

# Bayesian Mixed-effects Polychotomous Response Model with Application to Diverse Population Collaboration (DPC) Data

Fang Yang<sup>1</sup>, Xu-Feng Niu<sup>2</sup> and Jianchang Lin<sup>3\*</sup>

<sup>1</sup>Novartis Pharmaceuticals, Cambridge, MA 02139, USA

<sup>2</sup>Department of Statistics, Florida State University, Tallahassee, Florida 32306, USA

<sup>3</sup>Takeda Pharmaceuticals, Cambridge, MA 02139, USA

## Abstract

Polychotomous response models are commonly used in the clinical trials to analyze categorical or ordinal response data. Motivated by investigating of relationship between BMI categories and several risk factors, we carry out the application studies to examine the impact of risk factors on BMI categories, especially for categories of "Overweight" and "Obesities". In this study, we apply the Bayesian methodology through a mixed-effects polychotomous response model to the Diverse Population Collaboration (DPC) dataset. Using the mixed-effects Bayesian polychotomous response model with uniform improper priors, we would get similar interpretations of the association between risk factors and BMI, which are in great agreement with the results documented in literature. Our application showed that the Bayesian mixed-effects polychotomous response model with improper priors is a very useful statistical technique for solving real word problems.

**Keywords:** Bayesian hierarchical model; Mixed-effects; Polychotomous response; Ordinal data; Uniform Improper priors; Diverse population collaboration (DPC) data

## Introduction

Evidence in literature shows that there is an increasing prevalence of Overweight and Obesity across the entire world, especially in western counties this increasing prevalence of obesity is more significant [1-7]. Yet many researches have been done, using the body mass index score directly to investigate the relationship between obesity and some risk factors that may cause overweight and obesity. In this paper, instead of using the body mass index score, we will use a categorized body mass scores, and the Bayesian polychotomous mixed-effects model we introduced in previous chapter will be used on the new body mass index categories, to explore and examine the relationship between BMI and some of the risk factors concerning people's demographic and physical characteristics. The main reason we use the BMI categories for analysis is that this definition of categories was very commonly used in the literature and has been standardized by the National Institutes of Health. It is easy for people to understand what risk factors and how these risk factors affect people that fall in these categories, especially the higher categories of "Overweight", and "Obese". Besides the above reasons, the polychotomous response model is a very powerful and useful tool to analyze these BMI category data. When the logit link function is used in the model, the regression coefficients can be transformed to odds ratios thus is easy to interpret and understand. The study results could lead to a better understanding on the causes and preventions of becoming overweight and obese.

## Body mass index, obesity and risk factors

**Definition of body mass index and categorization:** Body mass index (BMI) or Quetelet index is defined by individual's body weight (kg) divided by the square of the height ( $m^2$ ), it is proposed by Quetelet to describe the relationship between the body weight and body stature in humans [8]. Given the properties that BMI is a simple, inexpensive, safe and practical measurement to acquire, it is now used widespread in most aspects of public health and epidemiology as a reasonable measure of a person's "fatness" or "thinness". This simple measurement also allow health professionals to easily assess the over and under weight problems in patients.

According to National Institutes of Health and the World Health Organization, a proper classification of BMI is: "Underweight" defined as  $BMI < 18.5$ ; "Normal weight" defined as  $BMI \geq 18.5$ , but less than 25; "Overweigh" defined as  $BMI \geq 25$  but less than 30; "Obese" defined as  $BMI \geq 30$ . Some investigators break up the "Obese" group by "Obese" ( $BMI \geq 30$  but less than 35) and "Severe obese" ( $BMI \geq 35$ ) [9].

The BMI definition itself is debated by professionals that it is not an accurate measurement if considering the distribution of muscle and bone mass of human body. It is argued that the error in body mass index is large and it is sometimes not useful in addressing health issues [10]. Evidence of the slightly lower effect of morbidity and mortality risk in the BMI overweight range on different age population group [11] also indicates that the cut-off point  $25 \text{ kg}/m^2$  for Overweight category are debatable. Despite the controversial discussion about accuracy of BMI and classification of BMI categories, the traditional classification is still considered proper in most cases, and we will use this classification through our study.

## Prevalence of obesity

Studies of the four separate national surveys: NHES I (1960-1962), NHANES I (1971-1974), NHANES II (1976-1980) and NHANES III (1988-1991), show that there is a significantly increasing trend for the prevalence of overweight and obesity in the United States for both children and adults [2,3]. The data indicate that there is a slightly increasing of Obese prevalence in NHES I and NHANES I surveys, comparing to a significant increase in HNANES II and NHANES III surveys [3,4].

\*Corresponding author: Yang F, Novartis Pharmaceuticals, Cambridge, MA 02139, USA, Tel: (850) 228-7421; E-mail: [fang.yang@novartis.com](mailto:fang.yang@novartis.com)

Received April 20, 2017; Accepted April 25, 2017; Published April 28, 2017

**Citation:** Yang F, Niu XF, Lin J (2017) Bayesian Mixed-effects Polychotomous Response Model with Application to Diverse Population Collaboration (DPC) Data. J Biom Biostat 8: 346. doi: [10.4172/2155-6180.1000346](https://doi.org/10.4172/2155-6180.1000346)

**Copyright:** © 2017 Yang F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

According to Campbell et al. [1,2], from NHANES II to NHANES III during the period 1988 to 1991, national survey data indicates approximately one third of adults 20 years of age or older were estimated to be overweight. Among all US adults, 31% of men and 35% of women are estimated to be overweight. The age-adjusted overweight prevalence increased 8% from 14.5% to 22.5% between the 1976-1980 and 1988-1994 periods. Mean level of BMI increased from 25.3 to 26.5 and the mean weight of adults aged 20 to 74 years increased about 3.6 kg between NHANES II to NHANES III [1].

The changes in obese prevalence also are significant and these changes of obese may related to the changes in the distribution of age groups in United States. Trioano et al. show that from previous national surveys to NHANES III, younger children and part of adolescents showed no significant changes in BMI. BMI for men of 20-49 years old have a small consistent increase over all levels of BMI categories, while BMI for men of 50-74 years old increase significantly in categories of overweight and obese, that is to say that the level of obese is become heavier. For women, the changes are greater than men, and the gradual increase is across the entire BMI distribution [2].

Overall, the BMI distribution is shifting upward in some extent among all age-sex groups, for both genders and across all races and smoke status, and the most marked change is on the upper end of the distribution for those overweight and obese people, with most of the rise occurring within the past decades [3,4]. Right now, approximately more than half of US adults are overweight (BMI>25) and nearly one quarter of US adults are obese (BMI>30) [3]. Other studies in England [5] and a research about Europe obesity epidemic [6] indicate the trend of an increasing prevalence of obesity not only in US but all over the world. In fact, The World Health Organization statistics showed that in 2005, approximately 1.6 billion adults over age 15 years were overweight and at least 400 million adults were obese. The World Health Organization further projects that by 2015, there will be approximately 2.3 billion adults being overweight and over 700 million adults being obese [7]. Obese is becoming a major health problem not only in United State but also a world-wide challenge.

