38384.34	38349.59

17.1% more chances of being diagnosed with diabetes.

For the smoothed numerical risk factors (variables) “age”, “education” and “income”, calculated as discussed in section 2.3, the results in the fourth column of **Table 2** suggest that only “age” and “education” with p-values (<0.05), have statistically significant effects on the incidence of diabetes [52]. These results are also supported by the smooth function estimates and 95% confidence bands on the variables as shown in **Figure 1**. In fact, the figure suggests that as age increases, the likelihood of being diagnosed with diabetes increases. On the other hand, as the number of years of education increases the likelihood of being diagnosed with diabetes decreases.

Hypertension equation results

Moving to the results from the hypertension equation, as shown in **Table 3**, the second column contains the results of the univariate probit model, the third column those of the fully parametric trivariate probit model, and the fourth column presents the results of the semi-parametric trivariate probit model.

As previously mentioned for the diabetes equation, the statistically significant correlation co-efficients $\hat{\theta}_{12}=0.276$, $\hat{\theta}_{13}=0.361$, $\hat{\theta}_{23}=0.353$, along with the AIC criteria ($AIC_{full}=38384.34 > AIC_{semi}=38349.59$) suggest that relaxing the independence assumption, along with the linearity assumption for the parametrically entering numerical risk factors yield a better

model for the description of hypertension incidence among 2008 MEPS respondents.

Focusing on our primary independent variable of interest, BMI, its effect on the incidence of hypertension is fairly stable and consistent across all three specifications. The statistically significant coefficient value of 0.050 on our preferred model in the fourth column of **Table 3** implies that a one unit increase in BMI raises the likelihood of the respondent being diagnosed with hypertension by 5.0%. With regards to the other risk factors entering the hypertension equation parametrically, the results in the fourth column of **Table 3** suggests that: Male respondents have 16.1% more chances of being diagnosed with hypertension than females. Blacks, native Americans, and other races have respectively 29.8%, 25.4%, and 11.9% more chances of developing hypertension than whites. Compared to respondents with excellent health conditions, those with very good, good, fair, and poor health conditions have respectively 31.8%, 52.8%, 88.8% and 112.4% more chances of developing hypertension. The coefficient value of -0.123 suggest that respondents for which health limits physical activities have relatively 12.3% less chances of being diagnosed with hypertension. The regional dummy variables suggest that compared to respondents from the northeast of the US, those living in the midwest and the west have respectively 9.0% and 8.3% less chances of being diagnosed with hypertension, while those living in the south have 3.4% more chances of being diagnosed with hypertension. The coefficient

value of 0.096 suggests that respondents with private health insurance coverage have 9.6% more chances of being diagnosed with hypertension than those who do not have private coverage. Finally, respondents with at least one visit to hospital outpatient departments have 16.3% more chances of being diagnosed with hypertension.

For the smoothed numerical risk factors “age”, “education” and “income”, calculated as discussed in section 2.3 [47], the results in the fourth column of **Table 3** suggest that only “age” and “income” with p-values (<0.05), have statistically significant effects on the incidence of hypertension. These results are also supported by the smooth function estimates and 95% confidence bands on the variables as shown in **Figure 2**. In fact, the figure suggests that as age increases, the likelihood of being diagnosed with hypertension increases. On the other hand, as income increases the likelihood of being diagnosed with hypertension decreases.

Hyperlipidaemia equation results

The results from the hyperlipidaemia equation are shown in **Table 4**, the second column contains the results of the univariate probit model, the third column those of the fully parametric trivariate probit model, and the fourth column presents the results of the semi-parametric trivariate probit model. As previously mentioned for the diabetes and hypertension equations, the statistically significant correlation coefficients $\hat{\vartheta}_{12}^* = 0.276$, $\hat{\vartheta}_{13}^* = 0.361$, $\hat{\vartheta}_{23}^* = 0.353$, along with the AIC criteria ($AIC_{full} = 38384.34 > AIC_{semi} = 38349.59$) suggest that relaxing the independence assumption, along with the linearity assumption for the parametrically entering numerical risk factors yield a better model for the description of hyperlipidaemia incidence among 2008 MEPS respondents. As such, this model is our preferred model in this analysis.

