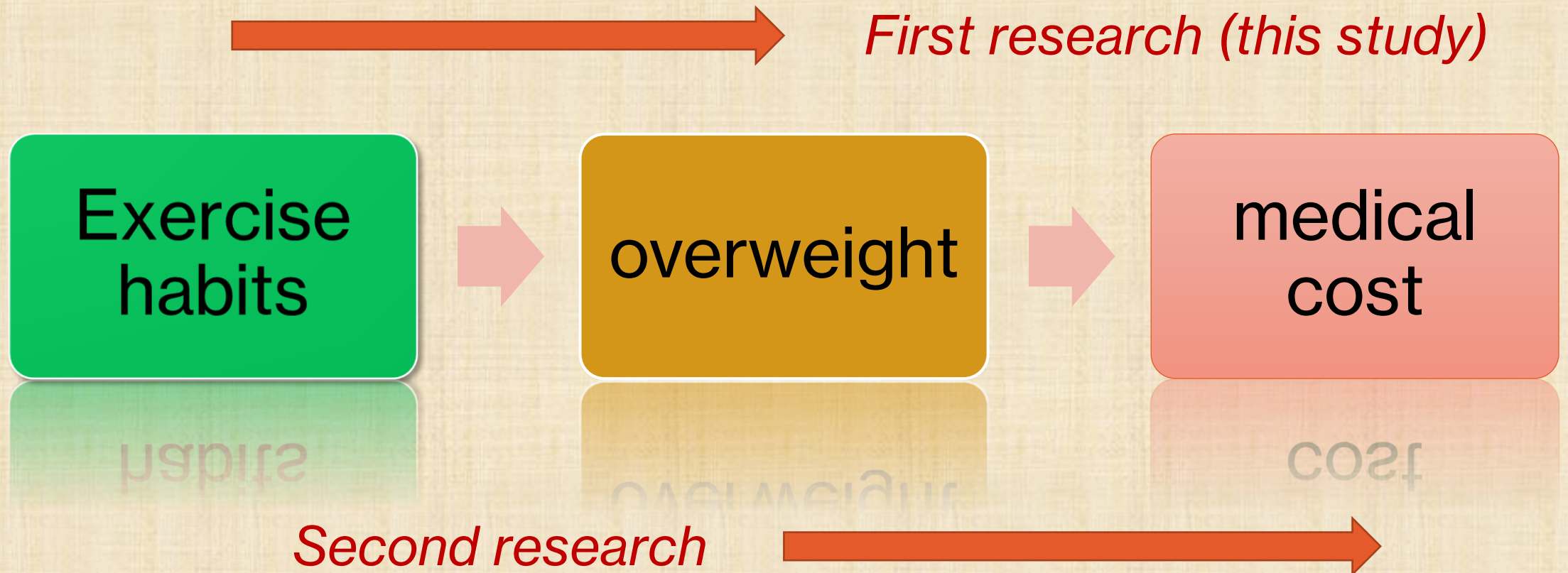


*A Study on the Correlation between Exercise and
Overweight in the 19 to 64 Age Group in Taiwan
- Application of the Bivariate Probit Model*