### Consequences of being overweight and obese

Obesity and Overweight are major risk factor for some chronic disease such as cardiovascular disease, Diabetes, certain type of cancers etc. Those serious health consequences of overweight and obesity are well documented. Cardiovascular disease including high blood pressure, coronary heart disease and stroke are now the leading causes of death in America. According to American Heart Association, in year 2006, it is estimated that 81,000,000 people in the United States have one or more forms of cardiovascular disease. Framingham study shows that the risk of CVD increased with increasing degree of obesity. Incidence of congestive heart failure for people younger than 50 years old increased 2.5-3 times from leanest to heaviest patients. The incidence for atherothrombotic stroke also increased in obese group [12]. Rexrode et al. discovered that the relative risk increased from 1.75 to 2.37 in women with BMI of 27 kg/m<sup>2</sup> to 28.9 kg/m<sup>2</sup>, and women with BMI of 32 kg/m<sup>2</sup> or more compared to BMI less than 21 kg/m<sup>2</sup> [13]. All research evidences indicate that the degree of obesity is an important predictor of CVD incidence.

Obesity is also an important risk factor in the development of insulin resistance, which can lead to type I and type II diabetes. Research found that the relative risk can increase from 11.3 to 17.3 when weight gain increased from 20 kg to more than 35 kg just in women. In men and children, the relative risk also increases with gaining weight [14-16].

Certain type of cancers were associated consistently with overweight and obesity across all case-control studies and cohort studies, especially cancers of the breast and endometrium in postmenopausal women, and cancers of colon and kidney are more prevalent among overweight and obese men. Moreover, studies show controversial associations between obesity and prostate cancer, pancreatic cancer, colorectal cancer. So being overweight or obesity may be a risk factor but not necessarily a dominant one [17-19].

Other disease like kidney stones are the major cause of morbidity, its formation is documented related to obesity [20]. Dementia, particularly Alzheimer disease in women, asthma (reported in children and women), fat liver hepatitis, knee osteoarthritis, infertility, miscarriage, and other reproductive disorder are reported to positively associate with overweight and obesity [21-25].

Besides all these health problems, overweight and obese people may also have social difficulties due to the society's discrimination against the adiposity. Studies have showed that an increase in depression and suicide among those obese people [26].

Furthermore, for the mortality cause by obesity, study shows that there are approximately 280000 of annual deaths attributable to obesity among US adults [27]. Obesity has become a major cause of mortality and morbidity in United States.

### Risk factors that can affect BMI

There are a lot of factors that can affect the level of BMI. All the risk factors can be classified into demographic factors, physical conditions, personal habits, geographic factors and genetic factors. Next, we will discuss each aspect in details.

Demographic factors include sex, age, race etc. It is well known that women have higher body fat level than men. The change of obese prevalence in the national health surveys also shows that the increase of obese is more significant for women than men [1,2]. The relationship of age with BMI in adults is not a linear one. The data from USA NHANES II and NHANES III show that the percentiles of BMI first increase with age get older, then after some middle age, the lower BMI percentiles start to keep at a constant level and higher BMI percentiles start to decrease as age get older. The decreasing rate is more rapid for very higher BMI percentiles such as the 95th percentiles or higher. The average age when the change of the BMI percentiles start is around 40-45 years old [28]. This factor can confound with other factors. Older people tend to have lower physical activities, which related with high level of body fat. Some disease caused by old age and the medicine for the disease sometimes can also cause the weight gain.

After controlling for age and sex, researchers found no significant association between race and BMI. However, there is a slightly higher prevalence of overweight and obese among black women than white women [8,28].

Physical conditions are the factors that relate to patients current health issues, such as the level of cholesterol, systolic blood pressure and whether have any disease that relate to the change of the BMI. From the literatures, we get the conclusion that the prevalence of high blood pressure and high blood cholesterol level are significantly increase as BMI increases for people 60 years old or younger. After controlling for age, race, education and smoking, the relationship of BMI with high blood pressure is still significant [29], which means BMI increases when systolic blood pressure increases.

Personal habits include smoking status, amount of physical

activities and personal diet habit etc. The impact of smoke on obesity is the most discussed issue because the relation between smoking and obese is complicated and controversial. On one hand, smoking tends to increasing energy expenditure, induce appetite and increase lung diseases, hence in some normal and light smokers, the mean weight is lower than non-smokers. On the other hand, quit smoking can lead to weight gain, and heavier smokers are more related to depression, poor diet and low physical activity. It leads to the result that there is higher mean weight among heavier smoker than other level of smokers [30]. And we all know that low physical activity is related to low speed of digest and increasing of cumulative body fat. People with low energy diet have relatively lower mean body weight than people with high energy diet.

Geographic factors refer to the neighborhood environment. Walkable neighborhoods, pedestrian-friendly designed neighborhoods are associated with lower mean BMI. Districts that need longer commutes may associate with higher BMI [31].

For genetic factor, there is a list of single-gene defects that can lead to human obesity including Leptin deficiency and Melanocortin 4 receptor mutations etc., which can cause severe obesity. And among the twins, monozygotic twins have higher body weight than dizygotic twins. But most of the time it is the interaction between genetic factor and environmental factors over time that cause the obesity [32,33].

All the above factors affect the distribution of BMI not independently but it is the combined effect that causes the change in the distribution of BMI. In this paper we use BMI categories to examine the association of being overweight and obesity with some risk factors such as smoking, health conditions as well as some demographic characteristic of patients such as age, sex, race etc.

In the next section, we will introduce the Diverse Population Collaborations (DPC) database, which is our source of data. The Bayesian mixed-effects polychotomous response model will be conducted on the selected data sets and results from this application will be discussed and interpreted.

**Diverse population collaboration (DPC) data:** After going through all the variables for each study in the DPC database, we choose seven variables as our potential explanatory variables and we want to explore their relationship with BMI categories. These seven covariates are age, total serum cholesterol level, systolic blood pressure, diabetes status, current smoke status, race and gender. Six studies are chosen from the DPC database that contains all these seven covariates: the Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (CHS), the Evans County Study (Evans), the Lipid Research Clinics Prevalence Study (LRC), the First National Health and Nutrition Examination Survey Epidemiologic Follow up Study (NHANES I), and the Second National Health and Nutrition Examination Survey and Mortality Follow-up (NHANES II). All six chosen studies are conducted in United States; the majority of the participants are Caucasian and African American. Since very few participants are minorities such as Hispanic, Asian, we exclude this few minorities in our study for easy interpretation.

For each study, we obtained the heights and weights of participants from the baseline visits, using this data we can calculate the body mass index directly from height/weight<sup>2</sup>. However, it is the BMI categories that we are interested in, so we classify the BMI according to the NIH standard that we described in the previous section. The BMI categories coding is listed in Table 1. We coded the BMI categories from 1 to 5

as body mass index increases from the leanest category to the heaviest category. Categorical risk factors are coded as in Table 2.

Table 3 lists the demographic statistics of these six studies. From Table 3, the variable age in ARIC and CHS studies has the smallest variance since CHS study has the oldest participants in average. This is mainly because these two studies are prospective studies that targeted to specific subgroups of the whole population: ARIC participants aged between 45 and 64 years old, and CHS study monitored participant's coronary heart disease and stroke in adults aged 65 years and older. All six studies have comparative gender and race ratios. There are certain numbers of people in each study that have diabetes or are currently smokers. Total serum cholesterol level and Systolic blood pressure among all six studies are relatively consistent except for Evans study, which has the highest systolic blood pressure in average. The BMI distribution is approximately normal with least number of people in the lower and upper level of BMI categories, and most number of people around the middle categories. For easy interpretation of the regression coefficients and programming convenience, we will use the logistic link function in the polychotomous response model through our application.

