EXPLORING ASSOCIATIONS BETWEEN LIFESTYLES AND METABOLIC SYNDROME IN MIDDLE-AGE CHINESE POPULATION

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Xiaoyi Zhou

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

May 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Exploring Associations Between Lifestyles and Metabolic Syndrome in Middle-age Chinese Population

By

Xiaoyi Zhou

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Gang Shen
Chair

Mark A. Strand

Rhonda Magel

Approved:

4/4/2018
Date

Rhonda Magel
Department Chair

ABSTRACT

Nowadays the prevalence of Metabolic Syndrome (MetS) affects many middle-age people in China. MetS is associated with the risk of type 2 diabetes and cardiovascular disease. Identifying the potential risk factors contribute to MetS is very important for preventing cardiovascular disease. The associations between lifestyles and prevalence of MetS are extensively studied by researchers.

A cross-sectional study, which was conducted by Strand, MA. surveyed 659 subjects in Yuci, China in 2012. The proportional odds model was applied to determine the associations between lifestyles and MetS in three Chinese middle-age groups.

The results demonstrated that doing daily exercise was one of the best method to treat MetS. Moderate alcohol consumption could prevent MetS in age group born in 1956. Occasionally milk consumption could prevent MetS in age group born in 1964, while it did not help age groups born in 1960-1961 and in 1956.

ACKNOWLEDGEMENTS

I would like to thank Dr. Gang Shen for advising me throughout the process of writing and reviewing this paper. And I want to thank Dr. Mark A, Strand. for offering data to me and giving me opportunities to work on this project. Finally, I want to thank all my committee members, Dr. Gang Shen, Dr. Mark A, Strand, Dr. Rhonda Magel for the time they took on my paper.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Nowadays the prevalence of metabolic syndrome (MetS) attracts extensive attention in medical field. According to the study *Menopause and the risk of metabolic syndrome among middle-aged Chinese women*, "MetS refers to a group of inter-related risk factors, which include abdominal obesity, dyslipidemia, insulin resistance, impaired glucose regulation, and hypertension (HTN)." From a previous study, MetS affects 20% to 30% population in developed countries, especially in the middle-aged group (men over age 45 and women over age 55) [1]. The risks of some diseases are associated with MetS, such as type 2 diabetes and cardiovascular disease [2,3]. Moreover, MetS is regarded as an important predictor of morbidity and mortality in cardiovascular disease [4-6]. Thus, exploring and identifying the potential risk factors that contribute to MetS is very important for preventing cardiovascular disease [7].

With the rapid development of industrialization and urbanization, MetS starts to be prevalent in middle-aged Chinese group. In recent decades, the lifestyles, including diet habits and physical activities, are changing dramatically in China [8]. The changes in the lifestyles significantly transformed metabolic profile of middle-aged Chinese people [9]. Alcohol consumption is a very popular habit in Chinese middle-aged group [7]. Alcohol consumption is reported to be associated with type 2 diabetes, hypertension, and dyslipidemia [7]. These are all the components of MetS [7]. The HERITAGE family study demonstrated that aerobic exercise training is a useful treatment strategy for patients with MetS [10]. The association between MetS and the diet habits, including consumption of milk, meat, night snack, fried food, egg, fruits, grain and vegetables, were examined by statistical methods， the result showed eating fruit and drinking milk are associated with reduce risk of Mets [2]. Therefore, the changes in lifestyles appear to be the potential factors either increase or reduce the risk of MetS. Furthermore, Some statistical methods, such as student T-test, Chi-square test and multinomial logistic regression. have been applied to analyze the relationship between lifestyles and MetS in previous studies [7, 10].

The objective of this paper was to determine associated risk factors for prevalence of MetS in middle-aged people in China. The specific objectives were as follows:

1. To determine how age, gender, alcohol consumption, exercise and diet habits are associated with MetS by proportional odds model.

2. To determine the most crucial factors that can contribute to prevalence of MetS or reduce the risk of MetS by proportional odds model.

In this paper, proportional odds models were constructed to analyze the factors that are likely to be associated with MetS. Based on previous studies, age, gender, alcohol consumption, amount of exercise and diet habits (including intake of egg, milk, meat, vegetable, fruit, fried food and night snack) were considered as covariates in initial proportional odds models. In addition, Chi-square test based backward stepwise selection was applied to ignore the covariates that were not significantly associated with MetS. The final models were reached by stepwise selection when all covariates in proportional odds model are significantly associated with MetS. Further, bootstrapping method was used to check the goodness of fit of final models.

2. LITERATURE REVIEW

Recent years, much of the literatures on preventing prevalence of MetS focus on seeking the factors that could have associations with MetS. At the same time, these literatures focus on determining the relationships between the factors and MetS.

Alcohol consumption is a factor that associates with MetS. Whereas, the relationship between alcohol consumption and MetS was inconsistent from previous studies. Koppes LL et al[11] and Baliunas DO et al[12] have claimed that moderate alcohol consumption has a beneficial effect on the risk of type 2 diabetes; while Xin X et al [13] has demonstrated that more alcohol consumption can increase risk of hypertension.

Jin et al [7] assessed the association of alcohol consumption with MetS in Chinese population. The researchers conducted a community-based cross-sectional study on 19,215 subjects' alcohol consumption situation, as well as other variables used in their research. Based on the amount of alcohol consumption, men were divided into five groups, and women were divided into two groups. Chi-square test was applied to compare the percentage difference of MetS rates between groups. The researchers also employed linear regression to examine associations between quality of alcohol consumption and metabolic traits. The alcohol consumption was proved to be associated with each risk determinates of MetS. In addition, multivariate logistic regression was used to assess the adjusted odds ratio of MetS and its individual components. Odds ratios were adjusted for age, gender, BMI, education levels, and exercise, due to age, gender, and exercise can be associated with MetS as well. The results of study showed that in male consumers, excessive wine consumption (alcohol > 50.0g/day) is associated with higher prevalence of MetS, however in women alcohol consumption did not have relationship with prevalence of MetS.

Katzmarzyk et al [10] conducted a study to explore the efficacy of aerobic exercise training in treating MetS. The researchers designed an aerobic exercising program for subjects. A total 126 subjects were identified as having MetS in researchers' sample. 105 out of 126 subjects who have MetS completed aerobic exercise training program. The effects of the aerobic exercise training program were then tested in the group of 105 subjects with MetS. Chi-squared test was applied to test the significant differences in the prevalence of the MetS and individual risk factors before and after training. From the test results, the researchers concluded that Aerobic exercise training is a useful treatment strategy for patients with MetS. In researchers' study, the aerobic exercise training program is a useful way to treat the MetS, it is interesting to know if a person's daily exercise could reduce the risk of MetS or not. In addition, the subjects in researchers' study were form African America and Caucasus group, it will also interesting to know how exercise is associated with MetS in Asian middle age population.