Focusing on our primary independent variable of interest, BMI, its effect on the incidence of hyperlipidaemia is fairly stable and consistent across all three specifications. The statistically significant coefficient value of 0.027 on our preferred model in the fourth column of **Table 4** suggests that a one unit increase in BMI raises the likelihood of the respondent developing hyperlipidaemia by 2.7%. With regards to the other risk factors entering the diabetes equation parametrically, the results in the fourth column of **Table 4** suggests that: Male respondents have 18.6% more chances of developing hyperlipidaemia than females. Compared to whites, blacks have 12.5% less chances of being diagnosed with hyperlipidaemia, while other races except for native Americans have 12.5% more chances of developing hyperlipidaemia than whites. Compared to respondents with excellent health conditions, those with very good, good, fair, and poor health conditions have respectively 26.9%, 42.0%, 76.0% and 79.7% more chances of developing hyperlipidaemia. The coefficient value of -0.114 suggest that respondents for which health limits physical activities have relatively 11.4% less chances of being diagnosed with hyperlipidaemia.

The regional dummy variables suggest that compared to respondents from the northeast of the US, those living in the midwest and the west have respectively 10.1% and 6.0% less

chances of being diagnosed with hyperlipidaemia, while those living in the south have 1.8% more chances of being diagnosed with hyperlipidaemia. The coefficient value of 0.161 suggests that respondents with private health insurance coverage have 16.1% more chances of being diagnosed with hyperlipidaemia than those who do not have private coverage. Finally, respondents with at least one visit to hospital outpatient departments have 29.0% more chances of being diagnosed with hyperlipidaemia.

For the smoothed numerical risk factors “age”, “education” and “income”, calculated as discussed in section 2.3, the results in the fourth column of **Table 2** suggest that only “age” and “education” with p-values (<0.05), have statistically significant effects on the incidence of hyperlipidaemia. These results are also supported by the smooth function estimates and 95% confidence bands on the variables as shown in **Figure 3**. In fact, the figure suggests that as age increases, the likelihood of being diagnosed with hyperlipidaemia increases. On the other hand, as the number of years of education increases the likelihood of being diagnosed with hyperlipidaemia decreases.

Finally comparing the marginal effects of BMI across all three disease conditions, using the semi-parametric trivariate probit model estimates in the fourth columns of **Tables 2-4** we can note that BMI has a relatively greater incidence on hypertension (5%), followed by diabetes (4.4%), and hyperlipidaemia (2.7%).

Conclusion

The motivation for this empirical analysis was the desire to understand the role that BMI plays in the joint likelihood of the three metabolic diseases of diabetes, hypertension, hyperlipidaemia. This was intended to test whether the independence assumption made by the past literature about the three disease processes is valid, and also to see if BMI has differing effects on them. To this end, the paper used three model specifications. The first was a univariate probit model, the second a fully-parametric trivariate probit model, and the third a semi-parametric trivariate probit model.

The study used data from the 2008 MEPS. The estimated correlation coefficients suggested that the unobserved risk factors affecting these three disease processes are positively related. Hence our semi-parametric bivariate probit specification is better model than the standard univariate probit which have been used by the past literature. In this past specification, the underlying assumption was that these processes are independent. However, as our estimations show taking into account the interdependencies in the processes generating those three metabolic diseases, allows for better more precise estimates of the marginal effects.

Furthermore, focusing on our preferred specification, it was shown that BMI has positive but differing marginal effects on the joint likelihood of diabetes, hypertension, and hyperlipidaemia. In fact, its marginal effect was found to be relatively greater on the incidence of hypertension, followed by diabetes, and then hyperlipidaemia. Overall, this study shed lights on the importance of joint modeling when investigating the incidence of potentially interrelated health conditions, so as to capture the

