This product is part of the Pardee RAND Graduate School (PRGS) dissertation series. PRGS dissertations are produced by graduate fellows of the Pardee RAND Graduate School, the world's leading producer of Ph.D.'s in policy analysis. The dissertation has been supervised, reviewed, and approved by the graduate fellow's faculty committee.

DISSERTATION

Three Essays on Health Behaviors and the Need for New Policy

Yang Lu

This document was submitted as a dissertation in September 2009 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Dana Goldman (Chair), Jeffrey Wasserman, and Jeanne Ringel.

PARDEE RAND GRADUATE SCHOOL

The Pardee RAND Graduate School dissertation series reproduces dissertations that have been approved by the student's dissertation committee.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Acknowledgments

I would like to thank my committee chair, Dana Goldman, and my two other committee members, Jeffrey Wasserman and Jeanne Ringel, for their insightful guidance and unwavering support. I would also like to thank James Smith for mentoring, outside reviewer Michael Grossman of CUNY for his comprehensive and insightful comments, and Darius Lakdawalla for his essential and constructive advice. I'm also grateful for a generous dissertation award of $25,000 from the RAND Center for Chinese Aging Studies.

I greatly benefited from research assistance from my colleagues and friends, including Ricardo Basurto, Ben Bryant, Richard Bowman, Meena Fernandes, Christine Holten, Erik Meijer, Mary Ann Murphy, Juan Pantano, Bodgan Savych, Baoping Shang, Rachel Swanger, Yuhui Zheng, Daniel Zimmerman, and many others. I have received unconditional love and support from my family; my father has always been a fountain of inspiration.

They have made this long, challenging dissertation process an enjoyable and wonderfully rewarding learning experience. I am very lucky to have them all.

Dissertation Abstract

This dissertation consists of three essays, each on one emerging public health issue that calls for new policy making. The first essay studies 15,000 adult individuals from a longitudinal dataset, the China Health and Nutrition Survey, collected in China 1991-2006. It explores the effects of food prices on obesity and shows evidence that while obesity corresponds to food prices changes, the effects might not always be accurately captured by Body Mass Index (BMI), but by a more direct measure of body fat – triceps skinfold thickness (TSF). The second essay extends the first essay and focuses on health implications of obesity on outcomes such as hypertension and diabetes. I limit the sample to non-obese individuals with a BMI less than 28. TSF, as a proxy for body fat, is shown to have significant independent effects on health. The third essay looks at unintended consequences of a new drug innovation, Viagra, and its successors, Cialis, and Levitra. It finds that erectile dysfunction (ED) medication users have a higher rate of STDs. Because most ED drug consumers are 40+ males, who are above the typical age range where routine STD tests are recommended, this finding reveals a new health threat to older populations and potentially the general public as a whole.

Table of Contents

List of Figures

List of Tables

Chapter I. The Effects of Relative Food Prices on Obesity - Evidence from China: 1991-2006

A driver of health care costs and a threat to quality of life, obesity has become an increasingly serious public health problem worldwide. Since obesity and its induced diseases and mortality are preventable, public policies targeting its causes and mechanisms can be used for effective mitigation and prevention. This paper explores the effects of relative food prices on body weight and body fat over time. I study a cohort of 15,000 adults from over 200 communities in China, using the longitudinal China Health and Nutrition Survey (1991-2006). The main outcomes are measurements of body mass index (BMI) and triceps skinfold thickness (TSF). The key independent variables are food prices indexed from major food groups of edible oils, grains, meat, and vegetables. OLS, Random Effects (RE), Fixed Effects (FE) and IV(GMM) models are employed for the panel data analyses. I find that (1) across all model specifications, decreases in the price of staple oils cause an increase in body fat, whereas, in the individual FE models, such price effects on BMI diminish, and (2) controlling for BMI, the effects of relative oil prices on body fat remain significant. These findings suggest that changes in food consumption patterns induced by varying food prices could increase individuals' percentage body fat to risky levels even without substantial weight gain. This increasing population of individuals with relatively normal weight and high level of body fat might soon become the next victim of the global obesity epidemic. The policy implications are twofold: (1) food prices and subsidies could be used to encourage healthier food consumption patterns and, thus, curb obesity, and (2) BMI is not the most accurate measure for obesity.

1. Background

Obesity is a condition under which the accumulation of body fat is high enough to adversely affect health (World Health Organization 2000). It is a risk factor for an array of non-communicable diseases, including cardiovascular disease (CVD), hypertension, diabetes, and certain types of cancer (World Health Organization 2000).

In industrialized countries, obesity has already been a serious public health issue. In the United States, two thirds of the adult population are overweight, and half of them are obese[1] in 2004 (Ogden *et al.* 2006). In fact, obesity and being overweight has become the second leading cause of preventable death in the U. S., accounting for 400,000 deaths in 2000 (Mokdad *et al.* 2004). Meanwhile in developing countries, an emerging nutrition transition is leading a rapid spread of overweight and obese individuals. China is no exception: between 1992 and 2002, the fraction of Chinese who are either overweight or obese had increased from 14.6% to 21.8% – a growth of 49% (Wang *et al.* 2007). What is even more worrisome is that, even without a BMI greater than 25 or 30, individuals can still be at risk for CVD and diabetes, if they have a high level of body fat – a situation that applies to a considerable portion of the Asian populations (He et al. 2001; Deurenberg-Yap et al. 2002; World Health Organization 2004). Thus, the magnitude of obesity as a public health threat might be underestimated if the BMI cut-off points of 25 and 30 are the only standard used to identify overweight and obese individuals.

[1] Defined here by the WHO body mass index (BMI) cut-off points. Overweight is defined as BMI >= 25 and obesity is defined as BMI >= 30.

This paper models the two main culprits of obesity, energy-dense diets and physical inactivity, with the main focus on the former, and measures the magnitude of their effects on obesity. Specifically, it explores whether variations in food prices can have any effects on obesity by influencing individuals' diet patterns. Given that humans have innate preferences for fat (Drewnowski et al. 1991), I hypothesize that when the price of cooking oil decreases relative to other foods, individuals will consume more of it. And because cooking oil is the most energy-dense[2] of all foods, such a food substitution can result in a much higher level of calorie intake, which subsequently leads to more body fat when excess calories are stored by the body. Findings in current literature support this hypothesis with empirical evidence. Lakdawalla and Philipson (2002) studied US individuals from 1976 to 1994 and showed that falling food prices (as compared to other commodities) due to technological innovation accounted for 40% of the growth in BMI within that time period. And Chou *et al.* (2004) found that reductions in the real and relative prices of fast foods (which are typically high-fat) and availability of fast-food restaurants had a significant impact on obesity. However, there is little research on comparing effects within food categories (such as oils, meats, grains, and vegetables) and there is a lack of literature on effects of food prices on obesity in developing countries including China, where the growth of obesity prevalence is new and fast-paced.

2. Theoretic framework: how relative prices affect diet choice

[2] The energy content of fat is ~9kcal/g, compared to ~4kcal/g for lean meat and grains, and ~0.25kcal/g for green vegetables. Source: The National Nutrition Laboratory of USDA.

2.1. The optimal case

Consider an individual who tries to maximize her utility from food consumption. Suppose she values the amount and taste of food, so high-fat food adds to the joy of eating, and the more calories the better. However, she is also concerned with her physical image and health status. Let's further assume that she is informed that both excess calorie intake and a high percentage of calories from fat will increase her body fat and subsequently the likelihood of developing certain diseases such as hypertension and diabetes. In addition, she understands that although she might damage her physical image by gaining weight from consuming too many calories, she might not gain any noticeable weight by consuming a high percentage of calories from fat, if she can limit the total sum of her calorie intake to a sustaining amount, say 1,800 calories/day. If she is informed of all the above, she solves

$$\max_{c,f} U[V(c-\underline{c}), Y(f), W(c), D(W(c), F(f))] \tag{1.1}$$

s.t.

$$I = P_f c^f + P_{nf} c^{nf} \tag{1.2}$$

$$c = c^f + c^{nf} \tag{1.3}$$

$$f = \frac{c^f}{c} \tag{1.4}$$

$$c \geq \underline{c} \tag{1.5}$$

In this model, c denotes the individual's calorie intake, $\underline{c}$ denotes the minimum of calories she needs to survive, and f denotes percentage calories from fat. V is the utility from consuming calories.

I assume the more calories she consumes, the happier she is, so $V'(\cdot) > 0$. Macro-nutrients, such as carbohydrates, protein, and fat, all contribute to calorie intake. Given humans' innate preferences for fat, I assume the higher f is, the happier the individual will be. So $Y_f' > 0$, where Y is the utility from consuming fat. W is body weight. The individual will try to maintain a normal weight by limiting her total calorie intake. $W(c)$ is monotonically increasing, so $W_c' > 0$. D denotes obesity-induced diseases. It's a function of both weight and body fat. Body fat, F, is a function of f, so I have $D_W' > 0$ and $D_F' > 0$.

According to the assumptions I also have $\frac{\partial U(\cdot)}{\partial V} > 0$, $\frac{\partial U(\cdot)}{\partial Y} > 0$, and $\frac{\partial U(\cdot)}{\partial W} < 0$. Further, if $D(\cdot)$ is addictively separable in W and F, I also have $\frac{\partial U(\cdot)}{\partial D} < 0$.

In the constraints, I is the individual's disposable income and P_f is the relative price of fat. I set the price of other goods P_f to be 1. I assume that fat is a normal good so when price is lower the individual will consume more. Because $c^f = cf$, and $c^{nf} = (1-f)c$, I can rewrite the budget constraint as $I = P_f f c + (1-f)c$. Then I have $c = \frac{I}{(P_f - 1)f + 1}$, and c is then a function of f: $c = \gamma(f)$. Substitute this $\gamma(f)$ into the utility function, I have

$$\max_{c,f} U[V(\gamma(f) - \underline{c}), Y(f), W(\gamma(f)), D(W(\gamma(f)), F(f))] \tag{1.6}$$

FOC w.r.t f:

$$\frac{\partial U(\cdot)}{\partial f} = \frac{\partial V(\cdot)}{\partial f} + \frac{\partial Y(\cdot)}{\partial f} - \frac{\partial W(\cdot)}{\partial f} - \frac{\partial D(\cdot)}{\partial f} = 0 \tag{1.7}$$

Let f^* be the optimal value of f that satisfies this FOC condition. Therefore, $U(f^*)$ is the optimal level of utility for the individual achieved by consuming f^* percent of calories from fat. And $\forall f \neq f^* | U(f) < U(f^*)$, given there are no multiple equilibria.

2.2. The suboptimal case

But, what would happen if the individual were to only observe weight, but not her body fat? This is quite common because it is much easier to weigh oneself than to get a direct measure of body fat. Thus, it is possible that the individual would not achieve the optimal level of utility, if she did not either observe or care about her body fat. In this case, she is ignorant about the potential harms a high body fat might cause her, and she solves her utility function by assuming $D = \tilde{D}(W(c))$ instead of $D = D(W(c), F(f))$, and she would choose $\hat{f}$ to satisfy the FOC condition as follows:

$$\frac{\partial U(\cdot)}{\partial f} = \frac{\partial V(\cdot)}{\partial f} + \frac{\partial Y(\cdot)}{\partial f} - \frac{\partial W(\cdot)}{\partial f} - [D'_W(\cdot)\frac{\partial W(\cdot)}{\partial c}\frac{\partial \gamma(\cdot)}{\partial F}] = 0 \qquad (1.8)$$

The difference between equation (1.8) and (1.7) is that, here, she sets $D'_F(\cdot)\frac{\partial F(\cdot)}{\partial f} = 0$. In the optimal case, $D'_F(\cdot)\frac{\partial F(\cdot)}{\partial f}$ is likely to be non-zero and positive when $f = f^*$, due to the concave shape of D_f. And also because $D'_f > 0$ and $D''_f < 0$, the $\hat{f}$ the individual chooses in the suboptimal case is likely to be higher than f^*. Subsequently, the disutility $\tilde{D}$ will be greater than D^*, and $\tilde{U} < U^*$. The individual will then suffer from a suboptimal level of utility by failing to link a high-fat diet with obesity-induced diseases.

3. Data

I use the China Health and Nutrition Survey (CHNS), which is an on-going panel of individuals from over 200 communities in 9 provinces.[3] CHNS collects longitudinal data on demographics, anthropometric measurements, health indicators, and community-level commodity prices. It began in 1989 and followed the participants subsequently in 1991, 1993, 1997, 2000, 2004, and 2006. I use data from all the waves except 1989 because most food prices in 1989 were determined by the central government rather than market-driven and only a limited sample of individuals were interviewed for health and nutrition information in that year.

For each wave, I limit the sample to adult individuals who are between 18 and 75 years old at the time of interview. For women, I exclude the waves when they were pregnant, to avoid irrelevant weight shocks. This results in 37,816 individual-year observations in total. It is an unbalanced dataset, and there are about 5,000 - 8,000 individuals in each interview year,[4] and 15,649 individuals in total.

The key outcome is obesity, or adiposity. Obesity is defined as the condition of having an abnormally high proportion of body fat (National Institutes of Health 1998, p174). In order to better capture effects on obesity, I adopt two measures to indicate the extent of body fat - body mass index (BMI) and triceps skinfold thickness (TSF), both of which are commonly used for measuring adiposity (Must et al. 1991).

[3] The nine participating provinces include Liaoning, Heilongjiang, and Shandong in the northeast, Jiangsu in the east, Henan, Hubei, and Hunan in the middle, Guizhou in the southwest, and Guangxi in the south. Heilongjiang was not surveyed in 1991 and 1993, and Liaoning was not surveyed in 1997. (Source: http://www.cpc.unc.edu/projects/china/design/sample.html).

[4] Numbers of observations vary by year due to attrition and addition of participating individuals and provinces.