*Chieh-Hsien Lee
Fo Guang University*



Research motivation



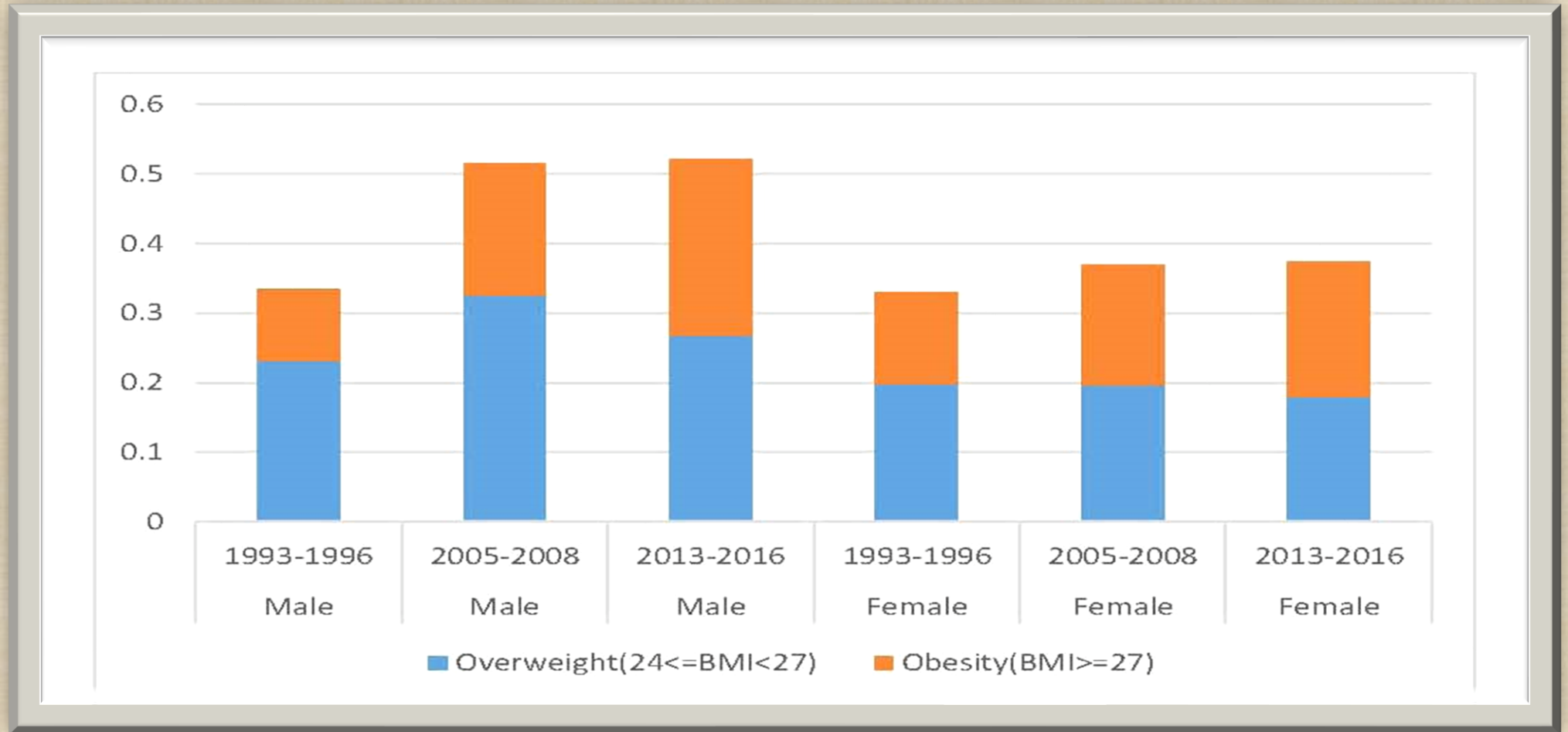
Introduction

- * " Insufficient Exercise " has become a topic of public health and welfare.
- * Overweight or Obesity has become one of the important issues in global health and medical treatment.
- * Many studies have found that exercise is not only beneficial for the development of physical functions, but also helps prevent overweight and chronic diseases.
- * Is exercise habit an exogenous or endogenous impact factor of overweight?

Nutrition and Health Survey in Taiwan

- * The Ministry of Health of the Republic of China has long commissioned the Academia Sinica to conduct a "Nutrition and Health Survey in Taiwan"(NAHSIT).
- * The purpose is to establish a national nutrition survey mechanism to monitor the trend of people's nutrition and health as an important reference indicator for the government to formulate nutrition and health policies.
- * Next figure compared with the prevalence of obesity surveyed in 1993-1996, 2005-2008 and 2013-2016, the total proportion of male and female who were overweight and obesity, from 33.4% and 33.0% of 1993-1996 increased to 51.5% and 36.9% of 2005-2008, and increased to 52.1% and 37.4% of 2013-2016.