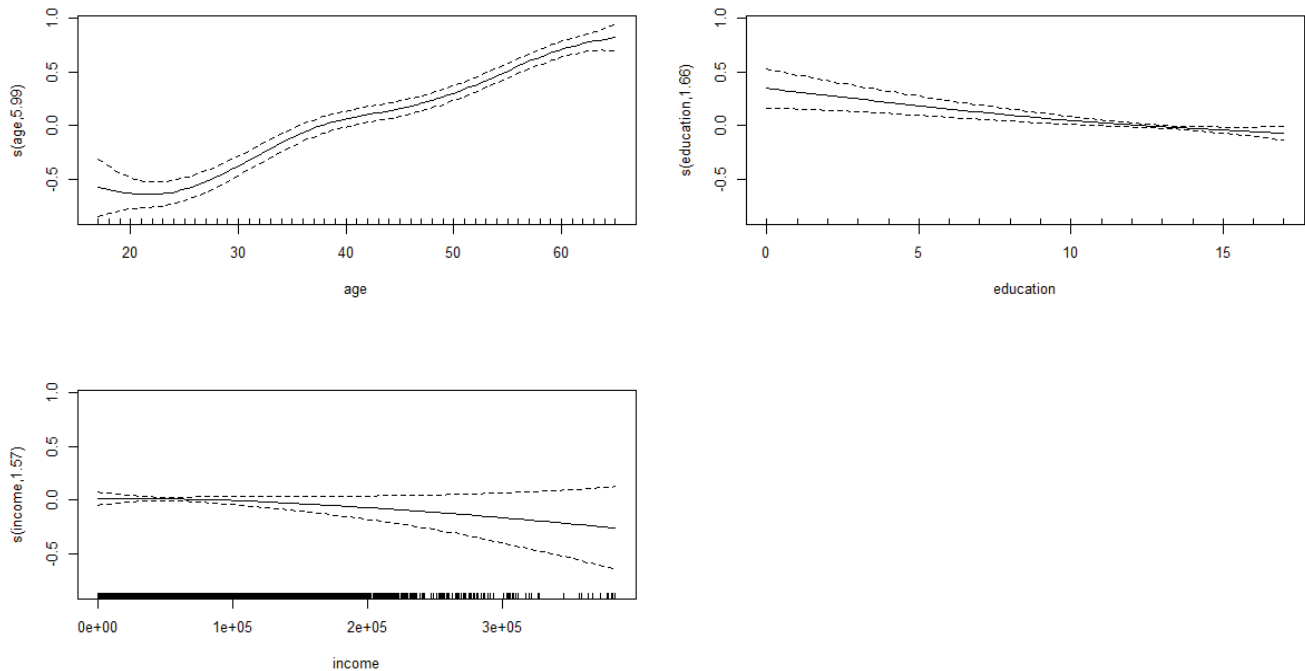


Figure 1 Smooth function estimates and 95% confidence bands for the numerical covariates (age, education, income) in the diabetes equation.