BMI is traditionally used to classify obesity as a surrogate measure for percentage body fat, mainly because it is convenient to obtain and easy to interpret. It is defined as weight in kilograms divided by the square of height in meters: BMI = weight(kg)/height squared (m^2). An individual with a BMI ≥ 25 is classified as overweight and is further considered obese if the BMI ≥ 30 (National Institutes of Health 1998). To construct BMI from the data, I use weight values from all interview waves and the first normal, non-missing height.[5] BMI is easy to calculate and widely used. However, because it is derived from only height and weight, it does not convey information on body composition, which is defined as the ratio of lean body mass to body fat mass. Thus, its accuracy in measuring body fat varies by muscularity, age, gender (Gallagher *et al.* 1996), and ethnicity (Calle *et al.* 1999). Specifically, BMI tends to overestimate body fat in muscular subjects and to underestimate percentage body fat of the elderly and of certain ethnicity groups such as Asians (Deurenberg et al. 1998). Due to this limitation, BMI can be a weak indicator for adiposity of certain population sub-groups (Piers et al. 2000; Deurenberg-Yap et al. 2002; Burkhauser and Cawley 2008). In fact, overweight and mild obesity indicated by BMI alone have shown to not be an independent risk factor for cardiovascular diseases (Romero-Corral *et al.* 2006).

Concerning the limitation of BMI as a proxy for obesity, I use a second measure for body fat - TSF. TSF is a vertical skinfold measured at the posterior midpoint between the acromion and the olecranon. It directly measures subcutaneous body fat, and is widely used to measure body composition in clinical studies. It is considered a more appropriate measure for obesity and body fat than BMI (Rolland-Cachera *et al.* 1997). Moreover, unlike BMI, TSF, or the logarithm of TSF, is consistently shown to be a linear function of body fat (Durnin and Womersley 1974; Lean et al. 1996). For example, Lean et al. (1996) showed that percent body fat increases by 0.76 percentage

[5] Height is measured in every wave but I found there are more abnormal values in the latest waves of 2004 and 2006. Because the height of an adult is pretty stable (in this sample the decrease in height is about 0.003cm every additional year of age), for each individual, we use the first non-missing height value that is within the normal range.

points for men and 0.67 percentage points for women, when there is a 1mm increase in TSF, holding all else equal. The limitation of TSF is that it does not directly calibrate visceral fat, and thus might underestimate the total body fat mass in obese individuals with high visceral fat (Williams *et al.* 1992, p604). Therefore I discard the very few records where BMI is higher than 35, and limit the analysis to non-morbidly obese individuals (with a BMI of 35 or higher). In addition, I also refer to the Anthropometric Reference Data from NHANES III (1988-1994)[6] and limit TSF values to a reasonable range of 3 – 40 mm.

The five key independent variables are food prices, in relative or absolute terms. They are: (1) price of staple oil relative to staple food, (2) price of staple oil to lean pork, (3) price of staple oil to commonly consumed vegetables, (4) price of staple oil by itself, and (5) price of staple oil to government-recommended intake of protein and carbohydrate, respectively.

These price variables are mainly relative and absolute prices of cooking oil. I construct the prices by taking into consideration how food is usually prepared and consumed. Grains (staple food), meats, and vegetables are the main food sources, and they are usually prepared with a certain amount of cooking oil. Because cooking oil is the most caloric of all foods, the fraction of oil used in food preparation helps determine the sum of calories a meal contains. And because the amount of oil to use in a dish is often discretionary, it allows variations in the portion of oil, and therefore variations in percentage calories from oil intake.

I use the five oil price variables separately, each in a different regression model. When the key regressor covers certain food categories, prices of other main food categories are also controlled for. For example, if the price of staple oil to staple food is the key regressor, then the price of lean pork and the price of commonly-consumed vegetables are being controlled for. And if the price of staple

[6] Source: http://www.cdc.gov/nchs/about/major/nhanes/Anthropometric%20Measures.htm.

oil by itself is the key regressor, then the price of staple food, the price of lean pork, and the price of vegetables are being controlled for.

In these prices effect models, oil and other foods are considered complementary to each other. My hypothesis is that consumption of foods corresponds to how expensive the foods are, and when the relative price of oil decreases, people will choose to use more oil when preparing meals, and vice versa. Subsequently, higher percentage of oil intake leads to more adiposity.

By analyzing five separate models, I can test whether the price effects are persistent and robust, and can obtain more information on how the magnitude of price effects changes when the comparison group is different. Below I describe in details how prices of each food category are coded.

Out of all edible oils with price infomation available, three (soybean, rapeseed, and peanut) are identified as staple oils as they account for more than 80% of the edible oil consumption in a certain region of China (Fang and Beghin 2002). Accordingly, I assign each participating province the price of its staple oil, depending on the region it belongs to. Namely, it is soybean oil for the northeastern provinces, including Liaoning, Heilongjiang, Shandong; rapeseed oil for the provinces in the middle and west, including Jiangsu, Henan, Hubei, Hunan, Guizhou; peanut oil for the southern province Guangxi.

The prices of staple food/grains are also region-specific, because the north and south have different staple foods. I use the average price of wheat and noodle for the northern region (Liaoning, Heilongjiang, Shandong, Henan), and the average price of rice and noodle for the south (Jiangsu, Hubei, Hunan, Guangxi, Guizhou).

I use lean pork to represent meats because it is the main source of animal protein in China. For vegetables, I use the average price of all the vegetables available in the survey: cabbage, rape[7], and "other commonly eaten vegetables".

The last price variable is the ratio of the price of staple oil to the price of a combination of protein and carbohydrates. I construct this key variable by acknowledging protein and carbohydrate as the main contributors of a regular meal in terms of calories. In theory, I can either model intake with real consumption patterns or with ideal proportions. Real consumption data change frequently and are hard to obtain, so I use ideal proportions by referring to the Dietary Guidelines for Chinese Residents 2007 by the Chinese Nutrition Society, in which the recommended daily intake of meats and gains in terms of calories for adults is about 1:2 to 1:3. In the analysis I use three patterns of intake combinations, 1:3, 1:2, and 1:1, mainly to check robustness of the price effects. I will describe in more detail in the results section but the three price ratios return similar and comparable coefficients on obesity. So due to space limitation, I only show the price ratio of 1:2 in this paper.

To summarize, these five food price variables can be written as: $\frac{P_{staple\ oil}}{P_{staple\ food}}$, $\frac{P_{staple\ oil}}{P_{lean\ pork}}$, $\frac{P_{staple\ oil}}{P_{vegetables}}$, $P_{staple\ oil}$, and $\frac{P_{staple\ oil}}{\frac{1}{3}P_{staple\ food}+\frac{2}{3}P_{lean\ pork}}$.

For all food categories, I use free market prices by default, and substitute with either state store market prices (from all interviews occurred in 1991, 1993, and 1997) or large store retail prices (from interviews occurred in 2000, 2004, and 2006) wherever free market prices are missing. I refer to the

[7] A leafy green vegetable in the mustard family, similar to Canola.

national food price data published by the Chinese Department of Agriculture,[8] in order to identify and to exclude abnormal price values as a way of data quality control.

Besides the key variables, I include covariates such as gender, age, age squared, type of residence (urban or rural), education attainment, income levels, physical activity levels, and year dummies.

I construct five levels of education attainment: (1) no education, (2) elementary school, (3) middle school, (4) high school diploma or equivalent (including technological and vocational school education), and (5) bachelor's degree and beyond.

There are two types of income variables in CHNS: personal income and household income. I use household income divided by household size to calculate household income per person, because (1) there are significantly fewer missing data points in the variable of household income than in personal income; (2) food preparation and consumption often do not vary within the household level in China, and per capita household income can better capture the level of disposal income for food than personal income.

Physical activity levels are important factors to include because they directly affect obesity. There are five levels of physical activity available in CHNS, based on how strenuous an individual's job is. I re-categorize them into three levels: light, moderate, and heavy, mainly to avoid a potential lack of power due to limited observations in certain activity levels. "Light" includes very light activities such as sedentary office work, and light activities such as jobs that have to be completed while standing, like those of a salesperson's or of a teacher's. "Moderate" includes occupations such as student,

[8] This dataset collects daily prices of several hundred different food items at large free markets nation-wide, for the past 10 years. All provinces are covered, though prices are not available for all free markets at all times. Source: http://www.agri.gov.cn/vegetable/sort.asp

electrician, and metal worker. "Heavy" includes heavy and very heavy activities, covering occupations such as farmer, dancer, athlete, miner, and logger.

Summary statistics are shown in Table I-1, for the whole sample, and the female and male subsamples. As we can see, there are about equal numbers of men and women. Rural residents constitute two-thirds of the sample, and urban residents one-third. Obesity-wise, both females and males have an average of BMI around 22.5, while females have a slightly higher index. In contrast, there is a much more distinctive difference between men and women in terms of TSF/body fat. Females have a TSF of 15.91 mm, 37% higher than that of men. The difference in body fat is nonetheless expected to a certain extent, due to biological differences between men and women.

The food prices provide a basic idea on how much each food item costs (note that they are a pooled average over a time period of 15 years). Lean pork is the most expensive of all, with the mean at 7 yuan per 500g. The second most expensive is staple oil, which costs slightly more than half of the price of lean pork. Staple food, representative of flour, rice, and noodles, cost about 1.2 yuan per 500g. The least expensive is vegetables, costing about half a yuan per 500g.

In terms of physical activities, almost half of all the individuals belong to the "heavy" category, mainly because most of the interviewees are rural residents, and among them many are farmers.

Perhaps the most striking demographic is education attainment. Only about 23% of all interviewees have an education attainment higher than or equivalent to a high school diploma, and only one in twenty interviewees have a bachelor's degree or higher. Males are more educated than women. 85% of males have at least some years of education, but only two-thirds of the females do.

Below I describe the time trend of the key regressors and obesity outcomes of interest, in addition to the mean values for the whole time period presented in Table I-1.

Figure I-1 shows the percentage changes of three relative oil prices from 1991 to 2006, including price of oil relative to staple food, price of oil relative to lean pork, and price of oil relative to vegetables. All food prices in nominal terms are increasing every year (not shown in the figure), due to inflation and economic growth, but the rate of increase in oil prices is so much lower than in the other three foods that the relative prices of oil unanimously decrease over time. If we use 1991 as the reference year, the percentage decrease in the relative prices of oil range from 24% in the price of oil relative to lean pork to 58% in the price of oil relative to vegetables in 2006. There is also a 35% decrease in the price of oil relative to staple food. The price ratio of oil to vegetables has most prominent variation of all: staple cooking oil was more than 13 times more expensive than commonly-consumed vegetables back in 1991, and that ratio has reduced to a little over 6 in 2006 (numbers not shown in the figure).

Within the same time period, the prevalence of overweight and obesity skyrocketed by the definition of BMI cutoffs: the fraction of overweight ($25 \leq BMI < 30$) individuals increased from 12.4% to 23.5%, and the fraction of obese ($BMI \geq 30$) individuals rose from 1.1% to 3.2%.

However, in continuous terms, the trend does not look as prominent, because one unit change in BMI often means a substantial change in body weight. In Figure I-2, it shows the trend of BMI and TSF from 1991 to 2006. (Note: the figure is not age-adjusted but the general increasing trend holds true either way.) The left chart shows changes over time in absolute terms, in which the mean BMI started at around 21.8 in 1991 and grew to 23.2 in 2006 (a1.4-point increase), and the mean TSF increased from 11.4mm in 1991 to almost 16.7mm in 2006. The chart on the right is the percentage changes over time. BMI is 6.5% higher in 2006 than in 1991. The increase in TSF is more substantial: it was 46.7% higher in 2006 compared to the 1991 level. Could this mean the growth of adiposity is faster-paced than that of body weight among Chinese individuals 1991-2006?

Also, was there a connection between the decrease in relative oil prices and the rise in obesity between 1991 and 2006? While the general time trend suggests that relative prices of energy-dense foods are negatively associated with obesity, the correlation could well have been due to chance. So Figure I-3 further looks at how changes in BMI and TSF are associated with relative oil prices at the community level. It tests whether the correlation between food prices and obesity would still remain strong, with the assumption that communities are heterogeneous enough that if "chance" had only occurred at the national level it would not have influenced on both prices and obesity at a finer level.

Figure I-3 shows two scatter plots with a common x-axis – percentage change in the price of oil relative to recommended protein and carbohydrate intake (or recommended meal in short). The plot on the left depicts the relationship of price with percentage change in BMI on the y-axis. Every plus sign in the plot represents one community. The plot on the right shows the correlation between percentage change in TSF and percentage change in the relative price of oil. Similarly, every circle represents one community. The majority of the communities experienced a decrease in the relative oil price and a rise in BMI and TSF, resembling the general trend. However a few of them witnessed an increase in the relative oil price and a drop in BMI and TSF, which is the opposite of the bigger picture, suggesting heterogeneity at the community level. In both scatter plots, the relationship between the relative oil price and obesity is negative and statistically significant, but the downward slope is steeper for TSF, or body fat, than for body weight.

4. Methods

In this section, I will describe the analyses I employed to explore a causal relationship between food prices and obesity, which involve a series of food price effects regressions at the individual person

level. To test whether the effects are persistent and robust, I use all the possible model specifications including pooled ordinary least squares regressions (OLS), random effects (RE) models, and fixed effects (FE) models. For all the regression models, obesity is a function of community-level relative oil prices, and individual person-level characteristics that can be time-invariant or time-varying, such as age, gender, education attainment, type of residence, physical activity levels, and year fixed effects.

"Obesity" as the outcome is measured separately by continuous variables BMI and TSF, in levels and logs, for each individual i in community j at time t. Taking the FE model as an example, "price" is the price of oil relative to staple food, to lean pork, to vegetables, to recommended meal, or by itself, respectively. X_{ijt} is the set of individual characteristics. τ_t is the time fixed effect. u_i is the individual fixed effect, and ε_{ijt} is the error term. Notation-wise, please see Equation (1.9) with BMI being the outcome of interest.

$$BMI_{ijt} = \beta_1 price_{jt} + \beta_k X_{ijt} + \tau_t + u_i + \varepsilon_{ijt} \quad (1.9)$$

When TSF is used as the outcome of obesity, for each model specification I estimate two regressions, one with BMI being a covariate, and one without. Please see Equations (1.10) and (1.11). This is because in the data, although BMI and TSF are both measures for obesity, they are not perfect substitutes for each other. In fact, the correlation between the two is only 0.45. Therefore, by controlling for BMI when the outcome is TSF, we are able to see whether body fat would still be affected by food prices when body weight is held constant.

$$TSF_{ijt} = \beta_1 price_{jt} + \beta_k X_{ijt} + \tau_t + u_i + \varepsilon_{ijt} \quad (1.10)$$

$$TSF_{ijt} = \beta_1 price_{jt} + \beta_k X_{ijt} + \beta_2 BMI_{ijt} + \tau_t + u_i + \varepsilon_{ijt} \quad (1.11)$$

In all models I estimate the coefficients with heteroscedasticity-robust standard errors clustered at the community level. This is the strictest assumption one can make with the dataset. Bertrand *et al.* (2004) suggested that even in differences-in-differences models (FE models with time fixed effects in this case), the standard deviation of the estimated treatment effects might be underestimated in the presence of serial correlation, when conventional standard errors are used. To be most cautious in interpreting the price effects, I use heteroscedasticity-robust standard errors clustered at the community level. Compared to the conventional standard errors, these clustered robust standard errors do not change point estimates of the coefficients, but often produce much larger standard errors and therefore decrease the likelihood for a coefficient to be statistically significant. However, this is necessary for correcting heteroscedasticity and serial correlation in at the community level where variation in the key regressors – food prices occurs.