Compare datas



- * The prevalence of overweight and obesity among adults aged 19 and over in Taiwan during 1993-1996, 2005-2008 and 2013-2016 (%)

Other surveys of advanced countries

- * Advanced countries such as the NHANES (National Health and Nutrition Examination Survey) survey in the United States and the National Maintenance Survey in Japan are also large-scale nutrition surveys organized by the government to monitor the dietary nutritional health trends of citizens.
- * Pan (2019) analyzes three data from the Taiwan Nutritional Change Survey (NAHSIT) of "1993-1996", "2005-2008" and "2013-2016" to track the situation of overweight or obese adult people in Taiwan.
- * From these studies, we can know that the population of many countries is currently overweight, and it is speculated that it may be highly related to exercise, and lifestyle. Among these issues, exercise and overweight are still the most popular research topics.

Research purpose

- * Using the data of NAHSIT, develop an evaluation model suitable for evaluating the prevalence of overweight with endogenous exercise variable.
- * Research exercise habit is an exogenous or endogenous impact factor of overweight.
- * Estimate that if other factors remain unchanged, the prevalence of overweight among those with exercise habit is lower or higher than that of those without exercise habit.

Theoretical framework

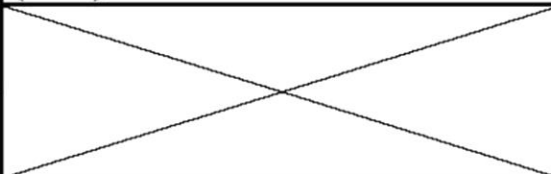
- * Data
- * Definition of "overweight" and "exercise habits"
- * Empirical model
- * Hausman test and J statistic
- * Variables
- * Data analyses

Data

- * In this study, use was made of second wave of Nutrition and Health Survey in Taiwan (NAHSIT)2005-2008. The NAHSIT survey has been described as follow research.
- * Tu SH, Chen C, Hsieh YT, Chang HY, Yeh CJ, Lin YC, et al. Design and sample characteristics of the2005-2008 Nutrition and Health Survey in Taiwan. Asia Pac J Clin Nutr. 2011; 20(2):225±37. Epub2011/06/15. PMID: 21669592
- * There were 1,710 people who had completed valid questionnaires and performed health checks at our research .

Definition of overweight

Adult healthy body position standards in Taiwan

Adult body position definition	definition body mass index — BMI (kg / m^2)	Waist circumference (cm)
Underweight	$BMI < 18.5$	
Healthy position	$18.5 \leq BMI < 24$	
Overweight	$24 \leq BMI < 27$	Male: ≥ 90 cm Female: ≥ 80 cm
Obesity	$27 \leq BMI$	

Note: The definition of overweight in this article is the definition of overweight and obesity, that is, those with a BMI greater than(or equal) 24 are considered overweight

Source: Health and Welfare Department "Adult Healthy Body Standards" (2018.10.24)

Definition of exercise habit

International Physical Activity Scale

-Classification of high, medium and low levels of physical activity

Physical activity status	Weekly indicators
Insufficient physical activity	Those who do not meet the indicators of adequate physical activity and high physical activity.
Adequate physical activity	1. Strenuous activities for 3 days or more and at least 20 minutes a day. 2. For 5 days or more and at least 30 minutes a day, moderate effort or walking. 3. Any activity up to 5 days with a total activity volume of 600 or more.
High physical activity	1. A strenuous activity that lasts for 3 days and has a total activity of more than 1500. 2. Any activity that lasts for 7 days or more and the total activity is more than 3000.

Note1:Resource : Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire(IPAQ)(2005)

Note2: The amount of activity per week (MET-min per week) = MET level × how many minutes are spent doing activities each day × how many days a week. Among them, the MET level is scores for personal exercise methods, such as walking for 3.3 points, moderate effort for 4.0 points, and effortless activities for 8.0 points. If an individual walks for 30 minutes a day and reaches 5 days a week, the weekly activity volume is $3.3 \times 30 \times 5 = 495$ MET-minutes/week.

Note3: The definition of with exercise habit in this article is the definition of adequate physical activity and High physical activity and no exercise habit is the definition of Insufficient physical activity and others.