From the demographic table, we know that there indeed have some varieties among these six studies, for example, the different age groups. Also we are only investigating seven of the many risk factors that can have effects on BMI. There are definitely some effects that cannot explain by these seven variables. Hence we will include a random effect in our polychotomous response model to incorporate the cluster variety and other unexplained effects on BMI.

Before we fit the model on these data, we gave a bar plots of the distribution of the BMI categories by each measurement of the risk factors. By looking at these bar plots, we would have some ideas of the distributions of BMI categories in associated with these risk factors. Because the exponential value of the coefficient of the regression in polychotomous response model can be interpreted as the cumulative odds ratio of corresponding risk factor increased by one unit, the probabilities of the BMI categories from the bar plot can give some information of the signs of the odds ratios for these risk factors. The bar plots are listed in Figures 1-7. Let us first look at the continuous predictors: age, total serum cholesterol level and systolic blood pressure. From Figures 2 and 3, we could clearly see that the probabilities for BMI in "Underweight or above" and "normal weight or above" increase when values of cholesterol and blood pressure increase, and the probabilities for BMI in "Overweight or above", "Obese or above" and "Severe obese" also increase when values of cholesterol level and blood pressure increase. For the distribution of BMI categories by age,

Body Mass Index		
Category	BMI range	Coding
Underweight	BMI<18.5	1
Normal weight	18.5 ≤ BMI<25	2
Overweight	25 ≤ BMI<30	3
Obese	30 ≤ BMI<35	4
Severe obese	35 ≤ BMI	5

Table 1: BMI categories coding.

Variable	Code
Gender	0=Male, 1=Female
Race	0=White, 1=Black
Diabetes	0=No diabetes, 1=With diabetes
Smoke	0=None smoker, 1=Smoker

Table 2: Code for categorical predictors.

	ARIC	CHS	Evans	LRC	NHANES I	NHANES II
Observations	15436	5665	2637	8459	12787	8852
Age	54.18 ± 1.7	72.8 ± 5.6	50.48 ± 12.3	47.79 ± 12.1	49.29 ± 15.4	54.42 ± 12.3
Gender						
Male	8485	3244	1384	3903	7583	4688
Female	6951	2421	1253	4556	5204	4164
Race						
Caucasian	11423	4812	1659	8160	10957	7902
Africa American	4013	853	978	299	1830	950
Diabetes status						
None diabetes	13923	4739	2532	8311	12304	8341
With diabetes	1512	926	105	148	483	511
Current smoke status						
None smoker	11400	4980	1716	5477	8611	5970
Smoker	4036	685	921	2982	4176	2882
Cholesterol level	215 ± 42	211 ± 39	215 ± 45	224 ± 47	222 ± 48	226 ± 49
Systolic blood pressure	121 ± 19	136 ± 22	152 ± 33	126 ± 19	135 ± 24	133 ± 22
BMI by categories						
Underweight	137 (0.89%)	24 (0.42%)	70 (2.65%)	111 (1.31%)	426 (3.33%)	226 (2.55%)
Normal weight	4987 (32.31%)	2173 (38.36%)	1182 (44.82%)	3602 (42.58%)	5948 (46.52%)	3852 (43.52%)
Overweight	6076 (39.36%)	2387 (42.14%)	820 (31.10%)	3440 (40.67%)	4224 (33.03%)	3214 (36.31%)
Obese	2835 (18.37%)	827 (14.60%)	363 (13.77%)	1013 (11.98%)	1530 (11.97%)	1098 (12.40%)
Severe obese	1401 (9.08%)	254 (4.48%)	202 (7.66%)	293 (3.46%)	659 (5.15%)	462 (5.22%)

Table 3: Demographic statistics for six studies.