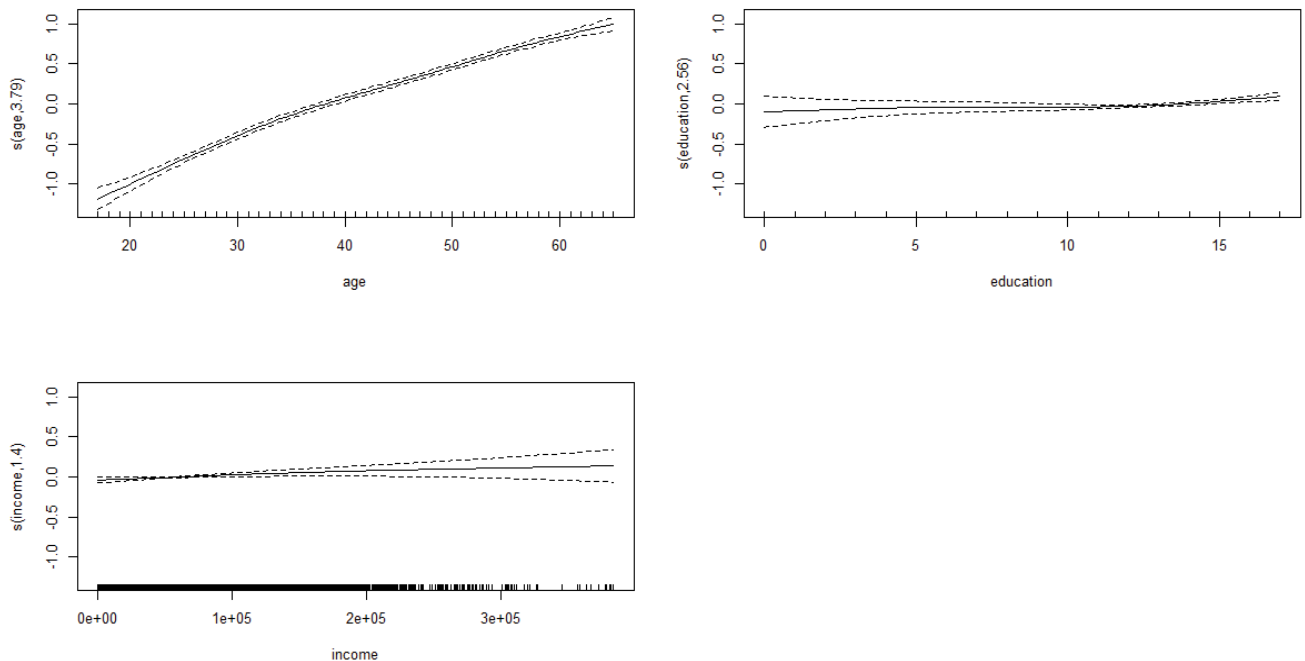


Figure 2 Smooth function estimates and 95% confidence bands for the numerical covariates (age, education, income) in the hypertension equation.

natures and strengths of the relations between them. As such, future studies interested in the incidence of several diseases in a given population, should consider our presented modeling

framework, as opposed to the single equation framework as has been accustomed in the literature.

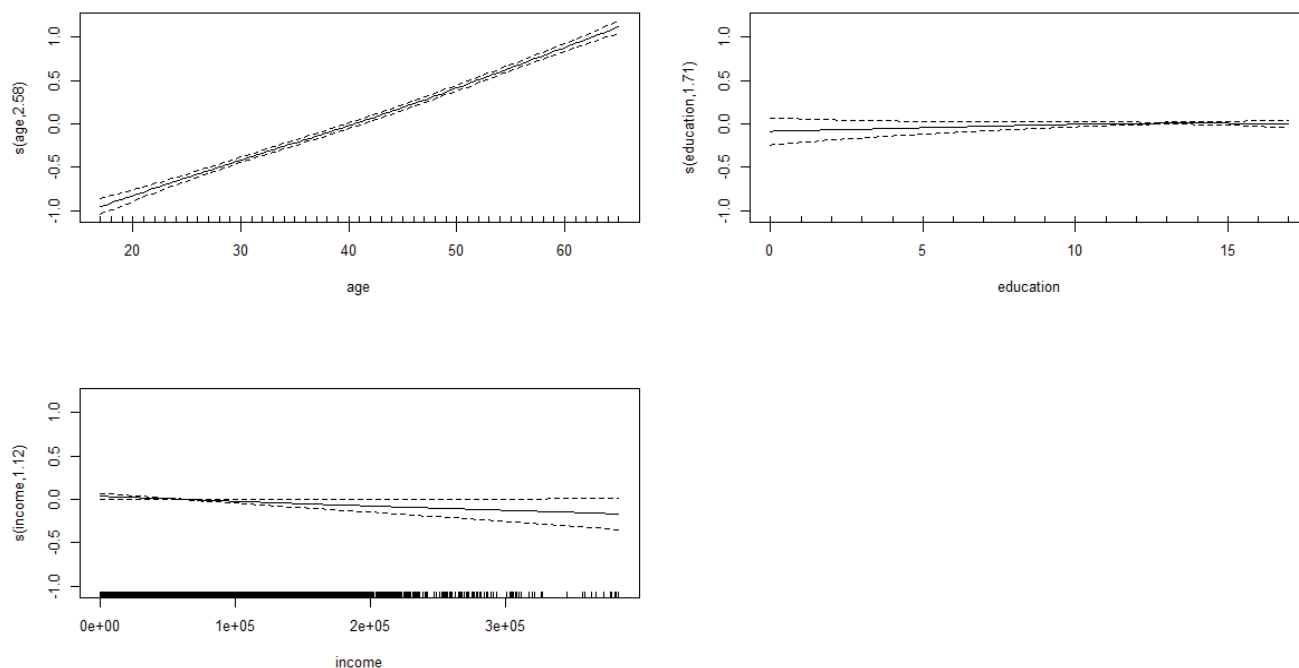


Figure 3 Smooth function estimates and 95% confidence bands for the numerical covariates (age, education, income) in the hyperlipidaemia equation.