Empirical model

- * In order to understand the impact of exercise on the prevalence of overweight, this study set "exercise habit" and "overweight" situation as two dependent **binary** variables. Since there may be potential (endogenous) correlation between the two variables, the bivariate probit model is chosen for estimation to avoid wrong estimation.
- * According to Maddala (1983)'s definition of bivariate probit model, assuming that Y_1^* and Y_2^* are two unobservable potential variables, then Y_1^* and Y_2^* have the following regression relationship with the explanatory (independent) variable X :

Bivariate Probit Model

$$Y_{1i}^* = X'_{1i}\beta_1 + \varepsilon_{1i} \quad (1)$$

$$Y_{2i}^* = X'_{2i}\beta_2 + \gamma Y_{1i} + \varepsilon_{2i} \quad (2)$$

$$Y_{1i} = \begin{cases} 1 & \text{if } Y_{1i}^* > 0 \\ 0 & \text{if } Y_{1i}^* \leq 0 \end{cases} \quad (3)$$

$$Y_{2i} = \begin{cases} 1 & \text{if } Y_{2i}^* > 0 \\ 0 & \text{if } Y_{2i}^* \leq 0 \end{cases} \quad (4)$$

Model description

- * β_1 and β_2 are the estimated coefficients corresponding to X . γ is also an estimation coefficient, which measures the potential effect of adults with exercise habits on overweight under other conditions not changed. β_1 and β_2 are error terms and meet the joint standard normal distribution (mean is 0, standard deviation is 1), and the correlation coefficient is represented by ρ . If $\rho \neq 0$ is significant, it means that there is a correlation between exercise habits and the prevalence of overweight, and this correlation is often determined by invisible common factors (for example, people's risk preferences, etc.). In addition, such models usually require that at least one independent variable in X_{1i} does not exist in X_{2i} (such as instrumental variable) during empirical analysis (Maddala, 1983), which is usually called an identification problem in the literature.

Endogeneity and instrumental variable

- * In this study, the selection of instrumental variables is mainly based on Cawley (2004) dealing with endogenous problem.
- * We think that exercise will also be affected by factors such as personal choice, knowledge, and culture, so exercise can be regarded as endogenous variable. In order to deal with the endogenous problem, this article uses "Pulmonary function measurement value" and "Anterior bending measurement value" as instrumental variables to estimate model.
- * The instrumental variable method must meet two conditions. First, the instrumental variable must have no correlation with the residual items in the overweight model. Second, the instrumental variable must have a high correlation with exercise habits.

Hausman test

$$H_0: (\hat{\gamma} - \tilde{\gamma}) = 0$$
$$h = \frac{(\hat{\gamma} - \tilde{\gamma})^2}{Var(\hat{\gamma}) - Var(\tilde{\gamma})} \sim \chi^2(1)$$

- * The null hypothesis and test statistics of Hausman test are as above:
- * In order to ensure the effectiveness of the instrumental variable and meet the first and second condition. First, we perform the overidentified test and set the null hypothesis is that the residual term of the overweight model and instrumental variable are not relevance. Second, we enter the weak IV test and set the null hypothesis is that the instrumental variable has no effect on the exercise habit model.

J statistic

- * we perform the J statistic on the three physical examination variables of the health indicators to test whether there are endogenous problems or not. When the probability value of the tested Difference in J-stats is <0.05 , the null hypothesis is rejected, which means that the physical examination variable has an endogenous problem.
- * After doing an endogenous test for SBP, LDL/HDL and GLUCOSE is found probability values of J-stats are 0.1524, 0.1271 and 0.1166 respectively, which are all greater than 0.05, indicating that SBP, LDL/HDL and GLUCOS of the samples in this study have no endogenous problems.