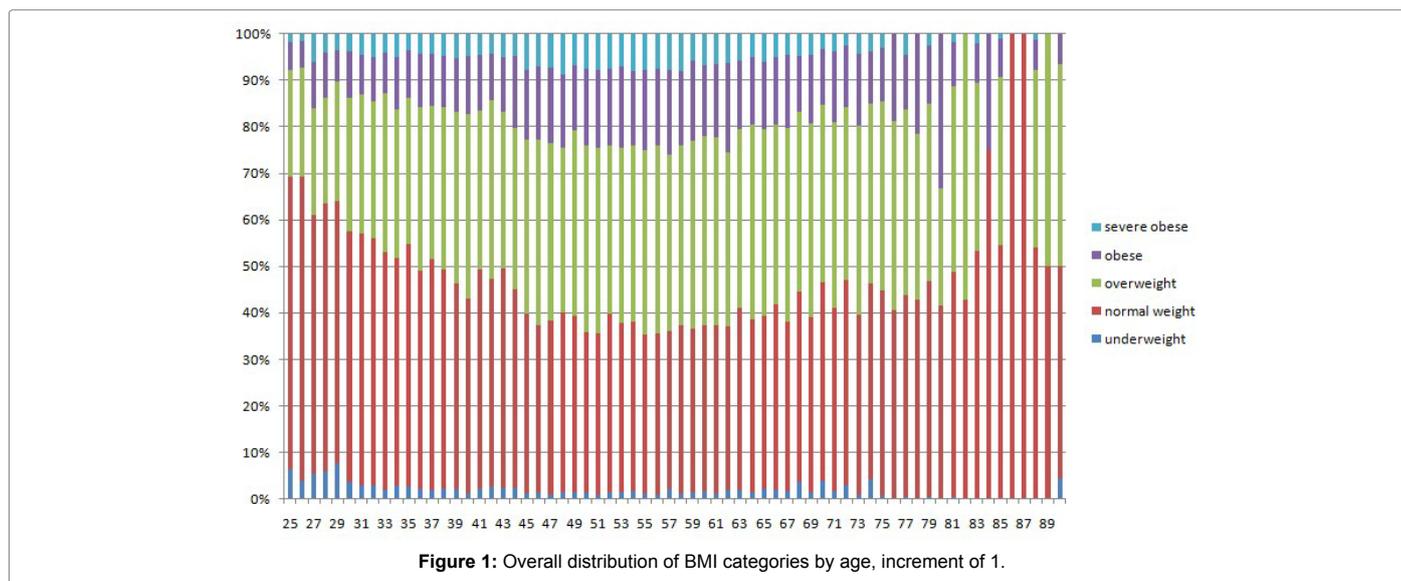


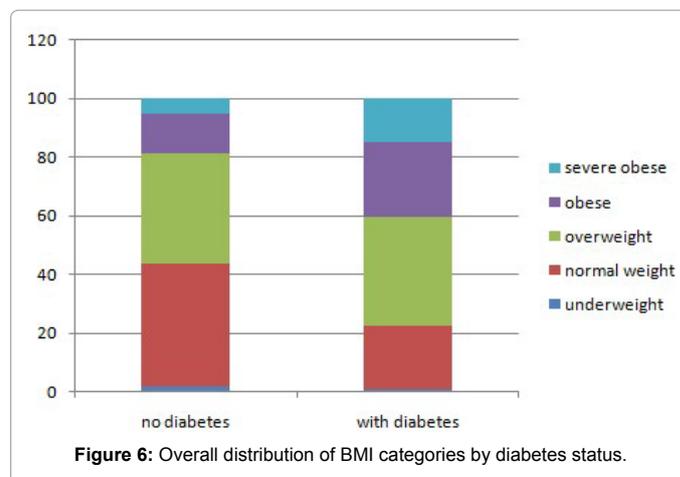
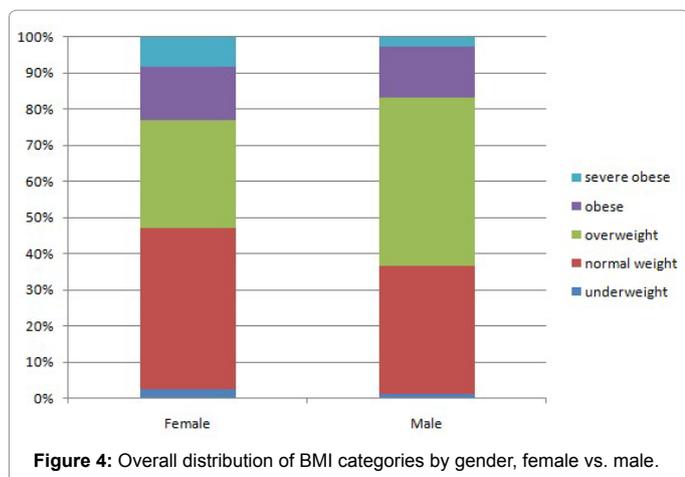
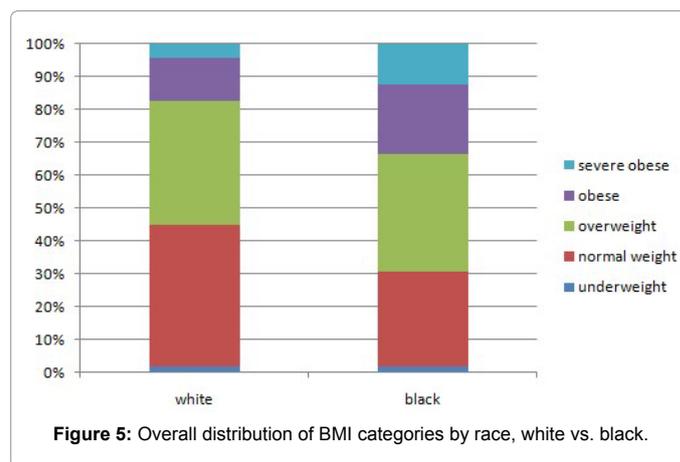
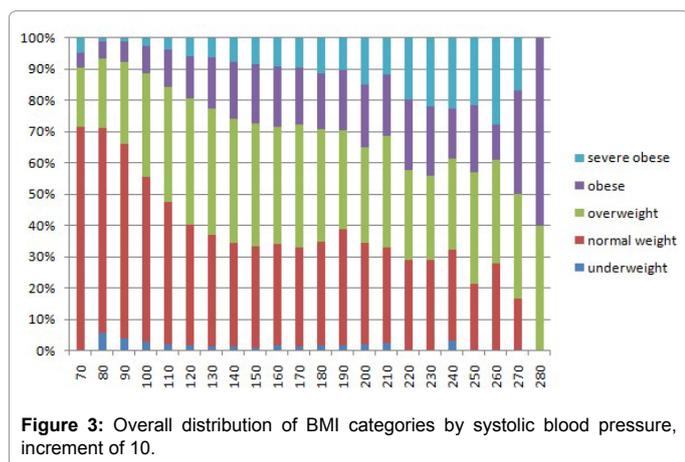
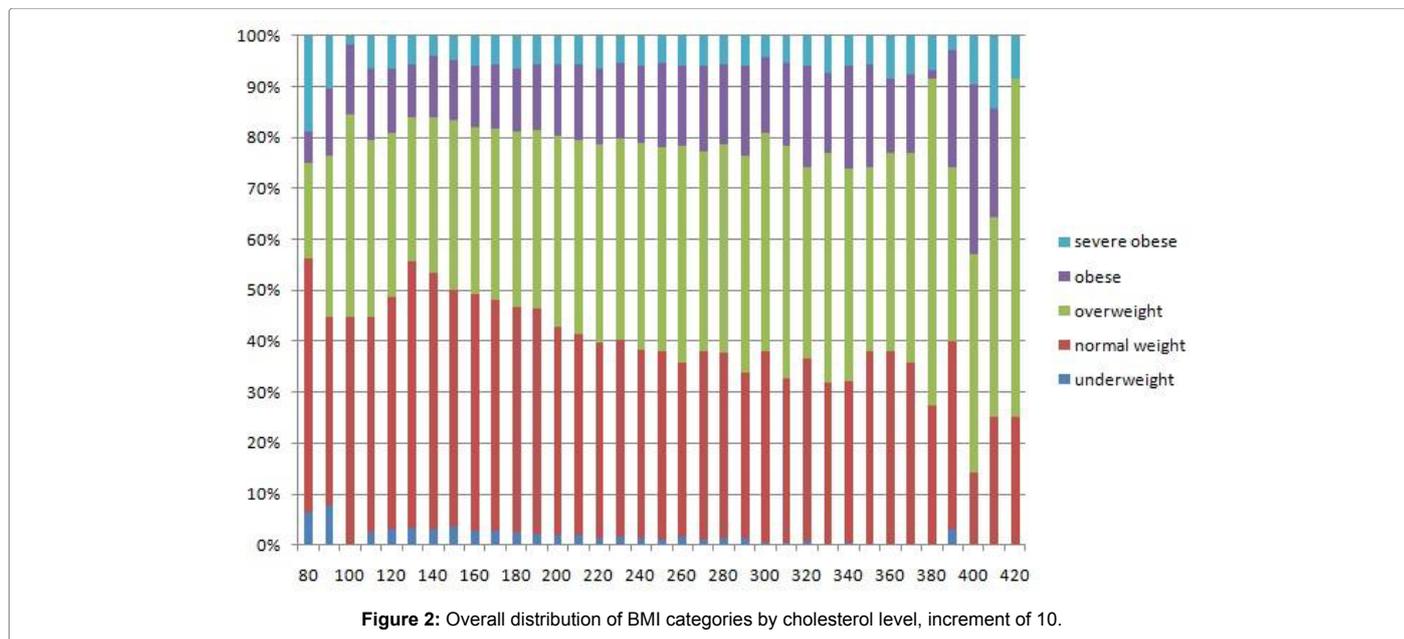
Figure 1: Overall distribution of BMI categories by age, increment of 1.

the situation is a little complicated and there is a curve likely effect of age on BMI categories. The probability of “Normal weight or above” first increases then decreases; the probability of “Overweight or above”, “Obese or above” and “Severe obese” also first increase then decrease. In the later part of this analysis, we will use second order polynomial regression in the polychotomous response model to fit this curve effect of age on BMI categories.

Then we look at the categorical predictor variables. Data show that Men have more percentage in the level of “Overweight”, however women have more percentages in both the “Obese” and “Severe obese” categories. Black people have consistently higher percentages in the “overweight”, “Obese”, and “Severe obese” categories. Participants diagnosed with diabetes also have higher percentage in the “overweight”, “Obese”, and “Severe obese” categories. Finally,

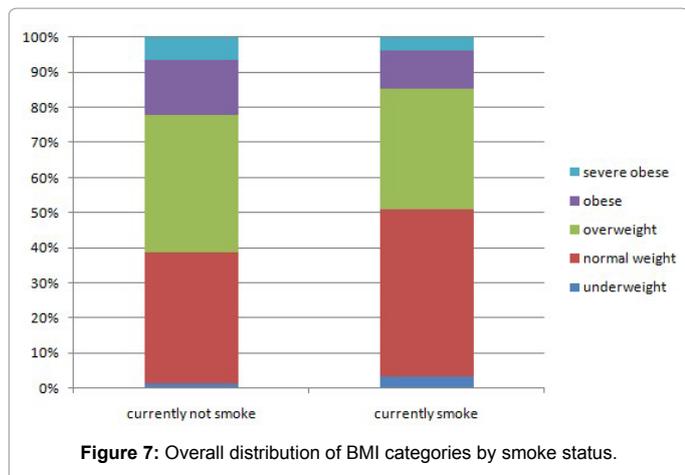
participants that currently smoking has lower percentages in the “overweight”, “Obese”, and “Severe obese” categories. Some of the risk factors maybe confound with each other and even the factor itself may show controversial effect. Since we are assuming the same effect across all the levels of the BMI categories, it is the dominant direction of the effect that we are estimating.