Strand, MA. et al [1] studied the relationships between menopause and MetS in Chinese middle age women. Age, gender, alcohol consumption and exercise were considered as covariates in the study. Multinomial logistic regression was applied to determine the odds ratio between the menopausal status and covariates. ANOVA and Chi-square tests were used to determine correlations between menopausal status and other diseases, such as diabetes and hypertension. The result of this paper showed post-menopausal women in China have higher risk of MetS. In Strand, MA. et al [2] research, the same statistical analysis methods were used as in 2015 study, the relationship between MetS and diet habits (intake of egg, milk, meat, vegetable, fruit, fried food and night snack) was analyzed. The researchers concluded rarely exercising (less than 60 minutes per week) was a significant risk factor of MetS, and abstaining from or occasionally consuming alcohol beverage could prevent MetS.

From above studies, age, gender, alcohol consumption exercise and diet habitats were known to be correlated with MetS, proportional odds model was applied to find more specific relationships between MetS and lifestyles (alcohol consumption, diet habits and exercise), risk factors contribute to MetS, and method to abate prevalence of MetS were determined.

3. METHODS

3.1. Data Collection

Strand, MA. conducted a cross-sectional study to compare the rates of MetS in 3 age groups (born in 1956, 1960-1961, and 1964 in the Yuci District, Jinzhong City, Shanxi Province, China). A questionnaire with 36 questions was administered by research staff to all subjects [2]. The questionnaire assessed demographic data, exercise, alcohol consumption, and diet habits. The inclusion criteria for subjects included being born in Jinzhong City and not currently being under treatment for tuberculosis or cancer, taking corticosteroids, or being pregnant [2]. A total of 806 subjects completed the cross-sectional study in 2008, and 659 subjects completed the follow-up study in 2012 [1]. Two thirds of the subjects were recruited through 16 of 19 community health centers (CHCs) in Yuci District [2]. The Health Record Database of each CHC contains the names of all enrolled individuals, and each CHC has approximately 16 000 persons. The subjects in CHC database were recruited by phone invitation, by posters in community, by word of mouth. At the same time, one third of the subjects were recruited through the Jinzhong Hospital Health Examination Center (JHHEC) [2]. Individuals were examined at JHHEC were primarily healthy individuals whose employer arranged an annual physical exam at this site. Individuals with disease were examined at outpatient department of hospital. The data collection process was done under the authority of the Shanxi Public Health Bureau and Yuci Public Health Bureau. The author has the completely data set collected in 2008 and 2012 from Strand, MA. The data were authorized to use in author’s master paper by Strand, MA. In this paper, only the data collected in 2012 were analyzed. 659 subjects completed the cross-sectional study in 2012, 16 of whom did not response to the questions which were relevant to the attributes used in this research. sample size was 643 after deleting invalid data.

3.2. MetS Definition

According to the U.S. National Cholesterol Education Program (NCEP) Adult Treatment Panel Ⅲ guidelines, MetS is defined using NCEP ATP Ⅲ [14] with the criteria for waist circumference adjusted for Asian people [15]. The NCEP ATP Ⅲ rule recommends that MetS is the presence of three or more of the following risk determinants:

1) Increased waist circumference (>90cm for men; >80cm for women),

2) Elevated triglycerides (>=150 mg/dl) or treatment for this abnormality,

3) Low HDL cholesterol (<40 mg/dl in men; <50mg/dl in women) or treatment for this abnormality,

4) Hypertension (>=130/>=80 mmHg) or treatment for hypertension, and

5) Impaired fasting glucose (>=100 or 110 mg/dl) or treatment for elevated blood glucose.

3.3. Alcohol Consumption

Alcohol consumption was assessed by inquiring each subject about their alcoholic beverage consumption situation in daily life (Strand, MA. designed the study survey 2012). The specific questions which were answered by subjects were: (1) What kind of alcoholic beverage you drink in daily life? (2) How many times you drink per week for each type of alcoholic beverage? (3) How much (the unit used was Liang, 1 Liang = 50 grams) you drink per time for each type of alcoholic beverage? The three alcoholic beverage types included in survey were: (1) beer, (2) wine, and (3) hard liquor (for example Chinese liquor). The alcohol by weight (ABW) in hard liquor was collected by questionnaire, however, ABW in wine and beer were not. According to a previous alcohol consumption study in China [16], the ABWs in beer and wine were estimated as 4% and 14%, respectively. Based on the ABW in each type of alcoholic beverage and quality of alcoholic beverage consumption, alcohol consumption in Liang per month for each subject was calculated. The specific formulas for calculations will be shown in the data transformation section later.

3.4 . Data Description

3.4.1. Response variable -- MetS risk of determinates

The data source for MetS risk of determinates in the original data set is named "x12atpMetS", the values of the variable are integers 0, 1, … ,5, which represent the number of the risk determinants.

3.4.2. Covariates

a) Gender

The data source for covariate gender in the original data set was named "Gender", the value of the covariate were integers 0 and 1 for male and female, respectively.

b) Age

The data source for covariate age in original data set is named “Age”, the value of the covariate were integers 1, 2, and 3 where each number represented a specific age group. Therefore, age was a categorical covariate and it was classified into three categories based on the definition of age groups:

1) Category 1—subjects who born in 1964,
2) Category 2 -- subjects who born in 1960-1961, and
3) Category 3 -- subjects who born in 1956.

c) Alcohol consumption

The data source for alcohol consumption including ABW in hard liquor, alcohol consumed in Liang per time for the three types of alcoholic beverage, and alcohol consumed times per month for the three types of alcoholic beverage. Source data about alcohol consumption was converted into a standard unit in Liang per month for each subject.

d) Exercise

The data source for exercise time in the original data set was named “ExerciseTime”, which recorded exercise time (minutes) for each subject per week.

e) Diet habits

The data source for diet habits including intake of egg, milk, meat, vegetable, fruit, fried food, and night snack. The intakes were represented by integers 1, 2, and 3 where 1 represented consumed rarely, 2 represented consumed occasionally, and 3 represented consumed four or more times a week.

3.5. Data Transformation

3.5.1. Response variable – Class of MetS

Based on the MetS definition, the response variable is ordinal which is in six categories. The response variable was re-classified into four ordinal categories according to the definition of MetS. The new ordinal categories were as follows:

Class 1—*"nomet"* (subjects who showed two or less MetS risk determinants),

Class 2—*"met1"* (subjects who showed three MetS risk determinants),

Class 3—*"met2"* (subjects who showed four MetS risk determinants), and

Class 4—*"met3"* (subjects who showed five MetS risk determinants).

The new re-classified response variable was in four ordinal categories, each class named by *"nomet", "met1", "met2", and "met3"*. The order of the four categories was *"nomet"<"met1"< "met2" < "met3"*.