Variable name	Definition	Average	Expected impact	
			Overweight	Exercise
EH	A. Health Index. B. Recall in the last week (including today), (1) Difficulty in sleeping, difficulty falling asleep, waking up easily or early. (2) Nervous and restless. (3) Feel easily distressed or angry. (4) Feeling melancholy and depressed. (5) Feel inferior to others. C. The degree to which the above problems make you feel distressed (1. Not at all 2. Slight 3. Moderate 4. Severe 5. Very severe) The total score is 25 points. If the total score is less than 5 points, it is normal, the EH value is 1; the total score is between 6-9 points is mild, the EH value is 2; the total score is between 10-14 points is moderate, and the EH value is 3. ; A total score of 15 is considered severe, and the EH value is 4.	1.836	-	?
Physical examination variables:				
GLUCOSE	Fasting blood glucose, unit:(mg/dL) °	107.5	+	?
LDL/HDL	Ratio of low-density cholesterol to high-density cholesterol	2.238	+	?
SBP	Systolic blood pressure, unit : mmHg °	113.9	+	?
Socio-economic variables:				
AGE	Age, unit: years.	43.64	?	?
SEX	Gender, dummy variable, 0 for male and 1 for female.	0.525	+	-
EDU	Education level, unit: year.	11.24	-	-
MARRY	Marriage status, dummy variable, married or cohabitation is 1, and other statuses are 0.	0.761	?	?
JOB	Work status, dummy variable, full-time job or family business worker is 1, and other statuses are 0.	0.561	?	?
HOUSE	Own house, dummy variable, owning is 1, not owning is 0.	0.886	?	?
INCOME	Personal Monthly Income, Unit: Thousand dollar.	22.81	+	?
SMOK	Whether to smoke for a long time, dummy variable, yes is 1, no is 0.	0.237	+	-
ALC	Whether to drink for a long time, dummy variable, yes is 1, no is 0.	0.140	+	-
CHU	Whether to eat betel nut for a long time, dummy variable, yes is 1, no is 0.	0.160	+	-

instrument variable:

PUN	A. Pulmonary function measurement value (L/min). B. The estimated value of normal peak expiratory flow rate is calculated based on individual height, gender, race and age. The formula for calculating the estimated adult value is as follows: (by Department of Thoracic Disease, Taipei Veterans General Hospital) Male: $3.89 \times \text{height (cm)} - 2.95 \times \text{age (full years)} + 43.59$ (L/min) Female: $4.10 \times \text{height (cm)} - 1.61 \times \text{age (full years)} - 173.55$ (L/min) The average measured value of the physical examination is greater than the estimated value is 1, otherwise is 0.	0.146	?	+
SRT	Anterior bending measurement value (cm), as shown on the website of the Ministry of Education. 19-64-year-old Taiwanese seated anterior bending percentile (unit: cm), bad and slightly worse grade is 0, ordinary, fair and very good grade is 1.	0.545	?	+
Dependent variable :				
OVERWT	Overweight or obesity, body mass index (kg/m^2) greater than or equal to 24 is 1, otherwise is 0.	0.511		
EXER	Exercise situation, with exercise habit is 1 and no exercise habit is 0.	0.313		
The number of samples	Population aged were 19 to 64.			1710

Data analyses(1)

Statistical Table of Average Value of Related Variables of Empirical Model (1)

	SAMPLES	OVERWT	EXER	EDU	JOB	HOUSE	INCOME	EH
all	1710	0.511	0.313	11.24	0.561	0.886	22817	1.836
male	811	0.578	0.314	11.93	0.676	0.890	30151	1.741
female	899	0.449	0.311	10.61	0.457	0.882	16201	1.922
[19,30]	358	0.349	0.327	13.95	0.550	0.796	17318	1.913
male	165	0.461	0.376	14.24	0.630	0.764	20379	1.739
female	193	0.254	0.285	13.71	0.482	0.824	14702	2.062
(30,40]	272	0.460	0.316	12.84	0.728	0.831	29320	1.890
male	126	0.579	0.254	13.28	0.881	0.857	38433	1.817
female	146	0.356	0.370	12.46	0.596	0.808	21455	1.952
(40,50]	480	0.529	0.298	11.11	0.635	0.898	27885	1.810
male	226	0.571	0.319	11.76	0.788	0.912	37146	1.796
female	254	0.492	0.280	10.53	0.500	0.886	19646	1.823
(50,60]	477	0.604	0.312	9.39	0.495	0.950	21148	1.788
male	232	0.642	0.306	10.57	0.599	0.953	29731	1.672
female	245	0.567	0.318	8.27	0.396	0.947	13020	1.898
(60,64]	123	0.659	0.325	7.45	0.187	0.976	11138	1.780
male	62	0.677	0.290	8.82	0.258	0.984	15403	1.645
female	61	0.639	0.361	6.05	0.115	0.967	6803	1.918

Data analyses(2)

Statistical Table of Average Value of Related Variables of Empirical Model (2)