Collinearity is defined as a strong correlation between independent variables. The existence of collinearity may cause insignificance of the independent variable or may cause change of signs from including each individual variable in the model to including both variables in the model. To examining if there exist collinearity among the risk factors, we take an extra step to look at the correlation coefficients of these risk factors and the scatter plot matrix of these variables. The correlation coefficients are showed in Table 4. From the table, we could see that



most of the coefficients are less than 0.2. This means that the linear effects among these seven risk factors are relatively weak. Except for one correlation coefficient between age and systolic blood pressure is 0.342. This moderate positive linear correlation indicates that systolic

blood pressure increases as age increases. The scatter plots did not show strong linear patterns among any pair of the risk factors. Since our sample size is very large, the moderate collinearity between age and systolic blood pressure may not cause any problems in the parameter



	Age	Gender	Race	Diabetes	Smoker	Cholesterol	Blood pressure
Age	1.000						
Gender	0.018	1.000					
Race	0.008	-0.054	1.000				
Diabetes	0.159	-0.004	0.109	1.000			
Smoker	-0.205	0.080	0.027	0.056	1.000		
Cholesterol	0.167	-0.078	-0.033	0.004	-0.013	1.000	
Blood pressure	0.342	0.030	0.142	0.098	-0.084	0.151	1.000

**Table 4: Correlation coefficients of seven risk factors.**

estimating. In the analysis, we use centered values for the continuous risk factors such as age, systolic blood pressure, and total serum cholesterol. This will also reduce the effect of collinearity among risk factors on the estimated results (Figure 8).

**Bayesian mixed-effects polychotomous response model:** The mixed-effects polychotomous response model is defined as following:

$$P(y_{ij} = 1) = \frac{\exp(\theta_1 + Z_{ij})}{1 + \exp(\theta_1 + Z_{ij})}$$

$$P(y_{ij} = 2) = \frac{\exp(\theta_2 + Z_{ij})}{1 + \exp(\theta_2 + Z_{ij})} - P(y_{ij} = 1)$$

$$P(y_{ij} = 3) = \frac{\exp(\theta_3 + Z_{ij})}{1 + \exp(\theta_3 + Z_{ij})} - P(y_{ij} = 2)$$

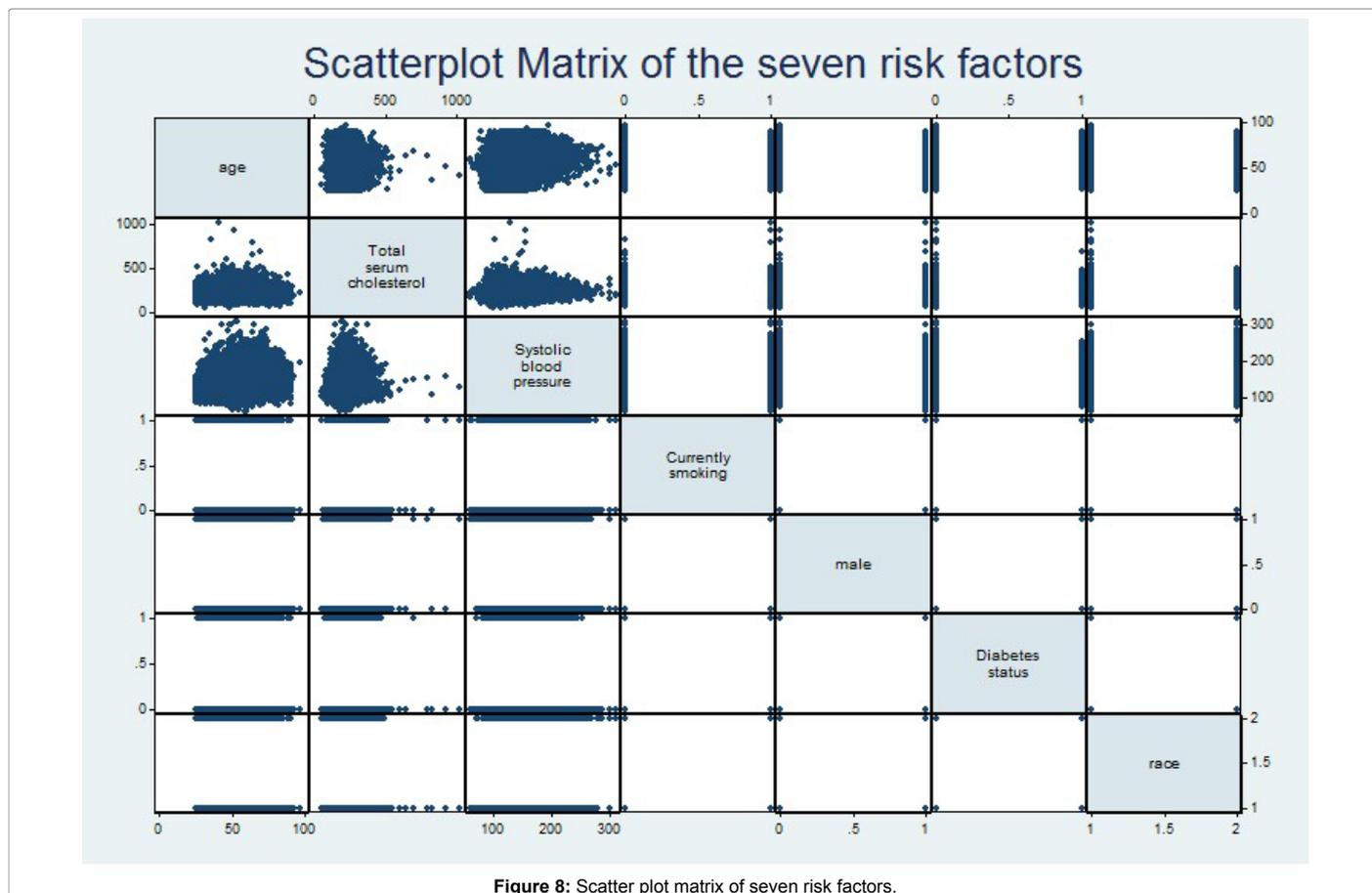
$$P(y_{ij} = 4) = \frac{\exp(\theta_4 + Z_{ij})}{1 + \exp(\theta_4 + Z_{ij})} - P(y_{ij} = 3)$$

$$P(y_{ij} = 5) = 1 - \frac{\exp(\theta_4 + Z_{ij})}{1 + \exp(\theta_4 + Z_{ij})}$$

Where  $y_{ij}$  is the response BMI category variable,  $Z_{ij} = \beta X_{ij}$ ,  $i=1, \dots, 6$ ,  $j=1, \dots, n_i$

In the Bayesian method, prior settings for parameters in the above model are as following:

$$b_i \sim N(0, \eta)$$



$\eta \sim \text{Gamma}(0.001, 0.001)$

each regression coefficient is set to have a uniform improper prior, that is  $\beta_i \sim dflat(\cdot), i=1, \dots, 6$ , where  $dflat(\cdot)$  is the notation in Winbugs for uniform improper prior and in the sampling process Winbugs with set it as 1 to form the conditional posteriors. After setting the priors for the parameters in the model, starting from a set of initial values, the Gibbs Sampling method generates samples of these parameters from their conditional posterior distributions. For each model fitting, we run 1000 iterations for the burn-in process and additional 5000 iterations for the sampling process.