3.5.2. Covariates -- Alcohol consumption

Because the ABW in the three types of alcoholic beverages were different, formulas were applied to convert measurement unit of alcohol consumption into Liang per month. Based on the analysis in the Alcohol Consumption section, a new numerical variable, alcohol consumption amount per month, was obtained according to the ABW in beer, wine, and hard liquor (ABW in beer and wine were 0.04 and 0.14, respectively, and ABW in hard liquor was collected in the survey). Therefore, the following formulas were used to calculate alcohol consumption amount per month:

$$\begin{cases} beer\ amount = (ABW\ in\ beer) \times (beer\ times) \times (beer\ amount/time) \\ wine\ amount = (ABW\ in\ wine) \times (wine\ times) \times (wine\ amount/time) \\ liquor\ amount = (ABW\ in\ liquor) \times (liqour\ times) \times (liquor\ amount/time) \end{cases}$$

$$alcohol\ consumption = beer\ amount + wine\ amount + liquor\ amount$$

(Eq.1)

The unit of alcohol consumption per month was *Liang*, named by *"MonthlyDrinkAmount"*. According to the frequency table of alcohol consumption per month (Table 1), 203 different values for alcohol consumption per month were recorded, 382 out of 643 subjects did not drink alcohol beverage. The range of alcohol consumption per month was from 0 to 14400, which was a very wide range. The frequency table and histogram of alcohol consumption per month, which were generated by R, are in Table 3.1. and Figure 3.1., respectively:

Table 3.1. Frequency Table of Alcohol Consumption Amount Per Month

NUM	VARIABLE VALUE	FREQUENCY
1	0	382
2	0.1800	1
3	0.7140	1
4	0.9792	1
5	1.1600	1
6	1.1590	1
7	1.6600	1
	....	
201	7500	1
202	14175	1
203	14400	1

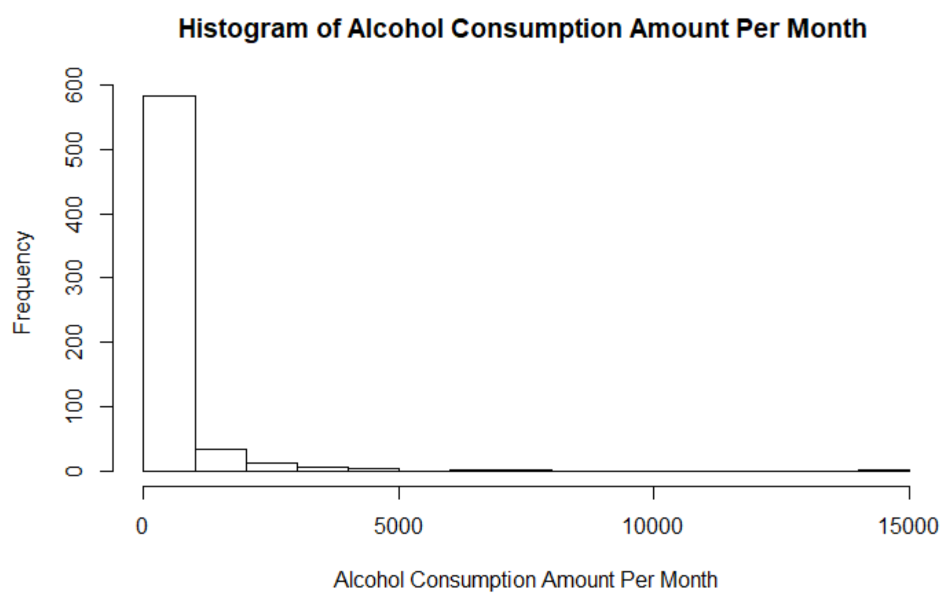

Figure 3.1. Histogram of Alcohol Consumption Amount Per Month

The histogram of alcohol consumption per month showed the distribution was right skewed. The scale of alcohol consumption was very large. Logarithm transformation method was applied to make it easier to visualize. The transformed covariate was called *Drink* and the formula used to transform was

$$Drink = log(MonthlyDrinkAmount + 1) \qquad \text{(Eq.2)}$$

Alcohol consumption per month, "MonthlyDrinkAmount", was added by 1, because the value of $\log(x)$ is a negative infinite at X=0. Since $\log(x+1) = 0$ at x = 0, the equation keeps the value of the logarithm transformed covariate, *Drink,* zero in the group of subjects who did not consume alcohol. The range of the logarithm transformed covariate, *Drink,* was much smaller than it was in the original form, and it was easier to visualize. The frequency table and the histogram of logarithm transformed covariate *Drink*, which were generated by R, are in Table 3.2. and Figure 3.2., respectively:

Table 3.2. Frequency Table of Transformed Covariate Drink

NUM	VARIABLE VALUE	FREQUENCY
1	0	382
2	0.1655	1
3	0.5388	1
4	0.6827	1
5	0.7701	1
6	0.9535	1
7	0.9783	1
	....	
201	8.9228	1
202	9.5593	1
203	9.5705	1

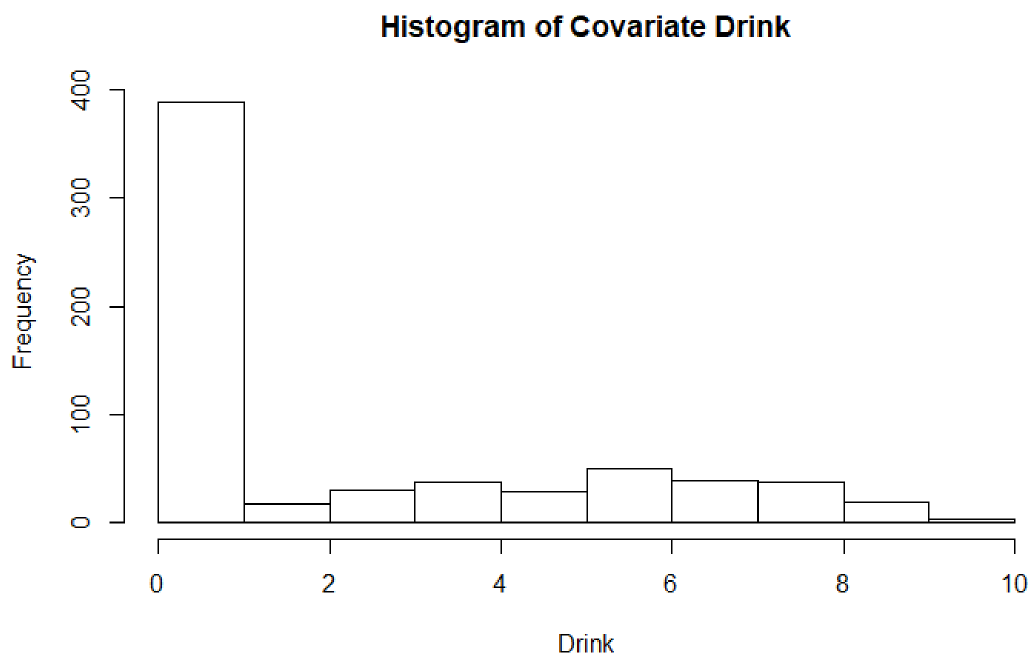

Figure 3.2. Histogram of Transformed Covariate Drink

3.5.3. Covariate -- Exercise

The original data record the exercise time (minutes) per week for subjects, named by "*ExerciseTime".* The frequency table of exercise showed 66 different values, and 284 of subjects did not have daily exercising. According to the histogram of original exercise time, the distribution of exercise time was also right skewed as distribution of original alcohol consumption. The frequency table and the histogram of exercise time per week, which were generated by R, are in Table 3.3 and Figure 3.3, respectively:

Table 3.3. Frequency Table of Exercise Time Per Week

NUM	VARIABLE VALUE	FREQUENCY
1	0	284
2	10	1
3	20	4
4	30	2
5	40	4
6	60	10
7	70	6
	....	
64	2520	1
65	3360	1
66	4200	1

Figure 3.3. Histogram of Exercise Time Per Week

Based on frequency table and histogram of exercise time per week, distribution of exercise time had the similar phenomenon as distribution of original alcohol consumption amount. Therefore, the same logarithm transform method was applied to original exercise time. A new covariate called *Exercise* was generated according to equation:

$$Exercise = log(1 + ExerciseTime) \quad \text{(Eq.3)}$$

Since log(x+1) = 0 at x = 0, the equation keeps value of logarithm transformed covariate, *Exercise,* zero in the group of subjects who did not do exercising. As before, range of logarithm transformed covariate, *Exercise*, was much smaller than it was in original form, it was easier to visualize. The frequency table and the histogram of logarithm transformed covariate *Exercise"* are in Table 3.4 and Figure 3.4 respectively:

Table 3.4. Frequency Table of Transformed Covariate Exercise

NUM	VARIABLE VALUE	FREQUENCY
1	0	284
2	2.3978	1
3	3.0445	4
4	3.4339	2
5	3.7135	4
6	4.1109	10
7	4.2626	6
	…	
64	7.8324	1
65	8.1199	1
66	8.3430	1

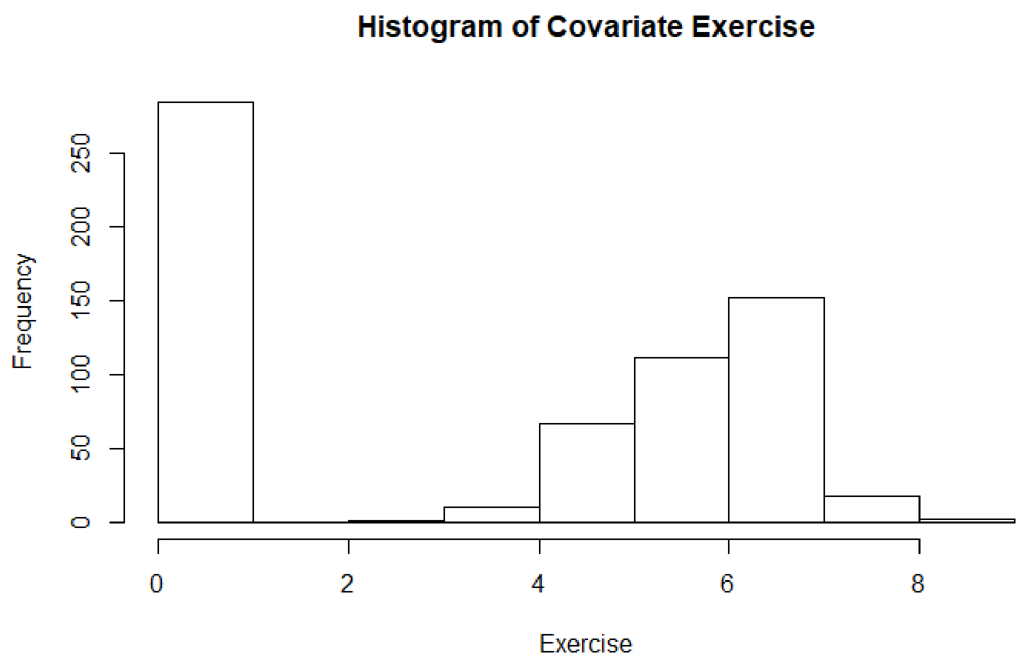

Figure 3.4. Histogram of Transformed Covariate Exercise

3.6. Proportional Odds Model

The method was applied to the proportional odds model [17]. The response variable of the proportional odds model is qualitative, which is similar to the multivariate logistic regression model. The characteristic of the proportional odds model is that the response variable is ordinal. The general form of the proportional odds model with *J* ordinal categories is:

$$\log \frac{P_j}{1-P_j} = \mu_j + \boldsymbol{x}'\beta \qquad \text{(Eq.4)}$$

Where *j = 1, 2, … (J-1).* Therefore, there are *J-1* log odds ratios to be estimated.

The most important reasons for choosing the proportional odds model are:

1. The response variable is qualitative and ordinal. The four categories of

response -- *nomet, met1, met2,* and *met3,* which indicated the four classes of subjects' health condition. *nomet* indicated the class of subjects who were identified as subjects not having MetS. *met1, met2,* and *met3* indicated three ordinal classes of subjects who were identified as patients having MetS. According to the number of MetS risk determinants showed in each class, the order was *nomet* < *met1* < *met2* < *met3*.

2. The effects of all covariates are the same to different response classes. There was not sufficient evidence to demonstrate that age, gender, alcohol consumption, or exercise time had different effects on subjects who had or did not have MetS, so all covariates were assumed to have the same effects on all response classes.

An example of the proportional odds model used for this paper can be written as:

$$\log\frac{P(y\le Y_j)}{1-P(y\le Y_j)} = \mu_j + a_i + g_k + \beta_1 d + \beta_2 e \qquad \text{(Eq.5)}$$

Where *j = 1,2,3,4; i=1,2,3; k=1,2.*

The response variable was in four ordinal classes, so three log odds ratios were calculated. The three log odds ratios can be rewritten as:

$$\log\frac{P_1}{1-P_1} = \mu_1 + a_i + g_k + \beta_1 d + \beta_2 e \qquad \text{(Eq.6)}$$

$$\log\frac{P_2}{1-P_2} = \mu_2 + a_i + g_k + \beta_1 d + \beta_2 e \qquad \text{(Eq.7)}$$

$$\log\frac{P_3}{1-P_3} = \mu_3 + a_i + g_k + \beta_1 d + \beta_2 e \qquad \text{(Eq.8)}$$

Where μ_i was intercept; a_i was main effect of age with age group born in 1964 as baseline; g_k was main effect of gender with male as baseline; d represented alcohol consumption; e represented exercise time; Each P_j represented the cumulative probability of individuals having the met status level no more than j. Moreover, assuming p_j represented the probability of an observation, which only belongs to *j-th* class. The following properties can be conducted: $P_1 = p_1$; $P_2 = p_1 + p_2$; $P_3 = p_1 + p_2 + p_3$; $P_4 = p_1 + p_2 + p_3 + p_4 = 1$.