5. Results

Table I-2, Table I-3, Table I-4, and Table I-5 show the effects on BMI and TSF by the price of staple oil relative to staple food, price of staple oil relative to lean pork, price of staple oil relative to vegetables, price of staple oil, and price of staple oil relative to recommended meal, respectively. All the regressions shown are log-log models, partly because it is easier to interpret the coefficients as price elasticities. In each table, there are six regressions. They are listed side by side for easy comparisons between treatment effects on body weight and that on body fat. The first three on the left are on the outcome of BMI, and the other three are on the outcome of TSF with BMI being controlled for. Regressions with the outcome of TSF without BMI being included are analyzed but

not included in the tables because (1) the price effects on body fat do not differ significantly when BMI is added as a control variable, and (2) space is limited.

For each outcome, an OLS, a RE, and a FE model is shown. Regression models in levels are also analyzed. The coefficients estimated have the same expected sign and the statistical significance is comparable to that in the log-log models. Those results are not shown due to space limitation, but are available upon request.

Across all model specifications, relative oil prices are generally shown to be negatively associated with TSF, and thus body fat, except in the FE model of oil price relative to vegetables. This confirms the hypothesis that when oil prices are lower, body fat will increase. I deduce that consumption of staple oil and energy-dense foods goes up as a response to changes in oil prices. This middle step, changes in food consumption, is not demonstrated in the reduced-form regressions, but it is reasonable to assume so because it is well documented elsewhere. Ng et al. (2008) found that increased consumption of cooking oil at the individual-person level was associated with decreasing prices – empirical evidence from the same CHNS dataset but with a shorter panel, from 1991 to 2000.

In contrast, the expected negative treatment effects on BMI only appear in some model specifications, all of which OLS regressions. Those effects vanish in the RE and FE regressions. However, the key regressor, price of oil relative to vegetables, appears to be an outlier with a positive and significant effect on body weight, which I will discuss in more detail when I describe results from Table I-4.

In Table I-2, there are six log-log models with one key regressor: the relative price of staple oil to staple food. In the three models on the left with BMI being the outcome, we see that the key

regressor does not appear to have any treatment effects on body weight. The sign is negative in the OLS model, which is expected, but it flips in RE and FE models, and is never statistically significant. However, the effects of non-food price control variables are reasonably in accordance with common beliefs. A moderate or heavy physical activity level is associated with less body weight as compared to light. Age and age squared are strong predictors for body weight, suggesting non-linear, quadratic age effects. Income is also positively associated with body weight, except the effects fade away in the FE model. In the OLS and RE model we are also able to see the effects of individual-specific, time-invariant variables such as type of residence, and gender. Both being an urban resident and a female are positively associated with body weight. Education attainment levels are included in the regressions but the coefficients are not shown here because most times they do not appear to be statistically significant, or interesting. Year fixed effects are considerable and statistically significant in all regressions. They are not shown here either because without further efforts we are not able to isolate and specify effects of individual events happened in those years that had an influence on obesity.

Still in Table I-2, the three regressions on the right are on the outcome of TSF, or body fat. There appears to be significant price effects on body fat in all three models, although the magnitude wanes when model specifications make more stringent assumptions about the data. The price coefficient is -0.15 in OLS, -0.13 in RE, and -.08 in FE. But what do price effects imply?

Suppose in one community, the price of staple food is 4 per 500 gram, and the unit price of staple oil is 8. The relative price of staple oil to staple food is 8÷4= 2. Then, if the price of staple oil increases from 8 to 9, while the price of staple food increases from 4 to 5, the price of staple oil to staple food will change to 9÷5= 1.8. And that is a (1.8-2) ÷ 2 = -10% change from the original relative price. In the OLS model, holding all else equal, a 10% decrease in the price of staple oil

relative to staple food leads to a 1.5% increase in TSF. With the same scenario, the treatment effect is reduced to a 1.3% increase in TSF in the RE model, and a 0.8% increase in the FE model.

Although the treatment effects are smaller in the FE model than in the OLS and RE models, they are in general much more reliable point estimates. There are a few reasons for that. First, in order for OLS estimators to be consistent, the error term has to be i.i.d. (independent and identically distributed), or in other words, random, and uncorrelated with the control variables, which is an ideal case that often does not exist in the real world. Second, RE models a composite error term $(u_i + \varepsilon_{ijt})$ instead of identifying each individual effect like FE model does, assuming orthogonality between individual effect u_i and control variables X_{ijt}.

However, unobservable individual characteristics absorbed by u_i can be correlated with observed variables and can lead to biased point estimates. There are at least two ways to test whether a RE model produces consistent results. One is to run a FE model where u_i will be estimated, and check the correlation between u_i and Xb. If the correlation is strong, it means that the orthogonality assumption a RE model makes is invalid, and therefore the RE estimator is inconsistent. If the correlation is weak or close to zero, then both the FE and the RE estimators are consistent, while the RE model is more efficient. The other way is to run both RE and FE models and conduct a Hausman test to see whether we can reject the null hypothesis that the RE estimator is consistent. I use the first method to evaluate whether the RE estimator is appropriate, because the statistic is automatically calculated for the regression model by Stata when a FE model is being estimated. In this specific FE model with price of staple oil relative to staple food being the key regressor, the correlation between individual fixed effects and control variables is -0.9988. The absolute value is very close to 1, suggesting the orthogonality assumption is invalid, and that the FE model should be

more appropriate in estimating the price effects. In fact, the correlation between u_i and control variables X_{ijt} in all the other FE models estimated is much close to 1 in absolute value, similar to the current example, so from now on, I will focus mainly on interpreting the FE estimators for all key regressors.

In the last column of Table I-2 where TSF is the outcome of interest in a FE model, the price of lean pork is negatively associated with body fat, while the price of vegetables has little, if any, effects on body fat at all. BMI is found to be a strong predictor for body fat, as expected. The estimated coefficient says that when BMI increases by 10%, body fat will increase by 12.6%, holding all else equal. Because the point estimate/elasticity is greater than one, it suggests that within individuals, body fat may be growing at a faster rate than is body weight, holding all else equal. Interestingly, within the same individual, changes in physical activity levels do not seem to affect body fat significantly, holding all else equal.

Lastly, by comparing the FE estimators between the two obesity outcomes, I find that effects of staple oil price relative to staple food on obesity are manifested in TSF, a more direct measure of body fat, but are not captured by BMI, an equivalent to body weight.

Table I-3 shows treatment effects of oil price relative to lean pork. The six regressions here explore to what extent individuals would substitute animal protein with cooking oil when price differences are noticeable, and how this would affect their body weight and body fat. The first three models model prices effects on the outcome of BMI. The relative oil price appears to be negatively associated with BMI, suggesting that when the relative price of oil becomes lower than that of lean pork, there would be a higher percentage of calories from oil consumption and a higher calorie

intake in total, which eventually lead to a higher body weight. However, the price effect loses its statistical significance in the FE model.

In contrast, when TSF is the outcome of interest, all three regression models indicate negative and significant treatment effects: the estimation of price effects varies from -0.16 in OLS to -0.14 in RE to -0.08 in FE. If we take the FE estimation at face value, holding all else equal, a 10% decrease in the price of staple oil relative to lean pork will result in a 0.8% increase in TSF, or body fat. The effects are comparable to that from Table I-2.

Results represented in Table I-4 are six regression models with the price of staple oil relative to vegetables as the key regressor. In the third model, the FE model on BMI, holding all else equal, a 10% decrease in the price of staple oil relative to vegetables results in a 0.04% decrease in BMI ($p<0.01$). Such a positive yet very small price effect is perplexing because it is contradictory to the common belief that lower oil price induces more oil consumption, and obesity. This effect is also quite different from what is presented in Table I-2 and Table I-3, where little or no relative oil price effects are shown on BMI. This positive effect is also contradictory to the general trend of food prices and BMI. The price of oil relative to vegetables has gone down between 1991 and 2006, as shown before, and at the same time BMI has increased, suggesting a negative correlation.

One possible explanation for this is that vegetables are a Giffen good, and when they are priced higher there will be more consumption. But there is little literature supporting this point. Alternatively, it could be price endogeneity – healthier, light-weighted individuals have a strong preference for vegetables, driving both prices and consumption up. A third possibility is that it is a type I error. None of these hypotheses can be tested here, however. But it would be easier to at least

explore the first hypothesis, when prices of each food category are being controlled for individually, as presented in Table I-5. I will discuss this further after I describe the TSF part of Table I-5.

Results on the outcome of TSF are also different from that of earlier model specifications. The price effects are negative and significant in OLS and RE but that statistical significance is lost in the FE model, suggesting that we cannot reject the null hypothesis the price of oil relative to vegetables has no effects on body fat.

Table I-5 shows results from six regressions with the key regressor being the absolute price of staple oil. Prices of staple food, lean pork, and vegetables are also being controlled for individually.

When BMI is the outcome of interest, in the FE model, it appears that the price of staple oil has no treatment effects on BMI. Actually, the only statistically significant price effect ($p<0.05$) here is from lean pork. When the price of staple oil and everything else is held constant, a higher price of lean pork predicts a higher body weight. The effect is small (point estimate 0.006), but significant. The price of vegetables is negative but only border-line significant ($p<0.1$), suggesting little to no effects on BMI, when other food prices are being held constant.

The results suggesting little price effects of vegetables are more reasonable than assuming negative effects or in other words a Giffen-good phenomenon. One good reason is that vegetables contain very few calories. In fact, there is only 0.25 calories per gram of vegetables on average. So even if the consumption of vegetables doubled within the same time period, it wouldn't add enough calories to significant increase calorie intake, or increase body weight, not to mention the fact that many fiber-rich vegetables are filling, so by consuming more of them people might eat less in general.

Also there is empirical evidence that consumption of vegetables do not increase when their prices go up. Aggregate-level data provided in the 2007 Agricultural Development Report by the Ministry of Agriculture in China show that annual consumption of vegetables per capita has been decreasing over time: from 127kg in 1991 to 101kg in 2006 for urban residents, and from 132kg in 1991 to 118kg in 2006 for rural residents.[9] Within the same time frame, prices of vegetables have increased. Connecting the two, I exclude the possibility that vegetables are a Giffen good. Price endogeneity or type I error are more plausible candidates.

Still in the same table, when TSF is the outcome of interest, prices effects are all statistically significant with an expected sign. The point estimate for oil price effect is -0.09 in the FE model ($p<0.01$), the magnitude quite comparable to that of the relate oil prices compared to staple food, or to lean pork. The price effects of staple food and lean pork on TSF are positive, when the pricing of oil is being held constant, which is in accordance with my hypothesis. The price of vegetables does not appear to have any treatment effects on TSF, or body fat.

Table I-6 is the last table displaying price effects regressions for the whole sample. The key regressor shown here is the price of oil relative to recommended meal. Again, in the FE models, the relative oil price appears to have no effects on body weight, and is shown to have a negative and significant ($p<0.01$) effect on body fat (TSF), even when BMI is being controlled for. The magnitude of the effect is even greater than earlier models: a 10% decrease in the price of oil relative to recommended meal will increase TSF by 1.2%, holding all else equal.

By look at results from all model specifications, it suggests that effects of both relative and absolute oil prices on TSF, or body fat, are persistent and robust. However, BMI does not seem to capture

[9] Source: http://www.agri.gov.cn/sjzl/baipsh/WB2007.htm#24.

these price effects. The only exception to the generalization is the price of staple oil relative to vegetables.

This implies that individuals consume more edible oils when the prices are lower, and additional oil consumption subsequently leads to a higher level of body fat. At the same time, individuals are likely to keep total calories consumed per day to a certain number, and reduce the consumption of other foods, such as rice, flour, meats, and vegetables, to compensate for more oil intake, such that the total calorie intake will not change substantially, but a higher fraction of calories will come from fat.

5.1. Extended Results by Gender

Although FE models are shown to be superior to RE in this paper (because observable individual characteristic are highly correlated with the control variables), the limitation of FE models remains – FE models do not estimate effects of individual characteristics that are invariant over time because those effects are modeled as part of individual fixed effects by definition. One way to overcome that limitation is to analyze fully-interacted models, or in other words, to run regressions using subsamples divided by a variable of interest. I employ this technique and estimate price effects within subsamples by gender to see if and how treatment effects differ between females and males.

Table I-7 shows the subsample results (log-log models only). The first column of results shows the regression on the outcome of BMI, and the second is on TSF. Each block of a coefficient and a standard error represents one regression. There are altogether 20 regressions in the table. They are all FE models with robust standard errors clustered at the community level.

When BMI is the outcome, all the price effects are not statistically significant except the price of oil relative to vegetables in the male subsample. When TSF is the outcome, mostly the price effects are comparable between females and males, except for the price of staple oil by itself. In the male subsample, the absolute oil price effect is not statistically significant, but in the female subsample, the p-value is less than 0.01.

6. Endogeneity

Price effects shown earlier cannot be considered causal if there is price endogeneity. In a free and competitive market, prices are co-determined by supply and demand. Self selection from the demand side and reverse causality are two legitimate concerns regarding the estimated price effects from earlier regression models. Therefore I use instrumental variables (IV) to test if price endogeneity is present and to further estimate causal price effects on obesity.