In the analysis, we first fit models using each individual risk factor. If any of the risk factors shows none significant effect, we will drop that factor in our later analysis. Results are showed in Tables 5-12. All the risk factors have significant effect on the BMI categories. For risk factor age, from the bar plot, we know that the effect of age on BMI categories is not linear. Hence we fit an additional model with second order polynomial term. The smaller DIC value showed that the quadratic polynomial model fit the data better.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{age}$	-0.009	6.929E-4	(-0.0106, -0.00079)
$\theta_1$	0	---	---
$\theta_2$	3.695	0.03103	(3.635, 3.759)
$\theta_3$	5.407	0.03227	(5.344, 5.471)
$\theta_4$	6.791	0.03547	(6.723, 6.862)
$\sigma$	3.789	1.069	(2.35, 6.56)

DIC: 133836.

Table 5: Polychotomous model with age effect only.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{age^2}$	-0.00663	7.335E-4	(-0.008, -0.0052)
$\beta_{age}$	9.206E-4	4.903E-5	(8.245E-4, 0.0010)
$\theta_1$	0	---	---
$\theta_2$	3.709	0.03715	(3.646, 3.791)
$\theta_3$	5.427	0.03903	(5.361, 5.515)
$\theta_4$	6.813	0.04261	(6.737, 6.907)
$\sigma$	4.025	1.25	(2.388, 7.059)

DIC: 133490.

Table 6: Polychotomous model with age and age<sup>2</sup> effects.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{chol}$	-0.0041	1.775E-4	(-0.0044, -0.0037)
$\theta_1$	0	---	---
$\theta_2$	3.716	0.03145	(3.661, 3.788)
$\theta_3$	5.437	0.03271	(5.38, 5.511)
$\theta_4$	6.824	0.03562	(6.76, 6.903)
$\sigma$	3.817	1.097	(2.331, 6.597)

DIC: 133466.

Table 7: Polychotomous model with serum cholesterol level effect only.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{diab}$	-0.4874	0.0168	(-0.522, -0.4549)
$\theta_1$	0	---	---
$\theta_2$	3.698	0.0294	(3.643, 3.762)
$\theta_3$	5.43	0.03042	(5.374, 5.498)
$\theta_4$	6.832	0.033373	(6.768, 6.902)
$\sigma$	4.204	1.22	(2.625, 7.337)

DIC: 133079.

Table 8: Polychotomous model with diabetes status effect only.

Then we fit the overall model with all seven risk factors together. An additional model with second-order polynomial effect of age is also fitted. The estimated results are listed in Tables 13 and 14. The smaller DIC value in Table 14 showed that with the additional polynomial terms in the model, the overall fitting is better.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{gender}$	-0.07278	0.0785	(-0.08828, -0.05775)
$\theta_1$	0	---	---
$\theta_2$	3.691	0.0291	(3.634, 3.743)
$\theta_3$	5.401	0.03065	(5.34, 5.459)
$\theta_4$	6.782	0.03425	(6.716, 6.848)
$\sigma$	3.852	1.198	(2.338, 6.601)

DIC: 133922.

Table 9: Polychotomous model with gender effect only.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{race}$	-0.3383	0.0115	(-0.3608, -0.3164)
$\theta_1$	0	---	---
$\theta_2$	3.707	0.03275	(3.648, 3.771)
$\theta_3$	5.439	0.03436	(5.377, 5.507)
$\theta_4$	6.838	0.03775	(6.769, 6.913)
$\sigma$	4.045	1.24	(2.465, 6.954)

DIC: 133092.

Table 10: Polychotomous model with race effect only.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{sbp}$	-0.0207	3.762E-4	(-0.02142, -0.01994)
$\theta_1$	0	---	---
$\theta_2$	3.759	0.03261	(3.693, 3.818)
$\theta_3$	5.537	0.03455	(5.469, 5.603)
$\theta_4$	6.955	0.03803	(6.879, 7.026)
$\sigma$	3.866	1.16	(2.384, 6.628)

DIC: 133019.

Table 11: Polychotomous model with systolic blood pressure effect only.

Simulation parameters	Mean	S.D.	95% Credible interval
$\beta_{smoke}$	0.2456	0.0092	(0.2276, 0.2641)
$\theta_1$	0	---	---
$\theta_2$	3.71	0.03095	(3.648, 3.775)
$\theta_3$	5.434	0.03199	(5.371, 5.501)
$\theta_4$	6.823	0.03541	(6.753, 6.897)
$\sigma$	3.727	1.14	(2.276, 6.497)

DIC: 133019.

Table 12: Polychotomous model with smoke effect only.

Simulation parameters	Mean	S.D.	95% Credible interval	Odd ratio
$\beta_{age}$	0.01305	8.113E-4	(0.01145, 0.01466)	1.0131
$\beta_{gender}$	-0.1018	0.00816	(-0.1178, -0.08556)	0.9032
$\beta_{race}$	-0.259	0.01148	(-0.2813, -0.2364)	0.7718
$\beta_{chol}$	-0.003669	1.857E-4	(-0.0040, -0.0033)	0.9963
$\beta_{sbp}$	-0.01916	4.247E-4	(-0.01997, -0.0183)	0.9810
$\beta_{smoke}$	0.2454	0.009333	(0.2275, 0.2639)	1.2781
$\beta_{diabetes}$	-0.3838	0.01615	(-0.4147, -0.352)	0.6813
$\theta_1$	0	---	---	---
$\theta_2$	3.815	0.03077	(3.757, 3.882)	---
$\theta_3$	5.652	0.03251	(5.591, 5.721)	---
$\theta_4$	7.102	0.03609	(7.033, 7.177)	---
$\sigma$	4.314	1.306	(2.6, 7.442)	---

DIC: 128755.

Table 13: Estimated parameters for six studies' data.

Simulation parameters	Mean	S.D.	95% Credible interval	Odds ratio
$\beta_{age}^2$	0.01596	8.37E-4	(0.01433, 0.01763)	---
$\beta_{age}$	9.93E-4	4.953E-5	(8.966E-4, 0.0011)	---
$\beta_{gender}$	-0.09897	0.008279	(-0.1148, -0.08281)	0.9058
$\beta_{race}$	-0.2577	0.01148	(-0.2801, -0.2356)	0.7728
$\beta_{chol}$	-0.00327	1.843E-4	(-0.00364, -0.0029)	0.9967
$\beta_{sbp}$	-0.01945	4.21E-4	(-0.0203, -0.01864)	0.9807
$\beta_{smoke}$	0.2542	0.0092	(0.2361, 0.2721)	1.2894
$\beta_{diabetes}$	-0.389	0.01632	(-0.4204, -0.3563)	0.6777
$\theta_1$	0	---	---	---
$\theta_2$	3.833	0.03146	(3.771, 3.895)	---
$\theta_3$	5.679	0.03293	(5.615, 5.746)	---
$\theta_4$	7.133	0.03638	(7.062, 7.205)	---
$\sigma$	4.515	1.28	(2.815, 7.59)	---
DIC: 128389.				