References

- World Health Organization (2009) Global health risks: mortality and burden of disease attributable to selected major risks.
- World Health Organization (2015) The utility of estimates for health monitoring and decision-making: global, regional and country perspectives. Background paper. Technical meeting WHO, Glion sur Montreux, Switzerland pp: 1-11.
- Omran AR (1971) The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Mem Fund Q* 49: 509-538.
- Ganz ML, Wintfeld N, Li Q, Alas V, Langer J, et al. (2014) The association of body mass index with the risk of type 2 diabetes: a case-control study nested in an electronic health records system in the united states. *Diabetol Metab Syndr* 6: 50-53.
- Pi-Sunyer FX (2002) The obesity epidemic: pathophysiology and consequences of obesity. *Obes Res* 10: 97-104.
- Krauss RM, Winston M, Fletcher BJ, Grundy SM (1998) Obesity impact on cardiovascular disease. *Circulation* 98: 1472-1476.
- Pan WH, Flegal KM, Chang HY, Yeh WT, Yeh CJ, et al. (2004) Body mass index and obesity-related metabolic disorders in taiwanese and us whites and blacks: implications for definitions of overweight and obesity for asians. *Am J Clin Nutr* 79: 31-39.
- Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, et al. (2006) Obesity and cardiovascular disease pathophysiology, evaluation, and effect of weight loss. *Arterioscler Thromb Vasc Biol* 26: 968-976.
- Bays HE, Chapman R, Grundy S, SHIELD Investigators' Group (2007) The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. *Int J Clin Pract* 61: 737-747.
- Schmiegelow MD, Andersson C, Køber L, Andersen SS, Norgaard ML, et al. (2014) Associations between body mass index and development of metabolic disorders in fertile women-a nationwide cohort study. *J Am Heart Assoc* 3: e000672.
- Madden DDP (2006) Body mass index and the measurement of obesity.
- Pi-Sunyer F, Becker DM, Bouchard C, Carleton RA, Colditz GA, et al. (1998) Clinical guidelines on the identification, and treatment of overweight and obesity in adults. The Evidence Report-National Institutes of Health.
- Garber A (2012) Obesity and type 2 diabetes: which patients are at risk? *Diabetes Obes Metab* 14: 399-408.
- Flegal KM, Carroll MD, Kit BK, Ogden CL (2012) Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. *JAMA* 307: 491-497.
- Cawley J, Meyerhoefer C (2012) The medical care costs of obesity: an instrumental variables approach. *J Health Econ* 31: 219-230.
- Wang Y, Beydoun MA, Liang L, Caballero B, Kumanyika SK (2008) Will all americans become overweight or obese? estimating the progression and cost of the us obesity epidemic. *Obesity* 16: 2323-2330.
- World Health Organization (2005) Chronic diseases and their common risk factors. Geneva: WHO pp: 1-3.
- Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN (2009) Hyperglycemic crises in adult patients with diabetes. *Diabetes Care* 32: 1335-1343.
- Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, et al. (2016) Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the global burden of disease study 2013. *BMJ* 354: 1-10.
- Sone H, Yamada N, Mizuno S, Ohashi Y, Ishibashi S, et al. (2004) Requirement for hypertension and hyperlipidemia medication in us and japanese patients with diabetes. *Am J Med* 117: 711-712.
- Arguedas JA, Leiva V, Wright JM (2013) Blood pressure targets for hypertension in people with diabetes mellitus. *Cochrane Database Syst Rev* 10: 1-47.
- James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, et al. (2014) 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the eighth joint national committee (jnc 8). *JAMA* 311: 507-520.
- Brunstrom M, Carlberg B (2016) Effect of antihypertensive treatment at different blood pressure levels in patients with diabetes mellitus: systematic review and meta-analyses. *BMJ* 352: 1-10.
- Swanton K, Frost M, Maryon-Davis A (2007) Lightening the load: tackling overweight and obesity: a toolkit for developing local strategies to tackle overweight and obesity in children and adults. National Heart Forum pp: 1-195.
- Carroll W (2011) Hypertension in America: estimates for the U.S. civilian noninstitutionalized population, age 18 and older, 2008. Medical Expenditure Panel Survey.
- World Health Organization (2006) Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: Report of a who consultation.
- Butalia S, Kaplan GG, Khokhar B, Rabi DM (2016) Environmental risk factors and type 1 diabetes: Past, present, and future. *Can J Diabetes* 40: 586-593.
- Tuomi T, Santoro N, Caprio S, Cai M, Weng J (2014) The many faces of diabetes: a disease with increasing heterogeneity. *The Lancet* 383: 1084-1094.
- Polonsky KS (2012) The past 200 years in diabetes. *N Engl J Med* 367: 1332-1340.
- American Diabetes Association (2013) Economic costs of diabetes in the U.S. in 2012. *Diabetes Care* 36: 1033-1046.
- Stewart WF, Ricci JA, Chee E, Hirsch AG, Brandenburg NA (2007) Lost productive time and costs due to diabetes and diabetic neuropathic pain in the us workforce. *J Occup Env Med* 49: 672-679.
- <https://www.niddk.nih.gov/health-information/health-statistics/Pages/default.aspx#category=diabetes>
- Carretero OA, Oparil S (2000) Essential hypertension part i: definition and etiology. *Circulation* 101: 329-335.
- Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, et al. (2012) Heart disease and stroke statistics-2012 update a report from the american heart association. *Circulation* 125: e2-e220.
- Centers for Disease Control (2011) High blood pressure fact sheet: Prevalence of hypertension.
- Lackland DT, Weber MA (2015) Global burden of cardiovascular disease and stroke: hyper-tension at the core. *Can J Cardiol* 31: 569-571.
- Mente A, O'Donnell M, Rangarajan S, Dagenais G, Lear S, et al. (2016) Associations of urinary sodium excretion with cardiovascular events in individuals with and without hypertension: a pooled analysis of data from four studies. *The Lancet*.