An instrument is considered valid and relevant in this particular case if (1) the instrument is associated with food prices from the supply side, such as a production cost, or a transportation cost, and (2) the instrument itself does not in any other way affect body weight or body fat except through food prices. The price of gasoline is available from the CHNS community survey and meets the two above requirements. First, gasoline is directly used for vehicles that transport and distribute food, so it is part of the transportation cost of food, from the supply side. The relationship between the price of gasoline and the price of food prices can be tested in the first stage of IV. Second, gasoline is unlikely to be directly linked with body weight and body fat. This condition might be violated in developed countries where car ownership is so common that when the price of gasoline is lower, individuals might choose to substitute walking with driving, and store more calories and fat

in the body. But in China, very few people have a car at their disposal, and most people commute by foot, bike, or bus. So, changes in gasoline prices supposedly would not change individuals' physical activity levels, or other behaviors related to obesity

For all the IV estimates, only FE-IV models are shown in Table I-8 and Table I-9 due to space limitation. There are altogether ten shown regressions, five on the outcome of BMI, five on TSF.

The first stage of these IV regressions is not shown, but for each regression I run a weak identification test and an under-identification test on the validity of the instrument. The weak identification test uses Kleibergen-Paap rk Wald F statistic and the under-identification test uses Kleibergen-Paap rk LM statistic (Baum *et al.* 2007). The statistics are shown near the bottom of the tables. In all models these statistics are large enough to reject the null hypothesis that instrument is not valid.

I also perform endogeneity tests of the price variables in OLS-IV models, which suggest that the price of staple oil is endogenous. But because it is in essence a Hausman test under conditional homoskedasticity (Baum *et al.* 2007), it is not available for the FE-IV models.

I find that in the FE-IV models, prices effects remain all negative and statistically significant on TSF, echoing what I find from earlier FE models. Even the price of oil relative to vegetables is shown to influence TSF. And the magnitude of effects is now larger. For example, the effect of staple oil price relative to staple food has risen from -0.08 in FE to -0.214 in the FE-IV model. The coefficient of -0.214 implies if the price of oil relative to staple food decreases by 10%, TSF will increase by about 2.1%, holding all else equal. However, does this instrumented point estimate sound too large to be reasonable? From 1991 to 2006, the price of oil relative to staple food decreased by 35%. Applying the instrumented point estimate, this price change led to a $35\% \times 0.214 \approx 7.5\%$ increase in TSF.

Within the same time period, TSF increased by 46.7%. So, the price of staple oil relative to staple food explains 7.5%÷46.7%=16% of the total increase in TSF. Given that excess food intake and physical inactivity are the two main contributors of high body fat, a point estimate indicating that food price accounts for 16% of the changes in TSF hardly sounds absurd.

On the other hand, the price effects on BMI are all shown to be negative and statistically significant as well, which is aligned with the findings on the oucome of TSF, but different from what I find in the FE models on BMI earlier. In the same example of the price of staple oil relative to staple food, the earlier FE model gives us a coefficient of -0.005 which is negative yet insignificant, but once instrumented, the coefficient is changed to -0.039 and is statistically significant ($p < 0.01$). It says, holding all else equal, if the price of staple oil relative to staple food increases by 10%, an average person's BMI will decrease by about 0.4%. If I take the results at face value, it implies that when relative oil prices are lower, individuals are likely to gain body weight and increase percentage body fat. However, price effects are much larger on body fat than body weight, suggesting body fat might have been growing at a faster rate than body weight.

I do not have enough information to conclude whether regression results from one model are definitively more reliable than those from another, but at least the IV method reconfirms the hypothesis that relative oil prices have negative treatment effects on body weight and body fat. Oil prices are demonstrated to have a greater impact on body fat than on body weight. And the results are consistently shown to be robust.

7. Conclusion

Results from the reduced-form price effects models suggest that food prices affect individuals' body fat levels, probably by changing the composition of meals they consume – when the relative price of

staple oil is lower, obesity increases, captured by changes in TSF. In addition, by using any constructed relative oil price, across all model specifications, price effects on body fat are significant and persistent. However, such price effects appear to be more subtle on BMI and cannot always be captured, especially in the FE models (though reflected in FE-IV models). This is partly because BMI is not a perfect measure of body fat. It is by definition the height-adjusted body weight and does not convey information on body composition such as percentage body fat and muscularity.

To summarize, all regressions results show that oil consumption can correspond to relative oil prices and it is shown to subsequently increase individuals' body fat at a faster rate than it affects body weight. The policy implications are two-fold. First, by using different pricing policies, it is possible to effectively induce healthier food consumption patterns and thus to control the growth of obesity. Second, BMI contains limited information on body fat and is a less accurate measure for obesity especially when it applies to certain populations, such as Asians.[10] It is important to raise public awareness that individuals need to be cautious when interpreting their degree of adiposity by using BMI alone, and that it goes beyond maintaining a normal body weight to ensure a healthy level of percentage body fat. If more direct measures of body fat such as TSF are too costly or simply unavailable, it would still help to watch and control daily oil consumption levels.

[10] The WHO and NIH definition of obesity indicated by a BMI between 25 and 30 is derived by measuring mostly young white males only.

Figure I-1 Price Trend

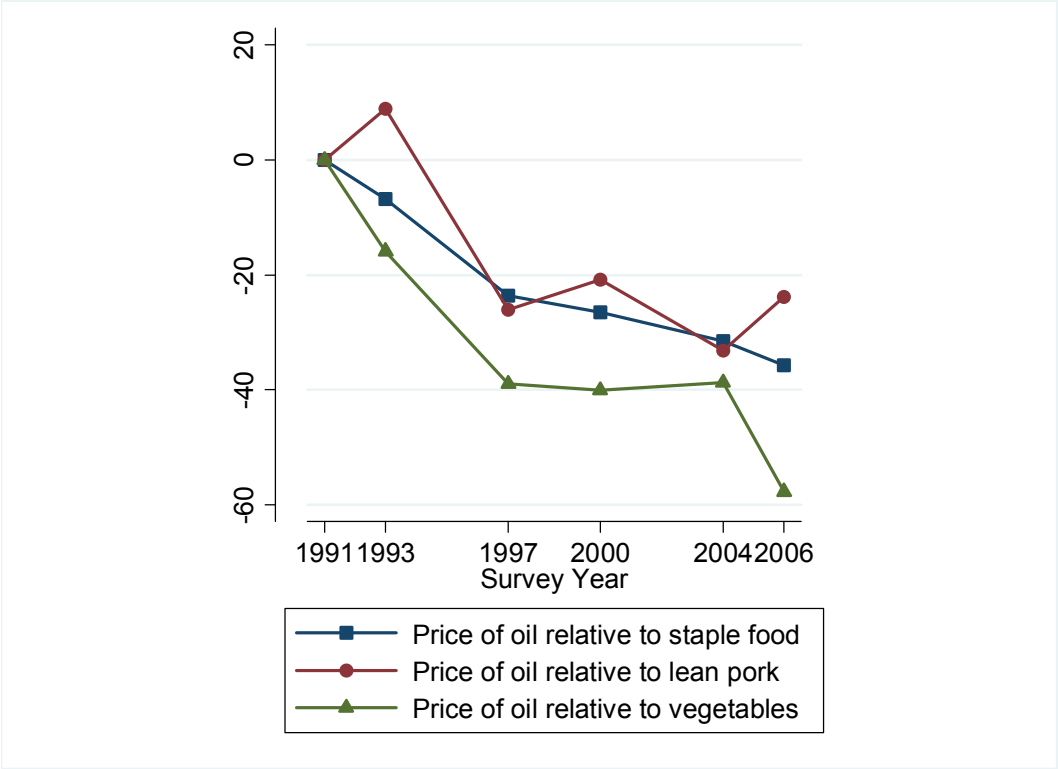

Figure I-2 BMI and TSF Trend

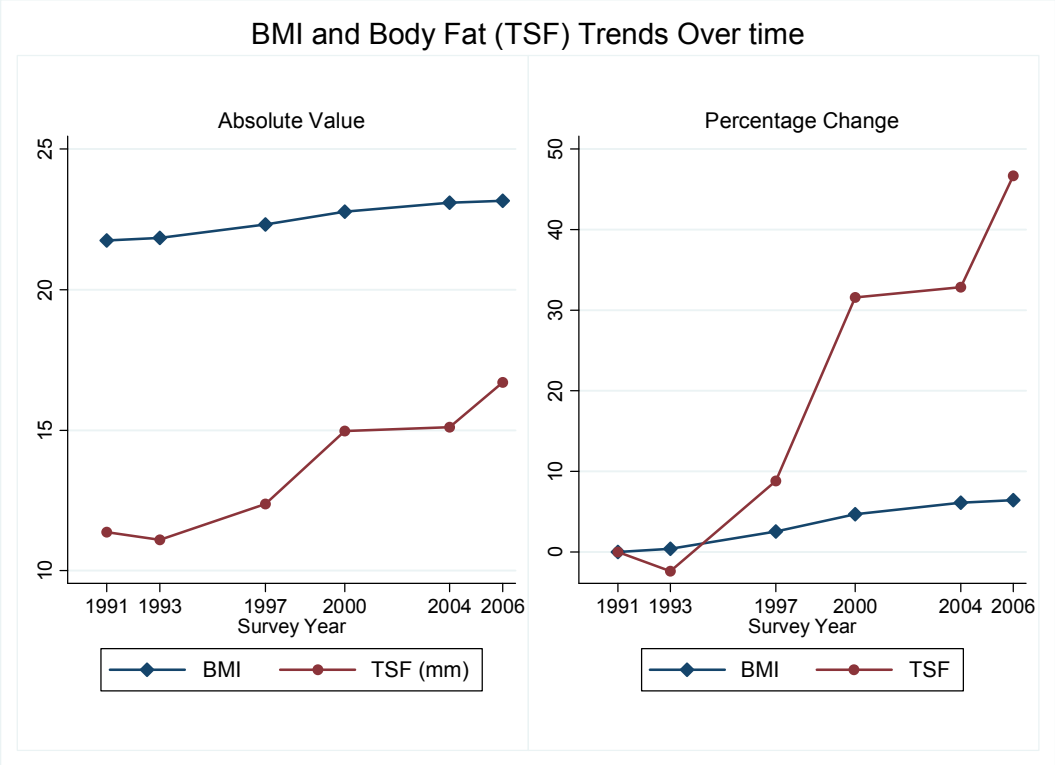

Figure I-3 Percentage changes in BMI and TSF Vs. percentage changes in relative oil prices

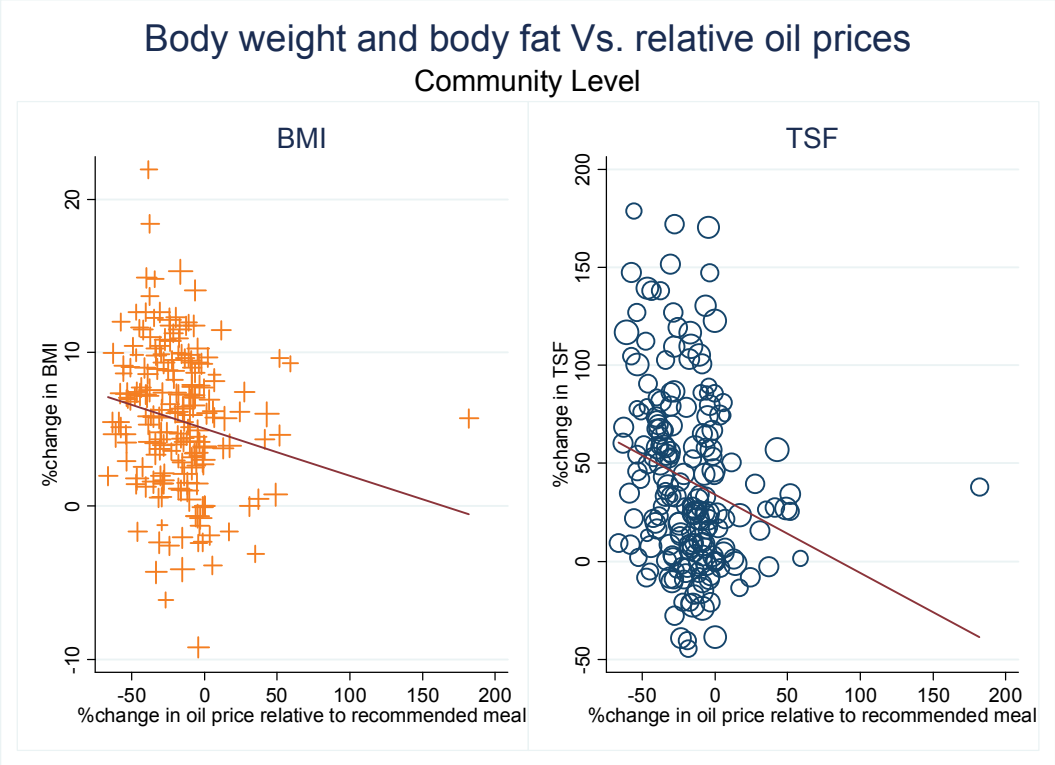

Table I-1 Demographics

	All		Female		Male	
Variable List	N	Mean	N	Mean	N	Mean
HH income	40164	13950	20826	13797	19338	14114
HH income pp	40164	3964	20826	3911	19338	4020
Female	40164	0.52	20826	1	19338	0
Urban	40164	0.34	20826	0.34	19338	0.34
Age	40164	44.06	20826	44.05	19338	44.08
Obesity indicators						
BMI	40164	22.55	20826	22.67	19338	22.42
TSF (mm)	40164	13.86	20826	15.91	19338	11.64
Food Prices						
Pr(staple oil)	40164	3.79	20826	3.77	19338	3.80
Pr(staple food)	40164	1.21	20826	1.21	19338	1.21
Pr(vegetables)	39217	0.54	20344	0.55	18873	0.54
Pr(lean pork)	37816	7.01	19610	6.99	18206	7.02
P(oil)/P(staple food)	40164	3.46	20826	3.46	19338	3.46
P(oil)/P(lean pork)	37816	0.59	19610	0.59	18206	0.59
P(oil)/P(vegetables)	39217	9.52	20344	9.47	18873	9.58
Physical Activity Level						
Light	40164	0.39	20826	0.42	19338	0.35
Moderate	40164	0.17	20826	0.14	19338	0.20
Heavy	40164	0.44	20826	0.43	19338	0.46
Education						
No education	40164	0.25	20826	0.34	19338	0.15
Elementary school	40164	0.22	20826	0.21	19338	0.23
Middle School	40164	0.31	20826	0.27	19338	0.35
HS diploma or equiv	40164	0.18	20826	0.15	19338	0.21
Bachelor +	40164	0.05	20826	0.03	19338	0.06
Fuel Prices (IV)						
Gasoline/L	29310	3.09	15178	3.09	14132	3.09