**Table 14:** Estimated parameters for six studies' data with second order polynomial term of age effect in the model.

Since the model has the following basic form:

$$y_i \sim \text{Categorical}(\pi_i)$$

$$\gamma_{ij} = P(y_i \leq j) = \pi_{i1} + \pi_{i2} + \dots + \pi_{ij}$$

$$\text{link}(P(y_i \leq j)) = \text{link}(\gamma_{ij}) = \theta_j + \beta X_i + \mu Z_i$$

Where  $\beta_i = (\beta_1, \beta_2, \dots, \beta_p)$ ,  $X_{i=} (X_{i1}, X_{i2}, \dots, X_{ip})$ ,  $i=1, \dots, N$ ,  $j=1, \dots, L$ . If we fix all the other explanatory variable  $X_{i2}, \dots, X_{ip}$ , as constants and set the random effects to be at its mean value 0, then the coefficient  $\beta_1$  for explanatory variable  $X_{i1}$ , can be interpreted as the estimated odds ratio between  $X_{i1}+1$  and  $X_{i1}$  of BMI below or equal to a fixed level. So if an odds ratio is greater than 1, that means participant have higher chance to be in the lower level of BMI with high value of risk factor than that participant have low value of risk factor, which also means that increasing risk factor causes the BMI decreasing when other factors are fixed as constants. On the other hand, if an odds ratio is less than 1, that means that the participant has lower chance to be in the lower level of BMI with high value of risk factor than a participant with low value of the risk factor, which also means that increasing risk factor causes the BMI increasing when other risk factors are fixed as constants.

Hence from the results, we know that given all the other variables as constants, the estimated odds of BMI below or equal to a fixed level for women is 0.9058 times of the estimated odds for men. In other words, women have higher chance of being at the higher level of BMI categories, which is consistent with current research findings that women have higher body fat than men [1,2]. This result is also consistent with the bar plot which shows that women have higher percentages in the categories of "Obese" and "Severe obese" than men.

The estimated odds of BMI below or equal to a fixed level for black people is 0.7728 times of the estimated odds for white people. This result means that black people have higher probabilities at higher level of BMI categories than white people. In other words, the race effect on BIM is significant here when only black and white people are included in our analysis. This result is consistent with the current research findings [8,18].

The estimated odds of BMI below or equal to a fixed level for smokers is 1.2894 times of that for non-smokers. This effect is also consistent with literatures findings that normal to light smokers have average lower mean weight than non-smokers [20]. Although we do not have the exact data about how many cigarettes these smokers had, we do know that the percentage of smokers among the participants

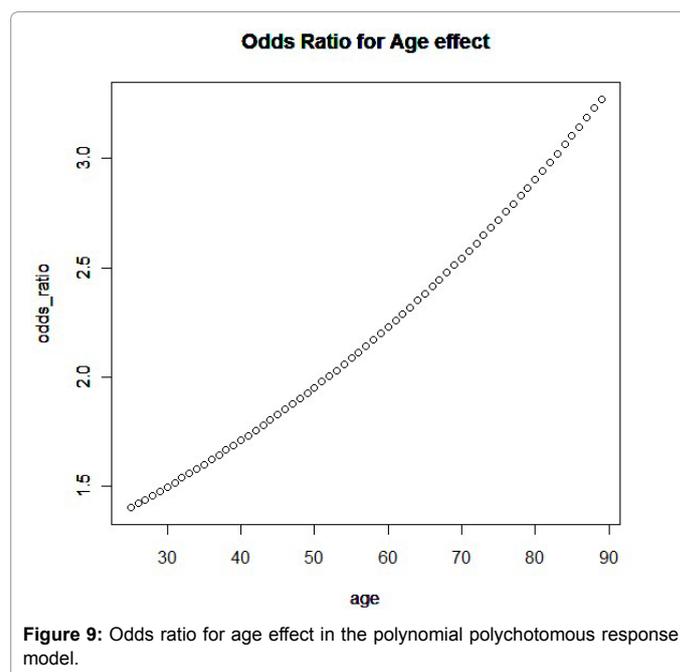
was quite high. Hence we assume that most of smokers were normal to light smokers.

The estimated odds of BMI below or equal to a fixed level for people with diabetes is 0.6777 times of the estimated odds of people without diabetes. This result is also consistent with the fact that obesity can cause insulin resistance thus lead to type I and type II diabetes [14-16].

For continuous predictors, the estimated odds ratio of BMI below or equal to a fixed level can be interpreted as the change of one unit. For age at the linear effect, when age grows older by 1 year, the estimated odds of BMI below or equal to a fixed level is 1.01 times of the estimated odds at one-year younger, after controlling for all the other variables. This result is reasonable for the average age in our six studies that was greater than 45 years old, the age when the BMI starts to decrease. For polynomial age effect, we got the odds ratio of age effect at the following plot Figure 9. From the plot we can see that the odds ratio of age follow a curve. But overall the odds ratio of age is great than 1, which means that participants with older age have more chance falling in the lower BMI categories. This result is consistent with current research findings that people's BMI percentiles decrease rapidly after 45 years old [28].

In the analysis we also confirmed that high cholesterol level and high systolic blood pressure result in increased BMI by looking at the odds ratios for both the variables. The estimated odds of BMI below or equal to a fixed level for participants with high systolic blood pressure is 0.9807 times of that for people with low systolic blood pressure. The estimated odds of BMI below or equal to a fixed level for patients with high serum cholesterol level is 0.9967 times of that for people with low serum cholesterol level. These results are consistent with literature findings [29].

In the model, we also add a random variable to measure the across studies effect that cannot be explained by the selected factors. The estimated variance for the distribution of the random effect is 4.515. Comparing to the relative small scales of the estimated coefficients, we say that this random effect is pretty significant, and hence there is a lot of difference among each studies and there is a large varieties among



**Figure 9:** Odds ratio for age effect in the polynomial polychotomous response model.

the data. We think this random effect may result from the different age groups that each study targeted along with different characteristic variables that confounded with the age variable. Also since we only pick seven risk factors from the whole variable list, there might be other important risk factors that we did not include in the model.

Figures 10 and 11 present the posterior kernel density of the parameters. All the kernel densities are uni-modal distributions and bell shaped. This indicates that the estimates are centered at their true values. We can also see that the density curves are quite smooth, which indicates that 5000 sampling iterations are sufficient for inference.

## Discussion

In this paper, we showed that generalized mixed-effects polychotomous response models using Bayesian hierarchical methods could be effectively used to analyze ordinal data. The Bayesian mixed-effects polychotomous response models were used to examine the relationship between and some risk factors. Results from the application were interpreted in terms of odds ratios. All of the estimated associations of risk factors with BMI categories are in great agreement

with the results documented in literature. The odds ratio we used to interpret the estimated coefficients are easy to be understood and the estimation process is not hard to be implemented using Winbugs. Our application showed that the Bayesian method for mixed-effects polychotomous response models with improper uniform priors is a very useful statistical technique for solving real world problems.