According to the proportional odds model assumption that covariates have the same effects on all response classes, the coefficients of covariates were the same in equations 6, 7, and 8. The only difference was the intercepts for equations 6, 7, and 8. The coefficients estimated in the proportional odds model were interpretable. For example, one unit increases in alcohol consumption, with other covariates staying the same, corresponded to a change in the log odds ratio where a subject could move from the no MetS class to a MetS class or move from a lower MetS class to a higher MetS class. The intercepts of each log odds equation could be interpreted as a threshold moving to the next class.

The proportional odds models were fitted into two different cases. The first initial proportional odds model included three covariates: age, gender, and alcohol consumption and all of the interaction terms between them were included. The second initial proportional odds model included covariates for: age, gender, alcohol consumption, exercise, and diet habits (consumption of egg, milk, meat, vegetable, fruit, fried food, night snack) and the interaction terms between age and diet habits.

3.7. Model Validation

After creating a full initial proportional odds model, a manual Chi-square test based backward stepwise selection was used to select the covariates significantly associated with MetS ($\alpha = 0.05$). Each time the least significant covariate, which had the largest P-value, would be abandoned. Stopped backward stepwise selection was used until the all remaining covariates were significant. Finally, the final fitted model was reached.

The bootstrapping method was applied to check the goodness of fit in the final fitted model. Bootstrapping is a test method that relies on random sampling with replacement. To check the goodness of fit in the proportional odds model, the basic idea was to check the distribution of deviance of refitted models. According to the final fitted model, each of the 643 observations was randomly classified into one of four ordinal classes, then a new allocation of 643 subjects was obtained. The proportional odds model, using the new allocation as observations, was refitted. The covariates used in the refitted model were the same as the final fitted model. Next a deviance of the refitted model was obtained. The procedure was repeated one thousand times, then the distribution of one thousand deviances was checked, the 95% confidence interval was found. If the deviance of final fitted fell into the 95% confidence interval, then the final model was a good fit.

4. RESULTS

The first initial proportional odds model included three covariates: age, gender, alcohol consumption, and all interaction terms between them, with age born in 1964 and . After Chi-square test based stepwise selection (P-value was 0.0483 for interaction term of alcohol consumption and age), age (with age group born in 1964 as baseline), alcohol consumption and interaction term remained in the final fitted (Model I). R output and histogram of bootstrapping deviances are in Table 4.1 and Figure 4.1 respectively:

Table 4.1. R Output of Model I

polr(formula = y ~ age * drink)			
Coefficients	Value	Std.Error	t-value
age2	0.5021	0.25065	2.003
age3	0.8117	0.24342	3.334
drink	0.1519	0.05333	2.848
age2:drink	-0.1149	0.06846	-1.678
age3:drink	-0.1590	0.06788	-2.342
Intercepts	Value	Std.Error	t-value
nomet|met1	0.6887	0.1954	3.5254
met1|met2	1.7426	0.2059	8.4617
met2|met3	2.9768	0.2344	12.6997
Residual Deviance: 1503.617			
AIC: 1519.617			

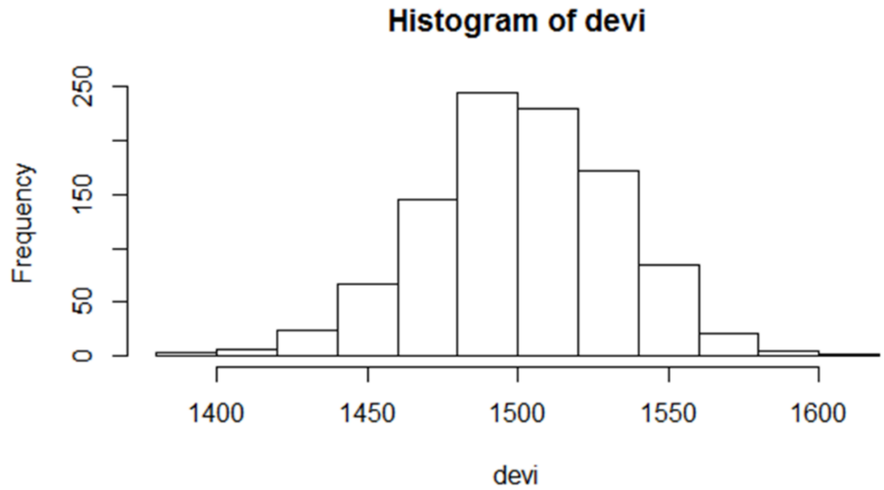

Figure 4.1. Histogram of Bootstrapping Deviances Based on Model I

The 95% quantile value of 1000 bootstrapping deviances was 1561.855. And the deviance of Model I was 1503.617, which was less than 1561.855. The bootstrapping test result showed Model I was good fitted.

The interaction term of age and alcohol consumption was included in the Model I, which indicated alcohol consumption had different effects on three age groups. The Figure 4.2 showed the effect of alcohol consumption on three age groups.

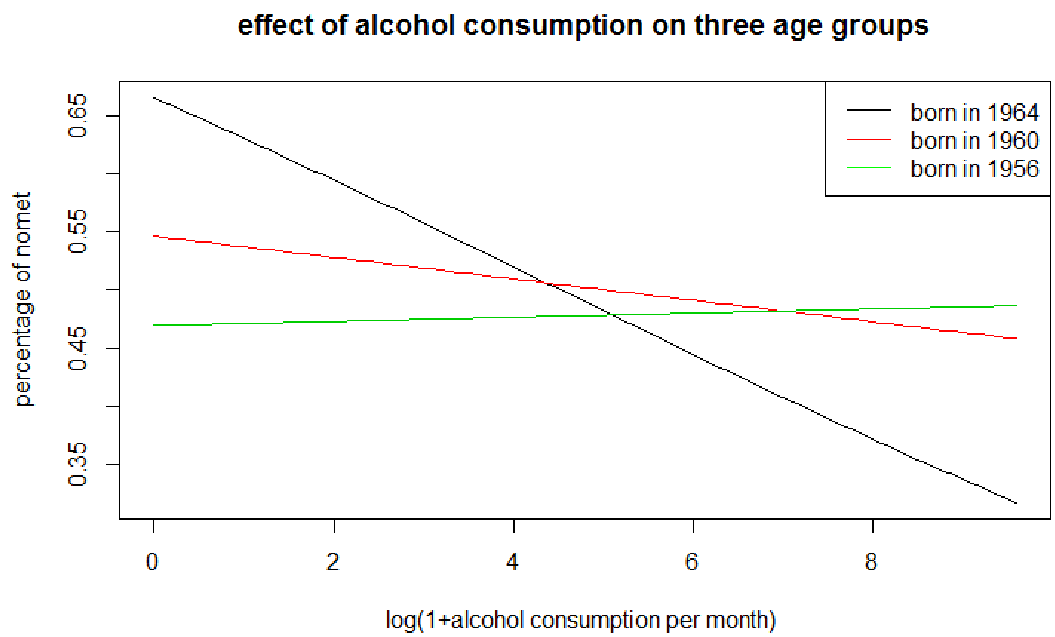

Figure 4.2. Effects of Alcohol Consumption on Three Age Groups

The second initial proportional odds model included covariates: age, gender, alcohol consumption, exercise and diet habitats (consumption of egg, milk, meat, vegetable, fruit, fried food, night snack), and all of interaction terms between age and diets habitats covariates. After Chi-square test based stepwise selection (P-value was 0.019 for interaction term between age and milk consumption, and P-value was less than 0.001 for exercise), age (with age group born in 1964 as baseline), milk consumption (with milk consumption described as rarely as baseline), exercise time and interaction term between age and milk consumption remained in the final fitted model (Model II). R output and histogram of bootstrapping deviances are in Table 4.2 and Figure 4.3 respectively:

Table 4.2. R Output of Model II

polr(formula = y ~ age * milk + exercise)			
Coefficients	Value	Std.Error	tvalue
age2	0.047571	0.26535	0.179277
age3	0.167373	0.26717	0.626461
milk2	-1.166874	0.40923	-2.851397
milk3	0.003173	0.38302	0.008285
exercise	-0.093952	0.02646	-3.550634
age2:milk2	1.216249	0.50981	2.385676
age3:milk2	1.401515	0.51539	2.719326
age2:milk3	-0.476120	0.49375	-0.964289
age3:milk3	-0.074448	0.46999	-0.158402
Intercepts	Value	Std.Error	tvalue
nomet|met1	-0.2170	0.2149	-1.0097
met1|met2	0.8575	0.2180	3.9333
met2|met3	2.0984	0.2417	8.6816
Residual Deviance: 1485.065			
AIC: 1509.065			

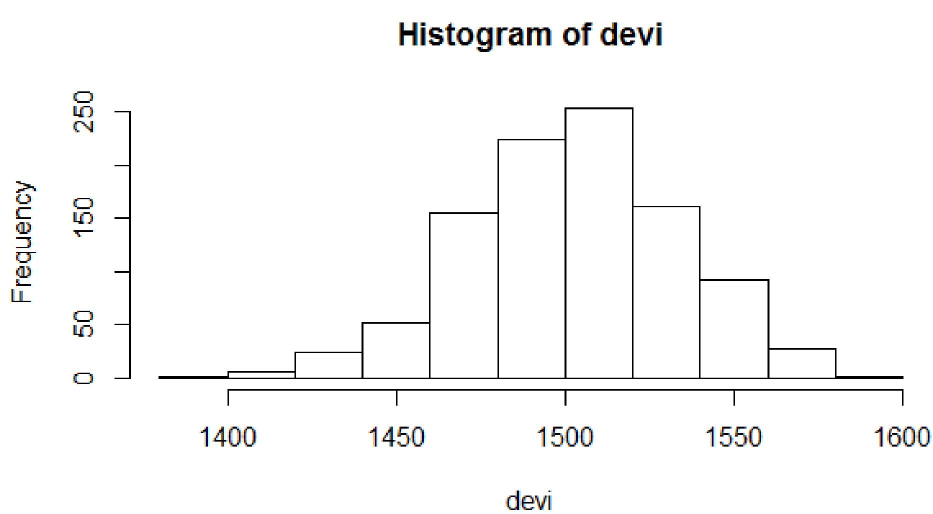

Figure 4.3. Histogram of Bootstrapping Deviances Based on Model II

The 95% quantile value of 1000 bootstrapping deviances was 1548.851. And the deviance of model II was 1485.065, which was less than 1548.851. The bootstrapping test result showed model II was good fitted.

After plugged into exercise and diet habitats covariates, the effect of alcohol consumption was not significant. exercise showed very significantly effect on health. Figure 4.4 showed the effect of exercise on three age groups, when considered milk consumption as a constant at baseline.

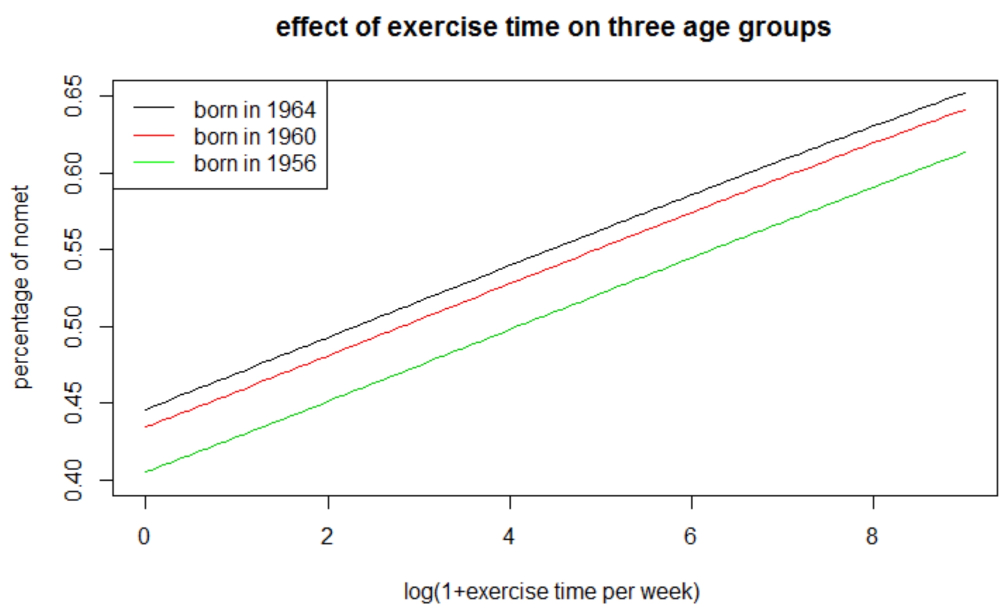

Figure 4.4. Effect of Exercise Time on Three Age Groups

Furthermore, as described in methods section, three log odds models were created. In three log odds models, the coefficients of covariates were the same, only intercept changed. For example, in Model II, if setting covariates: age group was born in 1956, occasionally consumed milk, and log of exercise time was one, then three log odds were calculated by:

$$\log\frac{P_1}{1-P_1} = -0.217 + \boldsymbol{x}'\beta \qquad \text{(Eq.7)}$$

$$\log\frac{P_2}{1-P_2} = -0.8575 + \boldsymbol{x}'\beta \qquad \text{(Eq.8)}$$

$$\log\frac{P_3}{1-P_3} = 2.0984 + \boldsymbol{x}'\beta \qquad \text{(Eq.9)}$$

Wher $\boldsymbol{x}'\beta = -0.1673 \times age3 + 1.1666 \times milk2 - 1.4015 \times age3 \times milk2 + 0.094 \times exercise$. Then from equation 7, 8 and 9, three odds ratios were 0.5915, 1.7323 and 5.9915. The corresponding cumulative probabilities were 0.3717, 0.6340 and 0.8570. based on properties $P_1 = p_1$; $P_2 = p_1 + p_2$; $P_3 = p_1 + p_2 + p_3$; $P_4 = p_1 + p_2 + p_3 + p_4 = 1$. The proportion of *nomet, met1, met2* and *met3* were 0.3717, 0.2623, 0.2230 and 0.1430, respectively.