Table I-2 Effects of staple oil price relative to staple food on BMI and body fat (TSF) – log-log model

	Outcome: Log(BMI)			Outcome: Log(TSF)		
Variables	**OLS**	**RE**	**FE**	**OLS**	**RE**	**FE**
Log(staple oil/staple)	-0.005	0.003	0.003	-0.153***	-0.131***	-0.079***
	[0.005]	[0.002]	[0.002]	[0.029]	[0.026]	[0.028]
Log(lean pork)	-0.004	0.005*	0.006**	-0.003	0.027	0.080***
	[0.007]	[0.003]	[0.003]	[0.029]	[0.026]	[0.029]
Log(vegetables)	-0.025***	-0.010***	-0.004*	0.035*	0.029	-0.012
	[0.004]	[0.002]	[0.002]	[0.020]	[0.019]	[0.020]
Log(BMI)				1.707***	1.617***	1.256***
				[0.043]	[0.042]	[0.065]
Moderate PAL	-0.015***	-0.005***	-0.001	-0.015	-0.011	0.010
	[0.003]	[0.002]	[0.002]	[0.012]	[0.011]	[0.011]
Heavy PAL	-0.042***	-0.015***	-0.005***	-0.090***	-0.071***	0.007
	[0.005]	[0.002]	[0.002]	[0.017]	[0.015]	[0.012]
Age	0.011***	0.010***	0.091***	-0.002	0.001	-0.687***
	[0.000]	[0.000]	[0.011]	[0.001]	[0.001]	[0.167]
Age squared	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	[0.007]	[0.006]	[0.008]	[0.028]	[0.027]	[0.049]
Log(HH income pp)	0.009***	0.003***	0.001	0.027***	0.023***	0.009
	[0.001]	[0.001]	[0.001]	[0.006]	[0.005]	[0.006]
Urban	0.014**	0.015**	dropped	0.036	0.039	dropped
	[0.006]	[0.006]		[0.026]	[0.026]	
Female	0.007**	0.006**	dropped	0.365***	0.359***	dropped
	[0.003]	[0.002]		[0.013]	[0.013]	
Education	included	included	included	included	included	included
Year FE	included	included	included	included	included	included
Constant	2.754***	2.791***	-0.124	-2.955***	-2.794***	23.246***
	[0.021]	[0.012]	[0.395]	[0.155]	[0.147]	[5.937]
Observations	36905	36905	36905	36905	36905	36905
R-squared	0.108	.	0.148	0.401	.	0.177
Number of perid		15472	15472		15472	15472

Robust standard errors in brackets

***** p<0.01, ** p<0.05, * p<0.1**

Table I-3 Effects of staple oil price relative to lean pork on BMI and body fat (TSF) – log-log model

	Outcome: Log(BMI)			Outcome: Log(TSF)		
Variables	**OLS**	**RE**	**FE**	**OLS**	**RE**	**FE**
Log(staple oil/lean pork)	-0.021***	-0.005**	-0.002	-0.164***	-0.141***	-0.083***
	[0.005]	[0.002]	[0.002]	[0.029]	[0.027]	[0.026]
Log(staple food)	-0.024***	-0.007**	-0.001	0.005	0.029	0.075**
	[0.005]	[0.003]	[0.003]	[0.028]	[0.028]	[0.032]
Log(vegetables)	-0.021***	-0.009***	-0.003	0.039*	0.030	-0.012
	[0.003]	[0.002]	[0.002]	[0.020]	[0.019]	[0.020]
Log(BMI)				1.689***	1.605***	1.256***
				[0.044]	[0.042]	[0.065]
Moderate PAL	-0.016***	-0.005***	-0.001	-0.021*	-0.015	0.010
	[0.003]	[0.002]	[0.002]	[0.012]	[0.011]	[0.011]
Heavy PAL	-0.042***	-0.015***	-0.005***	-0.087***	-0.070***	0.007
	[0.004]	[0.002]	[0.002]	[0.017]	[0.015]	[0.012]
Age	0.011***	0.010***	0.094***	-0.002	0.001	-0.687***
	[0.000]	[0.000]	[0.011]	[0.001]	[0.001]	[0.165]
Age squared	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.009***	0.003***	0.001	0.028***	0.023***	0.009
	[0.001]	[0.001]	[0.001]	[0.006]	[0.005]	[0.006]
Urban	0.011**	0.014**	dropped	0.025	0.032	dropped
	[0.005]	[0.006]		[0.026]	[0.026]	
Female	0.007**	0.006**	dropped	0.365***	0.359***	dropped
	[0.003]	[0.002]		[0.013]	[0.013]	
Education	included	included	included	included	included	included
Year FE	included	included	included	included	included	included
Constant	2.731***	2.800***	-0.220	-3.180***	-2.948***	23.244***
	[0.015]	[0.011]	[0.390]	[0.151]	[0.143]	[5.894]
Observations	36905	36905	36905	36905	36905	36905
R-squared	0.114	.	0.147	0.402	.	0.177
Number of perid		15472	15472		15472	15472

Robust standard errors in brackets
***** p<0.01, ** p<0.05, * p<0.1**

Table I-4 Effects of relative oil price to vegetables on BMI and body fat (TSF), log-logmodel

	Outcome: Log(BMI)			Outcome: Log(TSF)		
Variables	OLS	RE	FE	OLS	RE	FE
Log(staple oil/veg)	0.008**	0.007***	0.004**	-0.107***	-0.081***	-0.009
	[0.003]	[0.002]	[0.002]	[0.019]	[0.018]	[0.018]
Log(staple food)	-0.035***	-0.011***	-0.003	-0.010	0.015	0.053
	[0.005]	[0.003]	[0.003]	[0.028]	[0.028]	[0.033]
Log(lean pork)	-0.007	0.004*	0.006**	-0.011	0.017	0.060**
	[0.008]	[0.002]	[0.002]	[0.031]	[0.027]	[0.029]
Log(BMI)				1.710***	1.624***	1.256***
				[0.042]	[0.042]	[0.065]
Moderate PAL	-0.016***	-0.005***	-0.001	-0.018	-0.013	0.010
	[0.003]	[0.002]	[0.002]	[0.012]	[0.011]	[0.011]
Heavy PAL	-0.042***	-0.015***	-0.005***	-0.088***	-0.071***	0.006
	[0.005]	[0.002]	[0.002]	[0.017]	[0.015]	[0.012]
Age	0.011***	0.010***	0.091***	-0.002	0.000	-0.674***
	[0.000]	[0.000]	[0.011]	[0.001]	[0.001]	[0.169]
Age squared	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.009***	0.003***	0.001	0.028***	0.023***	0.010
	[0.001]	[0.001]	[0.001]	[0.006]	[0.005]	[0.006]
Urban	0.011**	0.014**	dropped	0.024	0.031	dropped
	[0.006]	[0.006]		[0.026]	[0.026]	
Female	0.007**	0.005**	dropped	0.364***	0.358***	dropped
	[0.003]	[0.002]		[0.013]	[0.013]	
Education	included	included	included	included	included	included
Year FE	included	included	included	included	included	included
Constant	2.755***	2.790***	-0.122	-2.958***	-2.819***	22.780***
	[0.021]	[0.012]	[0.395]	[0.153]	[0.146]	[6.034]
Observations	36905	36905	36905	36905	36905	36905
R-squared	0.107	.	0.148	0.401	.	0.175
Number of perid		15472	15472		15472	15472

Robust standard errors in brackets
***** p<0.01, ** p<0.05, * p<0.1**

Table I-5 Effects of oil prices on BMI and body fat (TSF), log-log model

	Outcome: Log(BMI)			Outcome: Log(TSF)		
Variables	OLS	RE	FE	OLS	RE	FE
Log(staple oil)	-0.033***	-0.003	0.004	-0.274***	-0.222***	-0.085**
	[0.007]	[0.003]	[0.003]	[0.041]	[0.036]	[0.035]
Log(staple food)	-0.018***	-0.008***	-0.003	0.055*	0.065**	0.075**
	[0.005]	[0.003]	[0.003]	[0.030]	[0.030]	[0.034]
Log(lean pork)	0.009	0.007***	0.006**	0.052*	0.065**	0.082***
	[0.006]	[0.003]	[0.003]	[0.030]	[0.029]	[0.031]
Log(vegetables)	-0.018***	-0.010***	-0.004*	0.063***	0.046**	-0.012
	[0.003]	[0.002]	[0.002]	[0.020]	[0.019]	[0.020]
Log(BMI)				1.675***	1.601***	1.256***
				[0.041]	[0.041]	[0.065]
Moderate PAL	-0.015***	-0.005***	-0.001	-0.016	-0.012	0.010
	[0.003]	[0.002]	[0.002]	[0.012]	[0.011]	[0.011]
Heavy PAL	-0.042***	-0.015***	-0.005***	-0.088***	-0.071***	0.007
	[0.004]	[0.002]	[0.002]	[0.017]	[0.014]	[0.012]
Age	0.011***	0.010***	0.091***	-0.002	0.000	-0.686***
	[0.000]	[0.000]	[0.011]	[0.001]	[0.001]	[0.167]
Age squared	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.009***	0.003***	0.001	0.027***	0.023***	0.009
	[0.001]	[0.001]	[0.001]	[0.006]	[0.005]	[0.006]
Urban	0.010**	0.014**	dropped	0.021	0.030	dropped
	[0.005]	[0.006]		[0.025]	[0.025]	
Female	0.007**	0.006**	dropped	0.366***	0.359***	dropped
	[0.003]	[0.002]		[0.013]	[0.013]	
Education	included	included	included	included	included	included
Year FE	included	included	included	included	included	included
Constant	2.767***	2.793***	-0.122	-2.816***	-2.711***	23.216***
	[0.021]	[0.012]	[0.395]	[0.156]	[0.146]	[5.953]
Observations	36905	36905	36905	36905	36905	36905
R-squared	0.115	.	0.148	0.408	.	0.177
Number of perid		15472	15472		15472	15472

Robust standard errors in brackets

***** p<0.01, ** p<0.05, * p<0.1**

Table I-6 Effects of oil prices relative to recommended protein and carbohydrate intake on BMI and body fat (TSF), log-log model

	Outcome: Log(BMI)			Outcome: Log(TSF)		
Variables	**OLS**	**RE**	**FE**	**OLS**	**RE**	**FE**
Log(staple oil/rec'd meal)	-0.021***	-0.002	0.001	-0.227***	-0.197***	-0.119***
	[0.006]	[0.003]	[0.002]	[0.031]	[0.028]	[0.028]
Log(BMI)				1.672***	1.588***	1.249***
				[0.044]	[0.042]	[0.064]
Moderate PAL	-0.015***	-0.005***	-0.001	-0.017	-0.013	0.011
	[0.003]	[0.002]	[0.002]	[0.012]	[0.011]	[0.011]
Heavy PAL	-0.041***	-0.015***	-0.005***	-0.089***	-0.070***	0.011
	[0.005]	[0.002]	[0.002]	[0.017]	[0.015]	[0.012]
Age	0.011***	0.010***	0.100***	-0.002	0.000	-0.673***
	[0.000]	[0.000]	[0.011]	[0.001]	[0.001]	[0.164]
Age squared	-0.000***	-0.000***	-0.000***	-0.000	-0.000**	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.007***	0.002***	0.000	0.028***	0.024***	0.009
	[0.002]	[0.001]	[0.001]	[0.006]	[0.005]	[0.006]
Urban	0.006	0.013**	0.000	0.039	0.045*	0.000
	[0.006]	[0.006]	[0.000]	[0.026]	[0.026]	[0.000]
Female	0.007**	0.005**	0.000	0.364***	0.357***	0.000
	[0.003]	[0.002]	[0.000]	[0.013]	[0.013]	[0.000]
Education	Included	included	included	included	included	included
Year FE	Included	included	included	included	included	included
Constant	2.805***	2.819***	-0.427	-2.998***	-2.787***	22.875***
	[0.016]	[0.011]	[0.389]	[0.147]	[0.140]	[5.850]
Observations	37816	37816	37816	37816	37816	37816
R-squared	0.103	.	0.146	0.404	.	0.175
Number of perid		15649	15649		15649	15649

Robust standard errors in brackets
***** p<0.01, ** p<0.05, * p<0.1**

Table I-7 Effects of relative oil price on BMI and body fat (TSF) by gender, log-log model

	Log(BMI)	Log(TSF)
Female		
Log oil price relative to staple food	0.003	-0.072***
	[0.003]	[0.026]
Log oil price relative to lean pork	-0.004	-0.083***
	[0.003]	[0.030]
Log oil price relative to vegetables	0.003	-0.018
	[0.002]	[0.016]
Log oil price	0.002	-0.098***
	[0.004]	[0.034]
Log oil price relative to rec'd meal	-0.000	-0.120***
	[0.003]	[0.030]
Male		
Log oil price relative to staple food	0.003	-0.086**
	[0.003]	[0.036]
Log oil price relative to lean pork	0.001	-0.085***
	[0.002]	[0.033]
Log oil price relative to vegetables	0.005**	0.000
	[0.002]	[0.024]
Log oil price	0.006*	-0.073
	[0.003]	[0.046]
Log oil price relative to rec'd meal	0.003	-0.118***
	[0.003]	[0.035]
***** p<0.01, ** p<0.05, * p<0.1**		