	SAMPLES	SMK	ALC	CHU	GLUCOSE	LDL/ HDL	SBP	PUN	SRT
all	1710	0.237	0.140	0.160	107.5	2.238	113.9	0.146	0.545
male	811	0.43	0.216	0.273	109.0	2.518	117.9	0.178	0.443
female	899	0.062	0.072	0.058	106.3	2.031	110.2	0.117	0.637
[19,30]	358	0.193	0.131	0.115	95.90	1.871	106.4	0.045	0.525
male	165	0.291	0.176	0.182	98.20	2.163	113.0	0.061	0.406
female	193	0.109	0.093	0.057	94.00	1.669	100.7	0.031	0.627
(30,40]	272	0.224	0.132	0.173	100.3	2.163	107.9	0.143	0.555
male	126	0.397	0.190	0.294	102.6	2.570	114.1	0.190	0.476
female	146	0.075	0.082	0.068	98.20	1.877	102.5	0.103	0.623
(40,50]	480	0.271	0.129	0.185	107.4	2.286	113.5	0.171	0.560
male	226	0.487	0.217	0.345	106.6	2.626	116.3	0.177	0.491
female	254	0.079	0.051	0.043	108.2	2.044	111.0	0.165	0.622
(50,60]	477	0.247	0.159	0.164	114.8	2.456	120.3	0.184	0.532
male	232	0.491	0.250	0.272	118.9	2.649	122.5	0.246	0.392
female	245	0.016	0.073	0.061	110.9	2.305	118.2	0.127	0.665
(60,64]	123	0.220	0.154	0.146	129.8	2.522	125.6	0.195	0.569
male	62	0.435	0.242	0.210	122.1	2.529	127.9	0.210	0.484
female	61	0	0.066	0.082	137.7	2.516	123.2	0.180	0.656

Result

Exercise habit is less likely to be overweight, which is consistent with common sense, but the estimated coefficient is not significant.

This is also the point of doubt in this study. Because most people regard exercise habit as an exogenous variable of overweight, and under this condition, exercise habit will not have a significant effect on overweight.

Overweight model estimation results without endogeneity

Independent variable	<i>OVERWT</i> (dependent variable)	
	Coefficient	<i>t value</i>
Constant	-4.8977	(-6.950)***
EXER	-0.2252	(-1.517)
AGE	-0.0038	(-0.610)
SEX	-0.0124	(-0.088)
EDU	-0.0607	(-3.151)***
MARRY	0.1819	(1.217)
JOB	0.1427	(1.081)
HOUSE	-0.4236	(-2.342)***
INCOME	-0.0011	(-0.502)
SMK	-0.5631	(-3.468)***
ALC	0.0164	(0.089)
CHU	0.4715	(2.485)***
EH	-0.1153	(-1.773)*
GLUCOSE	0.0091	(3.597)***
LDL/HDL	0.6361	(8.867)***
SBP	0.0286	(6.705)***
PUN	0.1669	(1.024)
SRT	-0.1066	(-0.925)
The number of samples		1710

Note: 1. *, **, *** indicate the significance level of 10%, 5% and 1% respectively.

2. This part is estimated by Limdep package software.

Source: compiled by this research.

Overidentified and weak IV test

Instrumental variables verification result

Item	Instrumental variable overidentified test		Weak instrumental variable test
	χ^2	P-value	χ^2
Overweight model	0.056	0.89	322.45***

Through the results of instrumental variable overidentified test and weak instrumental variable test, we can confirm the appropriateness of the instrumental variables selected in this study.

Result

As predicted in this study, the average measured value of PUN and SRT have a significant positive relationship with exercise habits.

Higher Pulmonary function measurement value and anterior bending measurement value have incentives to attract people to exercise more.

Estimation results of exercise habit model with instrumental variables

Independent variable	EXER (dependent variable)	
	Coefficient	t value
Constant	-1.0801	(-1.666)*
AGE	-0.0310	(-4.829)***
SEX	0.0971	(0.689)
EDU	-0.0811	(-4.241)***
MARRY	-0.0947	(-0.657)
JOB	-0.1986	(-1.515)
HOUSE	0.3001	(1.631)
INCOME	-0.0031	(-1.179)
SMK	-0.2699	(1.665)*
ALC	0.2377	(1.367)
CHU	1.1714	(6.854)***
EH	-0.0289	(-0.401)
GLUCOSE	0.0039	(2.363)**
LDL/HDL	0.0054	(0.089)
SBP	0.0141	(3.637)***
Instrumental variable		
PUN	0.4565	(2.911)***
SRT	0.3549	(3.066)***
The number of samples		1710

Note: 1. *, **, *** indicate the significance level of 10%, 5% and 1% respectively.