5. DISCUSSION

From the result of Model I, alcohol consumption had different effects on the three age groups. First, there was a significantly negative effect on the age group born in 1964, the proportion of subjects who were classified as not having MetS decreased apparently with the increase of alcohol consumption. In addition, the effect of alcohol consumption on subjects who were born in 1960-1961 was similar to the effect on subjects born in 1964. However, the effect on subjects who born in 1956 was positive. The average alcohol consumption (in unit Liang per month) in age group born in 1964, 1960-1961 and 1956 were 319.7289, 399.8107 and 263.979 respectively. Subjects born in 1956 had lower relative alcohol consumption compared to the two younger groups. This showed moderate alcohol consumption could reduce the risk of MetS. However, the limitation is that only the association between alcohol consumption amount and MetS was evaluated, and the association between types of alcohol beverage and MetS was not considered. Effects of different types of alcohol beverage may be various, one subject may drink a mix of alcohol beverage types. Therefore, the ratios of the three types of alcohol beverage consumption might be a potential factor, which could increase or reduce the risk of MetS.

After considered exercise and diet habitats as covariates in the model, the alcohol consumption was not significant. Based on the Model II, compared with alcohol consumption, the exercise and milk consumption were more crucial determinative factors, which contribute more to MetS.

Daily exercise is a very significant method to treat MetS according to the Model II. Exercise had the same effect on three age groups. For example, considering subjects in age group born in 1964 and rarely consumed milk, when exercise time changed from 0 to 60 minutes a week, the model predicted the proportion of subjects without MetS increased from 66.57% to 79.29%. The increased proportions of subjects without MetS were the same in three age groups when exercise changed one unit, because exercise had the same effect on three age groups. Further, for age group born in 1956 and rarely consumed milk, when exercise time increased from 0 to 60 minutes a week, the proportions of *nomet, met1, met2* and *met3*, changed from 40.50%, 26.09%, 20.74% and 12.66% to 50.05%, 24.54%, 16.45% and 8.97%. The predicted proportion of subjects without MetS increased when exercise time increased; Meanwhile, the predicted proportions of subjects, who showed three, four and five risk determinants, were all decreased.

Milk consumption had different effects on three age groups, which was similar to the effects of alcohol consumption. The result showed occasionally milk consumption can prevent MetS in age group born in 1964. when assuming conditions: occasionally consuming milk and 60 minutes' exercise a week, the proportion of subjects without MetS in age group born in 1964, 1960-1961 and 1956 were 79.1%, 51.8% and 44.2% respectively. Moreover, in age group born 1964, assuming 60 minutes' exercise a week, the proportions of subjects without MetS for rarely, occasionally, and often consuming milk were 54.2%, 79.1% and 54.1% respectively. Occasionally milk consumption was a good treatment for age group born in 1964. Whereas, in age group born in 1956 and 1960-1961, occasionally milk consumption did not reduce the risk of MetS.

From above results, the Model I and Model II demonstrated the following conclusions:

1. Amount of exercise was the most significant treatment method for MetS in the experiment. Increasing amount of exercise apparently reduce proportion of MetS in population.

2. Alcohol consumption had different effect on three age groups. Moderate alcohol consumption could prevent MetS in age group born in 1956. Two younger Age groups had higher alcohol consumption than 1956 group, the higher alcohol consumption could produce higher risk of MetS in these two groups.

3. Milk consumption also had different effect on three age groups. Occasionally milk consumption could prevent MetS in age group born in 1964, while it did not help age group born in 1960-1961 and 1956.

Finally, there are some limitations in this research. First, the generalizability was limited by the low response rates of each CHC. The selected subjects may could not represent the rest of population in the communities. Whether the general population was healthier than the selected subjects was unknown. Second, the amount of alcohol consumption was recorded based on subjects' subjective evaluation, instead of objective measurement, the bias was unavoidable.

REFERENCES

1. Strand, Mark A.; Huseth-Zosel, Andrea; He, Meizi; Perry, Judith. *Menopause and the risk of metabolic syndrome among middle-aged Chinese women.* Family Medicine and Community Health, Volume 3, Number 1, 1 March 2015, pp. 15-22(8)

2. Strand MA, Perry J, Wang P, Liu S, Lynn H. *Risk factors for metabolic syndrome in a cohort study in a north China urban middle-aged population*. Asia Pac J Public Health 2015;27(2):NP255–NP265.

3. Yang Z-J, Liu J, Ge J-P, Chen L, Zhao Z-G, Yang W-Y. *Prevalence of cardiovascular disease risk factor in the Chinese population: the 2007–2008 China national diabetes and metabolic disorders study.* Eur Heart J 2012;33:213–20.

4. Gami AS, Witt BJ, Howard DE, Erwin PJ, Gami LA, Somers VK, et al. *Metabolic syndrome and risk of incident cardiovascular events and death: a systematic review and meta-analysis of longitudinal studies*. J Am Coll Cardiol 2007;49:403–14.

5. Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, et al. *The metabolic syndrome and total and cardiovascular disease mortality in middleaged men*. JAMA 2002;288:2709–16.

6. Faramawi MF, Sall M, Abdul Kareem MY. *The association of the metabolic syndrome with T-wave axis deviation in NHANES III*. Ann Epidemiol 2008;18:702–7.

7. Lina Jin, Yun Huang, Yufang Bi, Liebin Zhao, Min Xu, Yu Xu, Yuhong Chen, Lina Gu, Meng Dai, Yaohua Wu, Jianing Hou, Xiaoying Li , Guang Ning. *Association between alcohol consumption and metabolic syndrome in 19,215 middle-aged and elderly Chinese*. Diabetes research and clinical practice 92(3):386-92 · March 2011.

8.Visscher TLS. *Public health crisis in China is about to accelerate the public health crisis in our world's population.* Eur Heart J 2012;33:157–9.

9.Muntner P, Gu D, Wildman RP, Chen J. *Prevalence of physical activity among Chinese adults: results from the international collaborative study of cardiovascular disease in Asia.* Am J Public Health 2005;95:1631–6.

10. Peter T. Katzmarzyk, Arthurh S. Leon, Jack H. Wilmore, James S. Skinner, D.C. Rao, Tuomo Rankenie, and Claude Bouchard. *Targeting the Metabolic Syndrome with Exercise: Evidence from the HERITAGE Family Study.* Medicine & Science in Sports & Exercise 35(10):1703-9 · November 2003.