Table I-8 Effects of instrumented food prices on BMI, log-log FE-IV model

	Outcome: Log(BMI)				
Variables					
Log(staple oil/staple food)	-0.039***				
	[0.012]				
Log(staple oil/lean pork)		-0.098***			
		[0.033]			
Log(staple oil/vegetables)			-0.053***		
			[0.017]		
Log(staple oil)				-0.066***	
				[0.020]	
Log(staple oil/rec'd meal)					-0.060***
					[0.019]
Log(staple food)		0.009**	0.000	0.012**	
		[0.004]	[0.002]	[0.005]	
Log(lean pork)	0.011***		0.007***	0.023***	
	[0.002]		[0.002]	[0.005]	
Log(vegetables)	-0.005***	-0.005***		-0.001	
	[0.001]	[0.001]		[0.002]	
Moderate PAL	0.001	-0.001	0.001	0.000	-0.001
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Heavy PAL	-0.004**	-0.002	-0.004**	-0.004**	-0.004**
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Age	0.106***	0.096***	0.155***	0.105***	0.104***
	[0.009]	[0.009]	[0.021]	[0.009]	[0.008]
Age squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.000	-0.000	0.000	0.000	-0.000
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Education	included	included	included	included	included
Year FE	included	included	included	included	included
Observations	25335	25335	25335	25335	26273
Number of perid	8626	8626	8626	8626	8842
WeakID test	347.6	54.93	93.98	224.5	185.2
UnID test	340.9	54.81	93.54	221.8	183.4
Standard errors in brackets					
*** p<0.01, ** p<0.05, * p<0.1					

Table I-9 Effects of instrumented food prices on TSF (body fat), log-log FE-IV model

	Outcome: Log(TSF)				
Variables					
Log(staple oil/staple food)	-0.214***				
	[0.070]				
Log(staple oil/lean pork)		-0.383**			
		[0.167]			
Log(staple oil/vegetables)			-0.187**		
			[0.076]		
Log(staple oil)				-0.277***	
				[0.106]	
Log(staple oil/rec'd meal)					-0.298***
					[0.106]
Log(staple food)		0.128***	0.072***	0.137***	
		[0.025]	[0.013]	[0.029]	
Log(lean pork)	0.094***		0.055***	0.124***	
	[0.014]		[0.013]	[0.027]	
Log(vegetables)	-0.024***	-0.028***		-0.013	
	[0.008]	[0.009]		[0.009]	
Log(BMI)	1.263***	1.240***	1.289***	1.261***	1.246***
	[0.048]	[0.050]	[0.051]	[0.048]	[0.048]
Moderate PAL	0.010	0.004	0.009	0.009	0.007
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Heavy PAL	0.016	0.025*	0.012	0.017	0.020*
	[0.012]	[0.013]	[0.012]	[0.012]	[0.011]
Age	-0.750***	-0.792***	-0.576***	-0.754***	-0.777***
	[0.055]	[0.052]	[0.098]	[0.054]	[0.050]
Age squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log(HH income pp)	0.010**	0.009**	0.011***	0.010**	0.011***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Education	included	included	included	included	included
Year FE	included	included	included	included	included
Observations	24245	24245	24245	24245	25083
Number of individuals	8327	8327	8327	8327	8538
WeakID test	371.3	68.27	148.1	273.7	212.2
UnID test	363.2	68.05	146.9	269.4	209.7
Standard errors in brackets					
*** p<0.01, ** p<0.05, * p<0.1					

Chapter II. Health Implications of Obesity

Based on the findings of the first essay, this chapter further explores health implications of obesity measured by TSF. I limit the sample to non-obese individuals with a BMI less than 28, and focus on health outcomes such as hypertension, diabetes, and overall health pertaining to everyday activities. The multivariate logistic regression results show that, TSF, as a proxy for body fat, is an independent and significant risk factor for health, among normal-weight individuals. When factors such as age, sex, income, education, residence type, and year fixed effects are adjusted for, a 1 percent increase in TSF at the mean will lead to a 0.06 percentage points increase in hypertension, a 0.003 percentage points increase in diabetes, and a 0.008 percentage points increase in the likelihood of being too sick to work, respectively. If TSF were to continue growing at the current rate, this single factor would lead 16.3 million more non-obese individuals to develop hypertension in 2021, according to the model results. The findings suggest that individuals with a normal body weight can be at risk for various obesity-related diseases if their body fat is high. Policy makers need to raise awareness among these individuals for better disease prevention and monitoring.

1. Background

The first chapter shows that relative food prices can lead to high levels of body fat, but it does not explore how high body fat affects health among people who maintain a normal weight.

It is known that obesity is a risk factor for hypertension, diabetes, and other cardiovascular diseases. The link between obesity and essential hypertension has long been confirmed by experimental, clinical, and observational studies (Hall 2003). Most of studies identify the risk factor with excess weight gain when they associate obesity with hypertension (Hall 2000). Garrison et al. also showed that adiposity is one of the strongest contributor to hypertension, by analyzing evidence from the Framingham Offspring Study (Garrison et al. 1987). Likewise, the link between obesity and diabetes has also been widely recognized (Mokdad et al. 2003; Hossain et al. 2007). However, little research has been done on the effects of varying body fat levels on health among normal-weight individuals.

This chapter focuses on health implications of obesity on three outcomes: hypertension, diabetes, and illness that disrupted work, using the same dataset that includes the same individuals (although the sample sizes vary depending on which health outcome is being analyzed). The main question addressed is, whether percentage body fat plays an independent role of affecting their likelihood of developing certain obesity-induced diseases, for individuals who are not considered obese by the standard of BMI cutoffs.

2. Data

This paper employs the same dataset, China Health and Nutrition Survey 1991-2006, for the health implication analyses. There are three outcomes of interest, which are all dummy variables. The first outcome is hypertension. An individual is considered to have hypertension if (1) the systolic pressure is greater than or equal to 140 mmHg or the diastolic pressure is greater than or equal to 90 mmHg, or (2) in the questionnaire one responds yes to the question "Has a doctor ever told you that you suffer from high blood pressure?" Because hypertension is a chronic health condition and is often not reversible, I code the variable as accumulative. In other words, if an individual is coded as having hypertension at one point, he will also be coded as having hypertension in later waves.

The second outcome is diabetes. It is available in interview waves 1997, 2000, 2004, and 2006, where the questionnaire asks, "Has a doctor ever told you that you suffer from diabetes?" If the individual answers "yes" to that question, he is coded as having diabetes. This independent variable is also treated as cumulative, because like hypertension, diabetes is a chronic health condition that is mostly likely not reversible.

The last health outcome is a measure of an individual's overall health status. The questionnaire asks every participating adult, "During the past 3 months have you had any difficulty carrying out your daily activities and work or studies due to illness?" For those who answered "yes", I assign them to the category of "too sick to work" in the past 3 months. This health measure is also only available in survey years 1997, 2000, 2004, and 2006, like diabetes. It is not coded as cumulative, because it is unclear what specific illness one had that interrupted work and daily activities, and whether it was a chronic condition.

Table II-1 shows the fraction of individuals who have hypertension, diabetes, or have been too sick to work in the past three months over time. There are about 15% interviewees with hypertension in 1991, and the fraction rises to almost 38% in 2006. The prevalence of diabetes in the sample is much lower than hypertension, yet the rate of growth is comparable. The fraction of individuals diagnosed with diabetes is 1.45% in 1997 and it increases to 2.96% in 2006. In contrast, the percentage of interviewees reported to be "too sick to work" in the past 3 months remain relatively unchanged over time. About 4.67% adults report that they have had difficulty to work or study or carry out daily activities due to illness in the past 3 months in 1997. The fraction is 6.25% in 2006.

The key independent variable is TSF (triceps skinfold thickness), which is a direct measure for subcutaneous body fat. The covariates include age, age quadratic, sex, type of residence, education attainment, per person household income, and year dummies.

In order to explore detrimental health effects of high body fat, I limit the sample to non-obese individuals who have a BMI that is at least 20 but no higher than 28. The main reason for limiting to such a range of BMI is because those individual are usually considered to have a normal weight, and when they are not labeled as "obese" by the standard of BMI, they might be more likely to ignore obesity-related health problems even if their percent body fat is high enough to put them at risk. For the same reason, policy makers might also be less likely to concern this subgroup of people with adverse health effects related to obesity. However, Chapter I establishes that individuals who maintain a normal weight can also have a high body fat at the same time. It would be policy relevant to see whether a high percent body fat is a risk factor for various health outcomes among individuals without a high BMI.

Figure II-1, II-2, and II-3 depict the unadjusted relationship between higher body fat and the three health outcomes, respectively. In each of the tables, I compare the prevalence of disease between individuals with a TSF in the top quartile (TSF ≥ 20mm) and the rest (TSF < 20mm). The x-axis is BMI, and I limit the range to 20 < BMI ≤ 28. The sample does not included non-obese individuals who have a BMI higher than 28 because there is an insufficient number of observations with a BMI between 28 and 30, and that leads to unreliable (spiky) point estimates of disease prevalence.

Figure II-1 shows that among individuals with a BMI greater than 20 but no greater than 28, having a TSF higher than 20mm does not elevate the likelihood of developing hypertension. According to the chart, except for individuals with a BMI between 20 and 21, individuals with a lower TSF are associated with a slightly higher fraction of hypertension. For both groups, the likelihood of having hypertension is positively associated with BMI. The rate of increase is gradual and smooth, and there is no noticeable spike near the traditional overweight cutoff at BMI = 25. However, the relationship shown is not adjusted by age, salt intake and other factors that influence blood pressure.

Figure II-2 shows the relationship of TSF and diabetes. In general, individuals with a TSF in the top quartile (≥ 20mm) are more likely to have diabetes than individuals with a TSF < 20mm. For individuals with a BMI between 21 and 24, the fraction is not consistently higher, mainly due to ups and downs of the curve for individuals with a TSF in the top quartile. But for individuals with a BMI ≥ 24, the difference in prevalence is constant and invariable.

Figure II-3 shows the fraction of individuals who are too sick to work, to go to school, or to carry out daily activities in the past 3 months, by TSF. It appears that (1) illness has a non-linear relationship with BMI. For individuals with a high TSF, the peak of illness takes place when BMI is between 24 and 25, and the prevalence decreases towards the two tails and rises again when BMI is

becoming very low (close to 20) or very high (close to 28). For individuals with a lower TSF, the prevalence of illness is relatively flat across BMI, but is minimized when BMI is between 23 and 25. Individuals with a TSF ≥ 20mm are more likely to become too sick to work than their counterparts when BMI is greater than 23.

To summarize, the three figures provide mixed evidence that high TSF leads to high prevalence of diseases, when BMI is restricted to a normal range. However, it is clear that the correlation between high subcutaneous fat and worse health outcomes holds true for some diseases within certain BMI ranges.

In the following section I will describe results from the regression analysis, when other relevant factors are being controlled for.

3. Regression Analysis and Results

Regressions are run on three health outcomes separately. For each outcome, there are two regression models, one with TSF in levels as the key regressor, and one with log-transformed TSF modeled instead. Results are shown in Table II-2, II-3, and II-4.

Table II-2 displays two regression tables on hypertension. The analysis is limited to a subsample of non-obese individuals with a BMI greater than 20 but no higher than 28. Model 1 shows the odds ratios of coefficients where TSF is modeled in levels. Model 2 shows the odds ratios where TSF is log transformed. From the table, one can see that TSF is significantly and positively related with the likelihood of developing hypertension. And among all covariates, age, and being an urban resident are positively correlated with having hypertension ($p<0.05$), and being female decreases the chance

of having hypertension ($p<0.01$), holding all else equal. It is interesting that income does not seem to have any effects on hypertension, according to the table, suggesting hypertension and income are unrelated.

In terms of marginal effects (not shown), at the mean value of TSF (2.523 in logs, and 12.53 in levels), a 1 percent increase in TSF is associated with a 0.06 percentage points increase in hypertension, holding all else equal. The likelihood of having hypertension would then change from the mean value of 22.77% to 22.83%, which is roughly an increase of 0.26%.

Table II-3 presents regression results on diabetes. Like the analysis of hypertension, the analysis is also limited to a subsample of non-obese individuals with a BMI greater than 20 but no higher than 28. Similarly, Model 1 shows the odds ratios of coefficients with TSF being modeled in levels, and Model 2 shows the odds ratios with TSF in logs. In both models, TSF is positively correlated with having diabetes ($p<0.01$), holding all else equal. Among other covariates, getter older, being a female, and living in an urban area are positively correlated with having diabetes. Education-wise, it seems that each additional educational degree further increases the likelihood of having diabetes, as compared to no education. However, the difference between having no education and having completed elementary schooling is not statistically significant.

According to the marginal effects table (not shown), at the mean value of 2.60 in logs (or 13.5mm in levels), a 1 percent increase in TSF will lead to a 0.003 percentage points increase in the likelihood of having diabetes, holding all else equal. In other words, if there is a 1 percent increase in TSF from 13.5mm, the chance of having diabetes will increase from 1.427% to 1.430%, which is roughly a 0.24% increase in diabetes.

Table II-4 shows the two regressions on the outcome of being too sick to work in the past 3 months. I limit the analysis to individuals with a BMI range from 22 to 28, which is slightly higher than the range of 20 to 28, because general illness might or might not be associated with obesity, and I intend to focus on obesity-related illness. Without knowing exactly what disease an individual has that disrupts work or daily activities, I can only exclude individuals with a relatively light weight, because they are more likely to fall sick due to reasons other than high body fat. Some might even suffer from being underweight or from malnutrition, in which case more body fat might play a positive role in one's health and thus could lead to confounding results.

Like the other two outcomes, Model 1 uses TSF in levels as the key regressor, and Model 2 uses TSF in logs the key regressor. TSF is shown to be positively correlated with illness in both models. The odds ratio is 1.01 in the levels model, and is 1.18 in the logs model. Age does not seem to be a risk factor for being too sick to work. However, being a female as opposed to male, or living in an urban area as opposed to rural has a positive relationship with illness, holding all else equal. In contrast, education attainment appears to reduce the likelihood of becoming too sick to work. On the other hand, household income per person does not have, if any, effects on illness.

In terms of marginal effects (not shown), a 1 percent increase in TSF from the mean value of 14.88mm (or 2.70 in logs) leads to a 0.008 percentage points increase in the likelihood of being too sick to work. The chance of falling sick rises from 5.14% to roughly 5.15%, which is an increase of about 0.15%.