## References

- Campbell SM, Kuczmarski RJ, Flegal KM, Johnson CL (1994) Increasing prevalence of overweight among us adults. The national health and nutrition examination survey. *The Journal of the American Medical Association*.
- Flegal KM, Troiano RP (2000) Changes in the distribution of body mass index of adults and children in the us population. *International Journal of Obesity* 24: 807-818.
- Carroll MD, Flegal KM, Kuczmarski RJ, Johnson CL (1998) Overweight and obesity in the united states: Prevalence and trends, 1960-1994. *The Journal of the American Medical Association* 22: 39-47.
- Dietz WH, Mokdad AH, Serdula MK (1999) The spread of the obesity epidemic in the united states, 1991-1998. *JAMA* 282: 1999.
- Foster K, Browne F, White A, Nicolass G, Carey S (1993) Health survey for england. Her Majesty's Stationary Office.
- Galdeano A (2005) The obesity epidemic in Europe.
- World Health Organization (2016) Obesity and overweight.
- Sepulveda D, Pierson RN, Harris T, Gallagher D, Visser M, Heymseld SB (1996) How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups. *Am J Epidemiol*.
- BMI classification. World Health Organization.
- Devlin K (2009) Do you believe in fairies, unicorns, or bmi? *Mathematical Associations of American*.
- Ian J (2007) Morbidity and mortality risk associated with and overweight bmi in older man and woman. *Obesity*.
- McNamara PM, Hubert HB, Feinleib M, Castelli WP (1982) Obesity as an independent risk factor for cardiovascular disease: A 26-year follow-up of participants in the framingham heart study. *American Heart Association, Circulation*.
- Willett WC, Colditz GA, Stampfer MJ, Rich-Edwards JW, Speizer FE, et al. (1997) A prospective study of body mass index, weight change, and risk of stroke in women. *JAMA*: 1539-1545.
- Stampfer MJ, Manson JE, Hennekens CH, Arky RA, Colditz GA, et al. (1990) Weight as a risk factor for clinical diabetes in women. *Am J Epidemiol*: 501-513.
- Lee IM, Cook NR, Manson JE, Buring JE, Weinstein AR, et al. (2004) Relationship of physical activity vs body mass index with type 2 diabetes in women. *JAMA*.
- Reinehr T, Stachow R, Grabert M, Scholbert E, Rascher W, et al. (2005) The accelerator hypothesis: relationship between weight, height, body mass index and age at diagnosis in a large cohort of 9248 german and austrian children with type 1 diabetes mellitus. *Diabetologia*.
- Carroll KK (1998) Obesity as a risk factor for certain types of cancer. *Lipids*.
- Calle EE, Kaaks R (1998) Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. *Nature Reviews: Cancer*.
- Calle EE, Thun MJ (2004) Obesity and cancer. *Oncogene*.
- Stampfer MD, Taylor EN, Curhan GC (2005) Obesity, weight gain, and the risk of kidney stones. *JAMA*.
- Miia Kivipelto (2005) Obesity and vascular risk factors at midlife and the risk of dementia and alzheimer disease. *Arch Neurol*.
- Brawer R, Brisbon N, Plumb J, Paxman D (2005) The asthma and obesity epidemics: The role played by the built environmenta public health perspective. *The Journal of Allergy and Clinical Immunology*, pages.
- Ian RW, John SL (1990) Fatty liver hepatitis (steatohepatitis) and obesity: An autopsy study with analysis of risk factors. *Hepatology*.

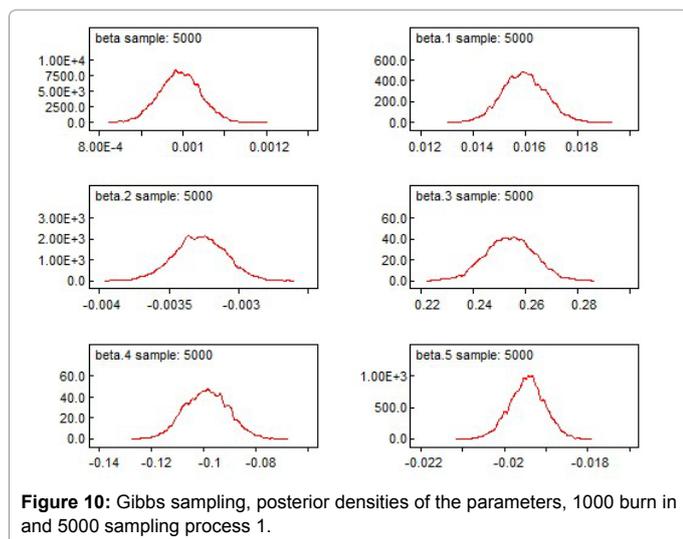


Figure 10: Gibbs sampling, posterior densities of the parameters, 1000 burn in and 5000 sampling process 1.

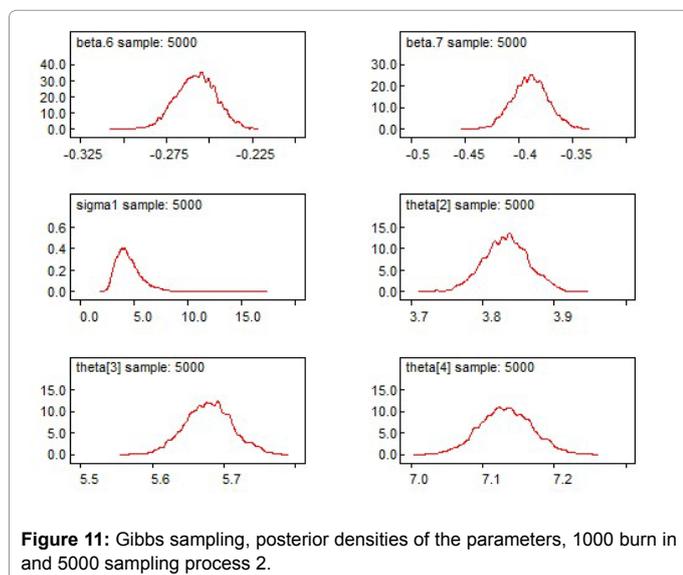


Figure 11: Gibbs sampling, posterior densities of the parameters, 1000 burn in and 5000 sampling process 2.

24. Naimark A, Walker AM, Felson DT, Anderson JJ, Meenan RF (1988) Obesity and knee osteoarthritis. The framingham study. *Ann Intern Med*.
25. Norman RJ, Clark AM (1998) Obesity and reproductive disorders: A review. *Reproduction, Fertility and Development*.
26. Onyike CU, Crum RM, Lee HB, Lyketsos CG, Eaton WW (2003) Is obesity associated with major depression? Results from the third national health and nutrition examination survey. *Am J Epidemiol* 12: 1139-1147.
27. Stevens J, Allison B, Fontaine KR, Manson JE, VanItallie TB (1999) Annual deaths attributable to obesity in united states. *JAMA*.
28. Ogden CL, Flegal KM, Carroll MD, Johnson CL (2002) Prevalence and trends in obesity among us adults. *JAMA* 14: 1723-1727.
29. Donato KA, Rohde FC, Garrison R, Obarzanek E, Ernst ND, et al. (2000) Body mass index and the prevalence of hypertension and dyslipidemia. *Obesity Research*.
30. Faeh D, Paccaud F, Chiolero A, Jacot-Sadowski I, Cornuz J (2007) Association of cigarettes smoked daily with obesity in a general adult population. *Obesity*.
31. Black JB, Saelens BE, Sallis JF, Chen D (2003) Neighborhood-based differences in physical activity: An environment scale evaluation. *American Journal of Public Health*, 9:1552-1558.
32. Clment K (1998) A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*.
33. Stunkard AJ, Foch TT, Hrubec Z (1986) A twin study of human obesity. *JAMA* 1:51-54.