11. Koppes LL, Dekker JM, Hendriks HF, Bouter LM, Heine RJ. *Moderate alcohol consumption lowers the risk of type 2 diabetes: a meta-analysis of prospective observational studies.* Diabetes Care 2005;28:719–25.

12. Baliunas DO, Taylor BJ, Irving H, Roerecke M, Patra J, Mohapatra S, et al. *Alcohol as a risk factor for type 2 diabetes: a systematic review and meta-analysis.* Diabetes Care 2009;32:2123–32.

13. Xin X, He J, Frontini MG, Ogden LG, Motsamai OI, Whelton PK. *Effects of alcohol reduction on blood pressure: a metaanalysis of randomized controlled trials*. Hypertension 2001;38:1112–7.

14. Grundy SM. *Metabolic syndrome scientifi statement by the American Heart Association.* Arterioscler Thromb Vasc Biol 2005;25:2243–4.

15. Heng D, Ma S, Lee J, Tai B, Mak K, Hughes K, et al. *Modifiation of the NCEP ATPIII defiitions of the metabolic syndrome for use in Asians identifis individuals at risk for ischemic heart disease.* Atherosclerosis 2006;186:367–73.

16. Li Y, Wang JG, Gao PJ, Wang GL, Qian YS, Zhu DL, et al. *Interaction between body mass index and alcohol intake in relation to blood pressure in HAN and SHE Chinese.* Am J Hypertens 2006;19(5):448–53.

17. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (1st ed.) Boca Raton, FL: Chapman & Hall/CRC.

APPENDIX. R CODE

```
file=file.choose()
mets=read.csv(file, header=T)
names(mets)
library(faraway)
library(MASS)
attach(mets)
age<-factor(Age)
gender<-factor(Gender,labels=c("M","F"))
drink<-cut(drinkamt_mo, breaks=c(-1,0.1,500,1000,200000),
labels=c("nodrink","ocassionally","often","intemperence"))
data.frame(table(drink))
exer<-log(1+X_12mex_18c)
hist(log(1+exercise))
exercise<-X_12mex_18c
fruit<-factor(X_12fruit_39)
fry<-factor(X_12fry_37)
bedsnack<-factor(X_12bed_38)
veg<-factor(X_12veg_40)
milk<- factor(X_12milk_41)
meat<- factor(X_12meat_42)
egg<- factor(X_12egg_44)

y<-factor(x12atpmets)
levels(y)<-c("nomet","nomet","nomet","met1","met2","met3")
drink<-log(drinkamt_mo+1)
hist(drink)
hist(drinkamt_mo)

intialmodi <- polr(y~ age*(drink+age+gender+exer+meat+veg+fruit+fry+bedsnack+milk)
initialmodc<-polr(y~ age*drink*gender)

drop1(initialmodc,test="Chisq")
modc<- polr(y~ drink*age)
summary(modc)
modc

drop1(initialmodi,test="Chisq")
modi<-polr(y~ age*milk+exer)
summary(modi)
```

```
modi

table(exercise)
data.frame(table(exercise))


devi<-newy<-NULL
an<-c("nomet","met1","met2","met3")
for(i in 1:1000){
   for(j in 1: 643){
      nid<-rmultinom(1,1,modc$fit[j,])
      pid<-which(nid==1)
      newy[j]<-an[pid]
   }
   newmod<-polr(factor(newy)~age+exer)
   devi[i]<-newmod$deviance
}
newy
length(z)
length(newy)
modc$fit
hist(devi)
quantile(devi,0.95)
quantile(devi,0.05)
summary(modc)

p11<-function(x) ilogit(modc$zeta[1]-modc$coef[1]*x)
p12<-function(x) ilogit(modc$zeta[1]-modc$coef[1]*x-modc$coef[2]-modc$coef[4]*x)
p13<-function(x) ilogit(modc$zeta[1]-modc$coef[1]*x-modc$coef[3]-modc$coef[5]*x)

pp11<-function(x) ilogit(modi$zeta[1]-modi$coef[5]*x)
pp12<-function(x) ilogit(modi$zeta[1]-modi$coef[5]*x-modi$coef[1])
pp13<-function(x) ilogit(modi$zeta[1]-modi$coef[5]*x-modi$coef[2])

z11<- ilogit(modi$zeta[1])
z12<- ilogit(modi$zeta[1]-modi$coef[3])
z13<- ilogit(modi$zeta[1]-modi$coef[4])

z21<- ilogit(modi$zeta[1]-modi$coef[1])
z22<- ilogit(modi$zeta[1]-modi$coef[1]-modi$coef[3]-modi$coef[6])
z23<- ilogit(modi$zeta[1]-modi$coef[1]-modi$coef[4]-modi$coef[8])
```

```
z31<- ilogit(modi$zeta[1]-modi$coef[2])
z32<- ilogit(modi$zeta[1]-modi$coef[2]-modi$coef[3]-modi$coef[7])
z33<- ilogit(modi$zeta[1]-modi$coef[2]-modi$coef[4]-modi$coef[9])
as.matrix(c(z11,z12,z13, z21,z22,z23,z31,z32,z33))
as.matrix(c(z11,z12,z13),c(z21,z22,z23))
ilogit(modc$zeta[1])
ilogit(modc$zeta[1]-modc$coef[1]*log(61))
summary(modi)

P1<- ilogit(modi$zeta[1]-modi$coef[2]-modi$coef[3]-modi$coef[7]-modi$coef[5])
P2<- ilogit(modi$zeta[2]-modi$coef[2]-modi$coef[3]-modi$coef[7]-modi$coef[5])
P3<- ilogit(modi$zeta[3]-modi$coef[2]-modi$coef[3]-modi$coef[7]-modi$coef[5])

plot(x,p11(x),type ="l", col=1, xlab="log(1+alcohol consumption per month) ",
ylab="percentage of nomet", main= "effect of alcohol consumption on three age
groups")
lines(x,p12(x),lty=1, col=2)
lines(x,p13(x),lty=1, col=3)

plot(x,pp11(x),type ="l", col=1, ylim= c(0.4, 0.65),xlab="log(1+exercise time per week) ",
ylab="percentage of nomet", main= "effect of exercise time on three age groups")
lines(x,pp12(x),lty=1, col=2)
lines(x,pp13(x),lty=1, col=3)

range(exer)
x <- seq(0,9,0.1)

e11<- ilogit(modi$zeta[1]-modi$coef[2])
e12<- ilogit(modi$zeta[2]-modi$coef[2])
e13<- ilogit(modi$zeta[3]-modi$coef[2])
e21<- ilogit(modi$zeta[1]-modi$coef[2]-modi$coef[5]*log(61))
e22<- ilogit(modi$zeta[2]-modi$coef[2]-modi$coef[5]*log(61) )
e23<- ilogit(modi$zeta[3]-modi$coef[2]-modi$coef[5]*log(61))
c(e11, e12-e11, e13-e12, 1-e13)
c(e21, e22-e21, e23-e22, 1-e23)
```