4. Discussion

All three model specifications suggest that individuals with a normal weight can be at greater risks for adverse health outcomes when their body fat increases. In the results section, I present the regression coefficients and briefly interpret the marginal treatment effects on health outcomes by changing TSF by one percent. However, marginal effects can not accurately predict how probabilities of those health outcomes when TSF changes by a considerable magnitude. If we ask, "How many more people would develop hypertension in 15 years, starting from 2006, if the TSF were to continue growing at the current rate and to increase by another 50%?", we can instead substitute new values of TSF in the fitted model after running a logistic regression.

Take Model 2 as an example where TSF is constructed as log transformed. In the estimated sample, log TSF has the mean value of 2.37 in 1991, and 2.74 in 2006. In levels, the mean values are 10.70mm in 1991 and 15.43mm in 2006. If we assume that TSF continues to grow at the same rate, then TSF will be 15.43mm × (15.43 ÷ 10.70) = 22.25mm in 2021, which is 3.10 in logs. Based on the estimation in Model 2, the predicted probability of hypertension is 28.77%, with TSF at the mean value in 2006, the year being 2006, and other control variables at their mean value of all years. And the predicted probability of hypertension increases to 31.26%, if TSF rises to 3.10 in logs (or 22.25mm in levels), and if the other variables stay constant. This is then a 31.26 – 28.77 ≈ 2.48 percent points increase in the prevalence of hypertension. To extrapolate this increase to the whole population of China, we would have 1.3 billion × 0.75 (fraction of adults 18 yrs old+) × 0.67 (fraction of adults with $20 \leq BMI < 28$) × 2.48% ≈ 16.3 million more people with hypertension in 2021, among non-obese adult individuals.

5. Policy Implications

This paper looks at whether and how levels of body fat adversely affect individuals' health outcomes, even when they are not considered "obese" by the traditional definition of having a BMI equal or greater than 30. Using three separate health outcomes, hypertension, diabetes, and illness, the analyses show that more body fat is positively associated with worse individual health in all three measures, when the range of BMI is limited to between 20 and 28 (or between 22 and 28). This means that individuals without a high body weight can be at risk for various obesity-related diseases when their body fat is high. However, this group of individuals that is traditionally considered non-obese might not be as alert for such health risks as desired because it is not the target group for obesity prevention and treatment at the policy level, and they indeed are maintaining a normal weight. If they have a high adiposity problem, the problem is rather silent, or invisible.

Thus, the main policy implication is to raise awareness among individuals with a normal weight whose body fat is climbing over time. Although the benefits of awareness are hard to quantify, there is empirical evidence that shows effectiveness of awarness. In fact, individuals that are labeled as "obese" might actually experience a lower mortality rate than their counterparts. For example, Frijters and Baron (2009) found that obese individuals are more likely to survive a stroke or lung disease. They concluded that the paradox is possibly due to the fact they are identified as obese. Thus their health status is more closely monitored, they go to the doctor more often, and they receive more attentive health care. It is possible that by applying the same level of attentiveness to normal-weight individuals with high adiposity, they will be more successful in controlling their percentage fat body as well as maintaining a normal body weight. And a large number of obesity-related diseases or even death can be prevented, or at least mitigated.

Figure II-1 Percent hypertension by TSF among non-obese individuals with a normal weight

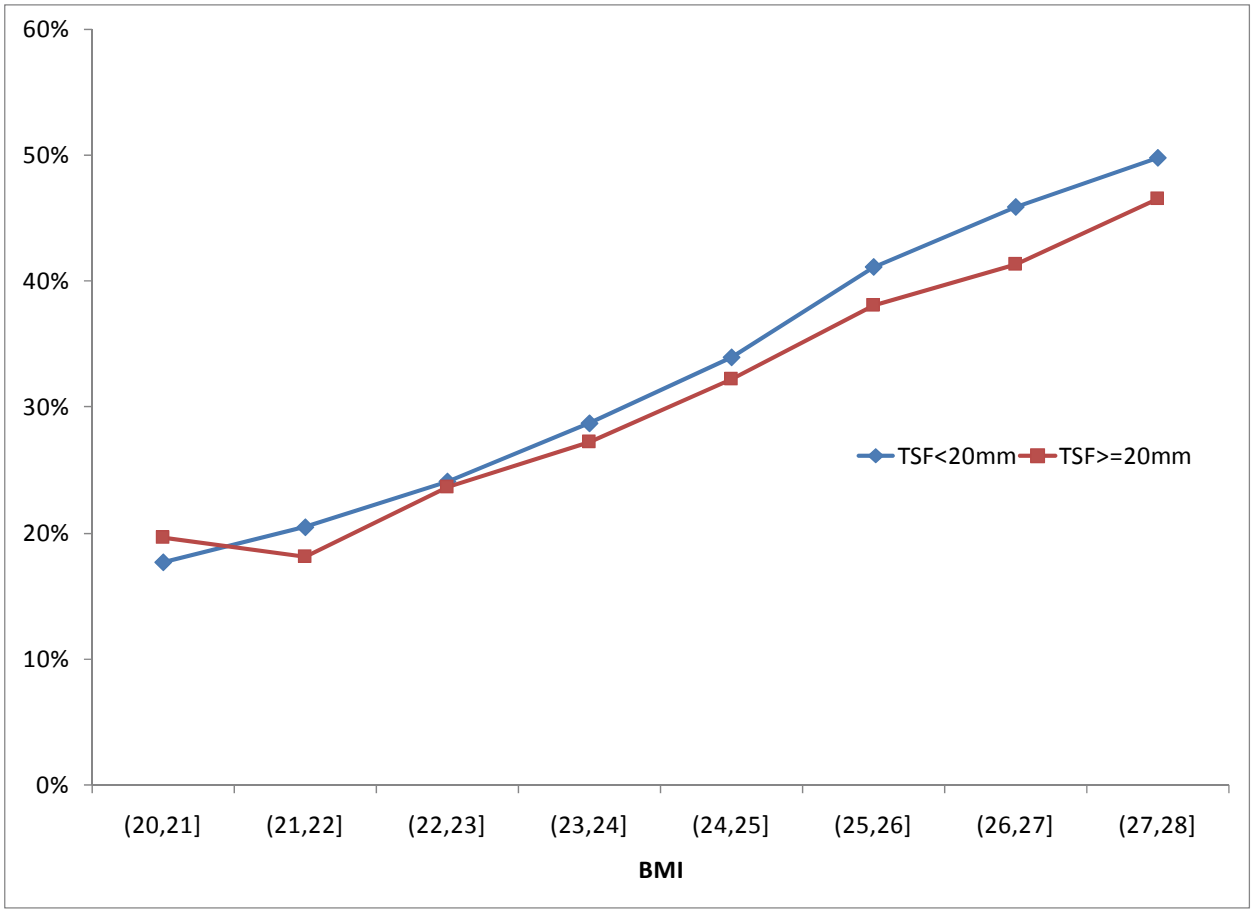

Figure II-2 Percent diabetes by TSF among non-obese individuals with a normal weight

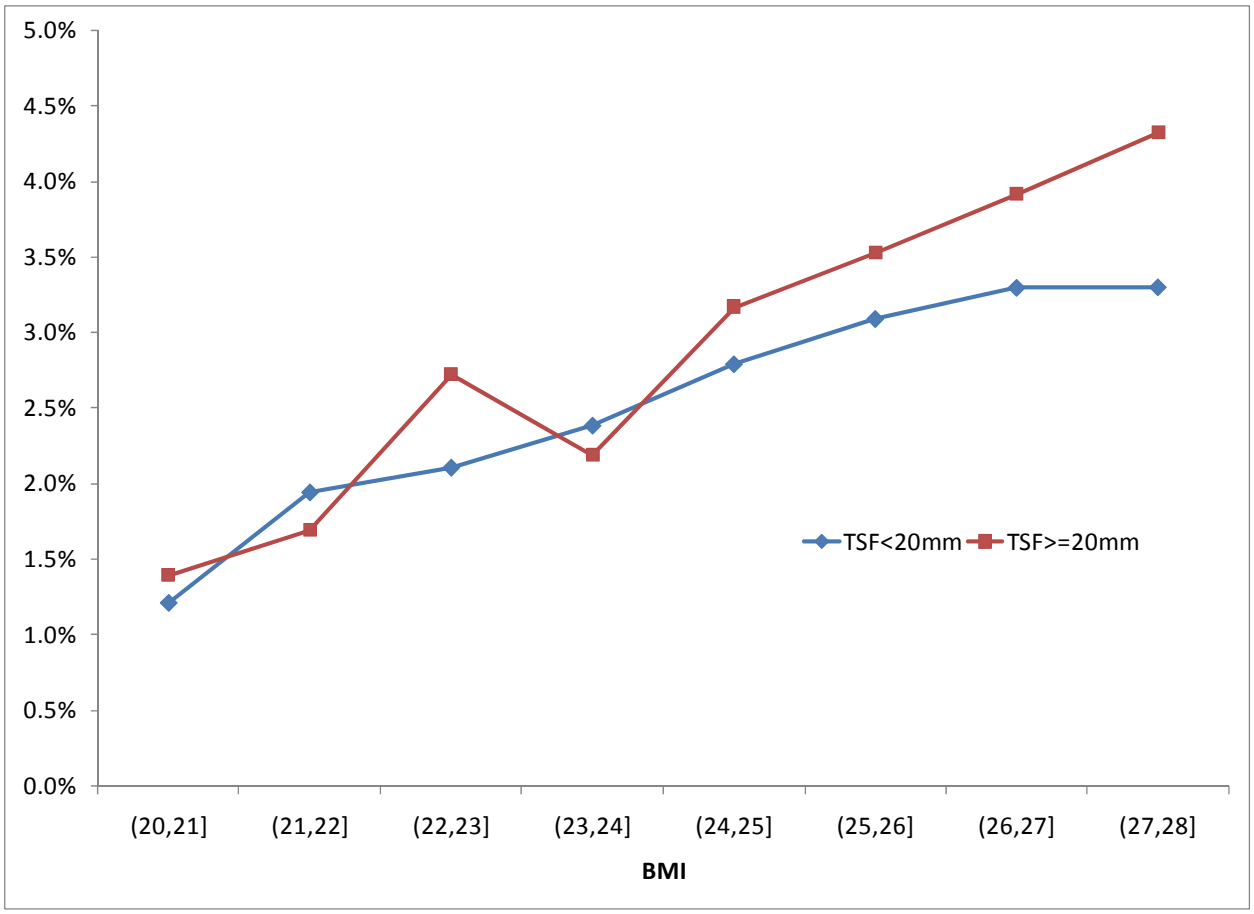

Figure II-3 Percent "too sick to work" by TSF among non-obese individuals with a normal weight

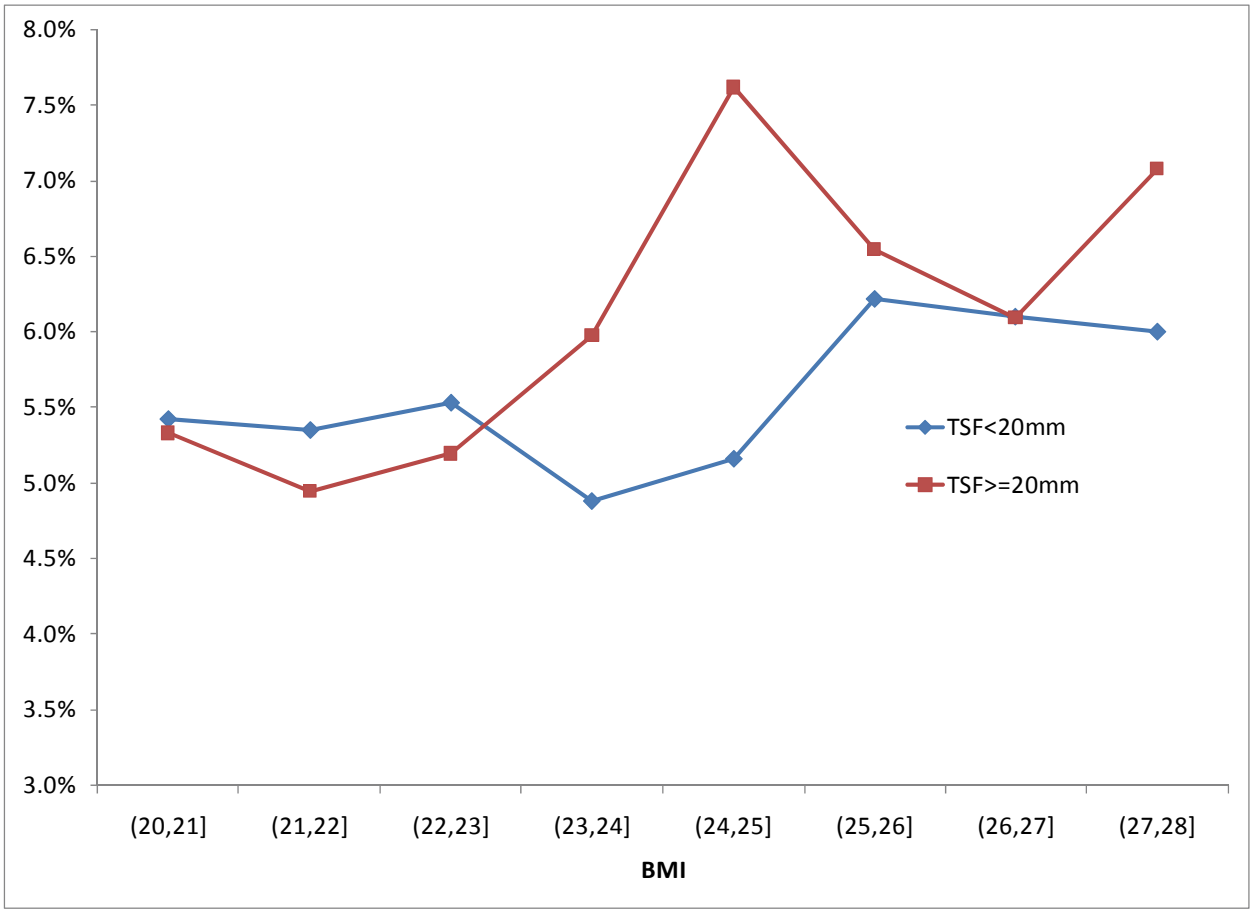

Table II-1 Time trend of health outcomes

	Hypertension		Diabetes		"Too sick to work"	
Survey Year	%	N	%	N	%	N
1991	15.06	5,193	n/a	n/a	n/a	n/a
1993	22.11	4,740	n/a	n/a	n/a	n/a
1997	25.01	5,485	1.45	5,366	4.67	3,295
2000	29.31	6,244	2.24	5,933	5.31	4,048
2004	33.82	5,981	2.54	5,980	6.49	4,098
2006	37.9	6,272	2.96	6,246	6.25	4,339