- 38 Wright JS (2013) Million hearts™: preventing a million heart attacks and strokes through public-private collaboration. *Future Cardiol* 9: 305-307.
- 39 Stone MS, Martyn L, Weaver CM (2016) Potassium intake, bioavailability, hypertension, and glucose control. *Nutrients* 8: 444-445.
- 40 Nagele E, Jeitler K, Horvath K, Semlitsch T, Posch N, et al. (2014) Clinical effectiveness of stress-reduction techniques in patients with hypertension: systematic review and meta-analysis. *J Hypertens* 32: 1936-1944.
- 41 Go AS, Bauman MA, Coleman King SM, Fonarow GC, Lawrence W (2014) An effective approach to high blood pressure control: a science advisory from the american heart association, the american college of cardiology, and the centers for disease control and prevention. *J Am Coll Cardiol* 63: 1230-1238.
- 42 World Health Organization (2016) A global brief on hypertension: silent killer, global public health crisis. *World Health Day* pp: 1-40.
- 43 Keenan NL, Rosendorf KA, National Center for Chronic Disease Prevention and Health Promotion, National Center for Health Statistics (2011) Prevalence of hypertension and controlled hypertension-United States, 2005-2008. *Morbidity and Mortality Weekly Report* pp: 94-97.
- 44 Harrison L (2014) Hypertension er visits surge 25% in five years. *Medscape*.
- 45 Heidenreich PA, Trogon JG, Khavjou OA, Butler J, Dracup K, et al. (2011) Forecasting the future of cardiovascular disease in the united states a policy statement from the american heart association. *Circulation* 123: 933-944.
- 46 Chait A, Brunzell J (1990) Acquired hyperlipidemia (secondary dyslipoproteinemias). *Endocrin and Metabolism Clinics of North America* 19: 259-278.
- 47 Chou R, Dana T, Blazina I, Daeges M, Bougatsos C (2016) Screening for dyslipidemia in younger adults: A systematic review for the us preventive services task force-screening for dyslipidemia in younger adults. *Ann Inter Med* 165: 560-564.
- 48 Mattar M, Obeid O (2009) Fish oil and the management of hypertriglyceridemia. *Nutr Health* 20: 41-49.
- 49 Raal F (2008) Management of dyslipidaemia. *Continuing Medical Education*.
- 50 Ezzati-Rice TM, Rohde F, Greenblatt J (2008) Sample design of the Medical Expenditure Panel Survey household component, 1998-2007. Agency for Healthcare Research and Quality, US Department of Health & Human Services.
- 51 Chib S, Greenberg E (1998) Analysis of multivariate probit models. *Biometrika* 85: 347-361.
- 52 Wojtys M, Marra G, Radice R (2016) Copula regression spline sample selection models: the r package semiparsamplesel. *J Stat Softw* 71: 1-66.
- 53 R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.