2. This part is estimated by Limdep package software.

Source: compiled by this research.

Result

The result shows that different from the estimation result of the overweight model that does not consider the endogenous exercise habit, after considering the endogeneity of the exercise habit, exercise habit has a significant negative impact on the overweight situation (ρ value also supports the discussion of this relationship). The result is the same as expected and influence of other factors on the overweight has not changed much.

We found that people with exercise habit are significantly less likely to be overweight than people without exercise habit. Education level, owning a house, smoking, eating betel nuts, health Index, Fasting blood glucose, ratio of low-density cholesterol to high-density cholesterol, systolic blood pressure also have a significant impact Its overweight condition.

Estimation result of overweight model with exercise habit as endogenous variable

Independent variable	OVERWT (dependent variable)	
	Coefficient	t value
Constant	-5.7328	(-7.210) ^{***}
EXER	-0.2256	(-1.782) [*]
AGE	-0.0030	(-0.489)
SEX	-0.028	(-0.200)
EDU	-0.0588	(-3.065) ^{***}
MARRY	0.1779	(1.192)
JOB	0.1571	(1.194)
HOUSE	-0.4203	(-2.324) ^{***}
INCOME	-0.0012	(-0.510)
SMK	-0.5616	(-3.462) ^{***}
ALC	0.0240	(0.130)
CHU	0.4724	(2.489) ^{***}
EH	-0.1133	(-1.796) [*]
GLUCOSE	0.0092	(3.650) ^{***}
LDL/HDL	0.6473	(9.079) ^{***}
SBP	0.0285	(6.692) ^{***}
ρ value	-0.364 ^{***}	
The number of samples	1710	

Note: 1. *, **, *** indicate the significance level of 10%, 5% and 1% respectively.

2. This part is estimated by Limdep package software.

Source: compiled by this research.

Marginal effect

Estimated value of the marginal effect of exercise habits on the prevalence of overweight

Influencing factor	Have exercise habit	No exercise habit
Overweight prevalence rate	43.76%	54.83%

When other factors remain unchanged, the prevalence of overweight among those with exercise habit is about **11.07%** lower than that of those without exercise habit.

Discussion

1. Conclusion

- * Data from the Academia Sinica (Nutrition and Health Survey in Taiwan, NAHSIT, 2005~2008), and collected 1710 valid samples.
- * Exercise habit is regarded as an endogenous variable of overweight. We use bivariate probit model to estimate the influence of 19 to 64-year-old population's exercise habits and related factors on their overweight situation, and analyze the marginal effect of exercise habits on the prevalence of personal overweight.
- * Through the Hausman-test endogeneity test, the hypothetical endogenous overweight model is verified.
- * The results show that after considering the endogeneity of exercise, exercise habits have a significant negative impact on the prevalence of overweight. We estimate that if other factors remain unchanged, the prevalence of overweight among those with exercise habit is about 11.07% lower without exercise habit.

Discussion

2. Expected benefits

- * **In terms of society**, through the establishment and quantitative analysis of basic data on exercise habits and overweight, it is helpful to describe and quantify the exercise habits and overweight appearance of the population aged 19 to 64.
- * **In academic terms**, we believe that the exercise habits of the 19-64-year-old population will be affected by non-genetic factors such as personal choice, knowledge and culture, so exercise habit is endogenous. This article uses the econometric model to deal with the problems of endogenous exercise habits and overweight prevalence, which is an important attempt different from other studies.
- * **In terms of economy**, we estimate that the prevalence of overweight is about 11.07% lower among those with exercise habit than those without exercise habit. **It can extend the evaluation model of added value such as the correlation between exercise and obesity medical costs.**

Thanks for your listening

Thanks Taiwan Ministry of Science and Technology for sponsoring this research.

MOST 107-2410-H-431-015 -