Table II-2 Body fat (TSF) as a risk factor for hypertension among non-obese individuals

Variables	Outcome: Hypertension	
	Model 1	Model 2
Triceps Skinfold Thickness (TSF)	1.020***	1.381***
	[0.002]	[0.039]
Age	1.081***	1.081***
	[0.008]	[0.008]
Age squared	1.000	1.000
	[0.000]	[0.000]
Female	0.574***	0.557***
	[0.017]	[0.017]
Urban resident	1.071**	1.064**
	[0.032]	[0.032]
Elementary school	1.138***	1.139***
	[0.046]	[0.046]
Middle School	1.091**	1.086*
	[0.046]	[0.046]
HS diploma or equiv	0.949	0.941
	[0.046]	[0.045]
Bachelor +	0.998	0.981
	[0.075]	[0.074]
Household income per person	1.000	1.000
	[0.000]	[0.000]
Wave 1993	1.642***	1.646***
	[0.094]	[0.095]
Wave 1997	1.772***	1.764***
	[0.098]	[0.097]
Wave 2000	1.968***	1.948***
	[0.105]	[0.104]
Wave 2004	2.031***	2.005***
	[0.110]	[0.109]
Wave 2006	2.275***	2.232***
	[0.124]	[0.122]
Constant	0.005***	0.003***
	[0.001]	[0.001]
Observations	33915	33915
Standard errors in brackets		
***** p<0.01, ** p<0.05, * p<0.1**		

Table II-3 Body fat (TSF) as a risk factor for diabetes among non-obese individuals

Variables	Outcome: Diabetes	
	Model 1	Model 2
Triceps Skinfold Thickness (TSF)	1.016***	1.281***
	[0.006]	[0.122]
Age	1.197***	1.197***
	[0.038]	[0.038]
Age squared	0.999***	0.999***
	[0.000]	[0.000]
Female	1.227**	1.215**
	[0.118]	[0.118]
Urban resident	1.390***	1.385***
	[0.130]	[0.130]
Elementary school	1.032	1.032
	[0.140]	[0.140]
Middle School	1.535***	1.535***
	[0.208]	[0.208]
HS diploma or equiv	1.659***	1.655***
	[0.242]	[0.242]
Bachelor +	2.405***	2.384***
	[0.469]	[0.465]
Household income per person	1.000	1.000
	[0.000]	[0.000]
Wave 2000	1.322*	1.319*
	[0.193]	[0.193]
Wave 2004	1.247	1.241
	[0.181]	[0.180]
Wave 2006	1.331**	1.320*
	[0.191]	[0.190]
Constant	0.000***	0.000***
	[0.000]	[0.000]
Observations	23525	23525
Standard errors in brackets		
***** p<0.01, ** p<0.05, * p<0.1**		

Table II-4 Body fat (TSF) as a risk factor for illness among overweight individuals

Variables	Outcome: Too sick to work in the last 3 months	
	Model 1	Model 2
Triceps Skinfold Thickness (TSF)	1.011**	1.178**
	[0.005]	[0.092]
Age	0.992	0.992
	[0.018]	[0.018]
Age squared	1.000*	1.000*
	[0.000]	[0.000]
Female	1.216**	1.215**
	[0.093]	[0.094]
Urban resident	1.279***	1.278***
	[0.095]	[0.095]
Elementary school	0.882	0.883
	[0.086]	[0.086]
Middle School	0.701***	0.702***
	[0.075]	[0.075]
HS diploma or equiv	0.629***	0.629***
	[0.077]	[0.077]
Bachelor +	0.615**	0.614**
	[0.122]	[0.122]
Household income per person	1.000*	1.000*
	[0.000]	[0.000]
Wave 1997	0.762**	0.987
	[0.083]	[0.151]
Wave 2000	0.869	1.315*
	[0.084]	[0.190]
Wave 2004	1.052	1.243
	[0.095]	[0.183]
Constant	0.036***	0.028***
	[0.017]	[0.014]
Observations	15780	15780
Standard errors in brackets		
***** p<0.01, ** p<0.05, * p<0.1**		

Chapter III. Sexually transmitted diseases among users of erectile dysfunction drugs[11]

Pharmacologic treatments for erectile dysfunction (ED) have gained widespread popularity among middle-aged and older males in recent years. Increased sexual activity among users of these treatments would raise concerns about sexually transmitted diseases (STDs) and help identify a group of individuals who may benefit from discussions about safe sexual practices. To examine the association between STDs and ED drugs, we conducted a retrospective study from 1997 to 2006 linking medical and drug claims data of 1,339,382 male employees above the age of 40 with private insurance from 49 large companies. We utilized differential plan coverage of ED drugs and clinical indicators of erectile dysfunction to control for selection into ED drug use. We found that among men above the age of 40, 3.44 percent of ED drug users had a STD compared to 2.19 percent among non-users ($p < 0.001$). After controlling for demographic and other co-morbid conditions, use of an ED drug was associated with an increased likelihood of sexually transmitted disease in the current year (OR = 1.52, $p < 0.001$) and subsequent year (OR = 1.44, $p < 0.001$). Individuals with a disease predisposing to ED were more likely to have a STD when their employer covered ED treatments (246 v. 233 per 10,000 beneficiaries, $p < 0.001$) but were no more likely to have a STD when their employer did not cover these treatments (180 v. 179 per 10,000 beneficiaries, $p < 0.001$). The increased availability of ED treatments through employer coverage was estimated to directly increase the prevalence of STDs (OR = 1.12, $p < 0.05$). We concluded that users of erectile

[11] This paper is co-authored with Anupam B. Jena, Dana P. Goldman, Amee Kamdar, and Darius N. Lakdawalla.

dysfunction drugs have higher rates of STDs. Men using ED drugs should be monitored more closely for STDs, and discussions about safe sexual practices are recommended. Furthermore, wider availability of ED drugs may increase STD rates.

1. Motivation

Pharmacologic treatments for erectile dysfunction (ED) have gained widespread popularity among middle-aged and older males in recent years. Driven largely by the high prevalence of erectile difficulties in this population (Laumann et al. 1999; Johannes et al. 2000; Bacon et al. 2003; Araujo et al. 2004; Enzlin et al. 2004; Nicolosi et al. 2004; Laumann et al. 2005; Lindau et al. 2007), rates of sildenafil use reportedly reached 1.4% in the commercially insured population by 2002 (Delate et al. 2004). This is perhaps not surprising, since 37% of men aged 57 to 85 have some degree of erectile dysfunction (Lindau *et al.* 2007). While their clinical efficacy has been well documented, little is known about the relationship between ED treatments and the prevalence of sexually transmitted diseases (STDs). In light of growing evidence for rising STD cases, including AIDS cases, at older ages (Ory 1998; Karlovsky et al. 2004; Bodley-Tickell et al. 2008; Sharples 2008; Schmid 2009), ED drugs have received attention for their possible contribution to these trends (Sharples 2008; Schmid 2009).

Although middle-aged and older adults generally take fewer risks with their health, their decreased need for contraception may imply less than optimal safe sexual practices compared to younger populations (Lindau *et al.* 2006). For example, prior research suggests that condom use declines with age (Patel et al. 2003; Lindau et al. 2006) and, among at-risk populations, individuals older than 50 years are one sixth as likely to use condoms during sex and one-fifth as likely to have been tested for HIV compared to individuals in their twenties (Stall and Catania 1994). Moreover, a survey of primary care physicians revealed that most physicians rarely or never discuss sexual risk factor

reduction with their middle-age or older patients (Skiest and Keiser 1997). These facts are particularly important in light of the emergence of ED drugs, which have improved sexual function among older adult males.

Several small studies in the MSM (men who have sex with men) community have investigated the connection between pharmacologic ED treatments and STDs (Kim et al. 2002; Chu et al. 2003; Paul et al. 2005; Swearingen and Klausner 2005). In this community, ED drug use is associated with high-risk sexual behavior, such as unprotected anal sex (Chu *et al.* 2003; Paul *et al.* 2005; Swearingen and Klausner 2005). ED drug users also report a greater number of recent sex partners and higher rates of STDs than non-users (Kim et al. 2002; Paul et al. 2005; Swearingen and Klausner 2005). While the measured outcomes of these studies likely reflect selection bias among users rather than the effect of ED drugs per se, these studies nonetheless highlight a group of individuals within the MSM community who are at high risk of contracting STDs.

In light of these findings and the growing use of pharmacologic treatments for ED, we investigated the relationship between STDs and ED drug use in a comprehensive, large sample of privately insured, middle-age and older adult male beneficiaries. To account for selection bias among users of ED treatments in our sample, we conducted additional analyses to examine whether the differential availability of ED drugs – generated by variation in employer coverage of these treatments – may have had a direct impact on STDs as well.

2. Methods

We assembled a data set of pharmacy and medical claims from 1997 to 2006 for 49 large US employers covering 4,178,779 beneficiaries continuously enrolled for up to 10 years (n = 12,158,049 person-years). Because we were interested in STD rates among males, we restricted our sample to include only male beneficiaries (n = 1,970,160). This led to a sample including 5,797,675 person-years. In addition, to focus on the population most likely to use erectile dysfunction drugs, we created a separate sub-sample of men above the age of 40 (n = 1,339,382) which included 4,292,445 person-years.

The pharmacy claims incorporated all prescription drug claims, each with information on the type of drug, drug name, national drug code, dosage, and days supplied. The medical claims included the date of service, diagnosis, and procedure code. These data have been used elsewhere to examine the impact of benefit design on pharmacy spending23, use of medication by the chronically-ill (Goldman et al. 2004; Goldman et al. 2006), and specialty drugs (Goldman *et al.* 2006). Although all types of health care encounters were captured - including inpatient, emergency, and outpatient services - our claims data excluded both the informal provision of prescription drugs by online or "black-market" suppliers, as well as prescriptions that were filled but not reported to the health plan. This is, of course, a possibility for drugs used to treat erectile dysfunction.

In each year, we classified an individual as using an erectile dysfunction drug if they filled one or more prescriptions in that year for either sildenafil (Viagra), tadalafil (Cialis), or vardenafil (Levitra). Use of any of these drugs was identified by searching the pharmacy claims data for both the generic and branded names of these drugs, as well as the national drug codes associated with them.

We classified an individual as having a sexually transmitted disease if they had at least one medical claim for one or more of the following: chlamydia, gonorrhea, haemophilus ducreyi, herpes, HIV/AIDS, human papilloma virus, lymphogranuloma venereum, or syphilis. Disease indicators for these STDs were identified in the medical claims according to International Classification of Disease, Ninth Revision (ICD-9) diagnoses. (Full list of ICD-9 codes used are available from the corresponding author).

We also constructed disease indicators for co-morbid conditions, some of which might be associated with the use of erectile dysfunction drugs and the likelihood of sexually transmitted disease. Separate disease indicators identified the following conditions: anxiety, asthma, cancer, cardiac disease, congestive heart failure, chronic obstructive pulmonary disease, depression, diabetes, hypercholesterolemia, hypertension, stroke, vascular disease. A beneficiary was determined to have one of these conditions if their medical claims included 2 or more office visits with the corresponding ICD-9 code (available upon request). For each person-year, we also constructed a separate indicator for whether a beneficiary had a condition known to predispose towards erectile dysfunction: anxiety, diabetes, depression, hypertension, and vascular disease.

3. Statistical Analysis

In the baseline analysis, we estimated logistic models of STD prevalence among users and non-users of erectile dysfunction drugs. The unit of observation was a person-year. We controlled for age of the beneficiary, co-morbid diseases, employer, and year. In addition to estimating the relationship between current ED drug use and current STD, we estimated logistic models in which the dependent variable was the presence of a STD in the following year.

To investigate whether the use of ED drugs might directly lead to increases in STDs, we conducted additional analyses to mitigate the selection bias that might occur if high-risk individuals are more likely to take ED drugs. While ED drug users might be systematically more inclined toward sexual risk, we hypothesized that individuals with diseases that elevate the risk of ED (e.g. diabetes, hypertension, anxiety, etc.) would arguably be no more inclined toward risky sex. Under this hypothesis, any increased prevalence of STDs among individuals with diseases predisposing to ED would be attributable to the use of ED treatments. Moreover, if this hypothesis were true, only those employers covering ED treatments would be expected to display a pattern of higher STD prevalence among individuals with greater likelihood of ED.

This hypothesis motivated our comparison of mean STD rates among those with high and low "disease-based" likelihood of having ED. And, as a further validity test, we refined this comparison by computing the difference in mean STD rates within employers that covered ED drugs, and within employers that did not cover them. To convert these estimates into odds-ratios, we then estimated an analogous logistic model: the estimated impact of ED drug coverage on STDs was the coefficient on the interaction between employer coverage and an indicator for high likelihood of ED drug use. If the use of ED treatment drives the differences in STD prevalence between high- and low-likelihood patients, we would expect to see little to no difference in STD rates within employers that do not cover ED drugs, but significant differences in those that do. Conversely, if the results are driven by an underlying correlation between disease status and the propensity for sexual risk-taking, we would expect to see differences in STD rates within both kinds of employers.

An employer's coverage rules were determined empirically using the claims data: an employer was defined as never covering an ED drug if the rate of ED drug use was below 0.01 percent; this can be compared to the mean rate of use among employers of 5.21 percent. Occasionally, employers had a

handful of claims for ED drugs that we interpret as special exceptions rather than full employer coverage of the drug (n = 26 claims out of 801,349 person-years in plans defined as not covering ED drugs). This resulted in 5 firms (16 firm-years) that never covered an ED drug and 44 firms that did (171 firm-years).

A subtle concern with our comparison across employers is the potential for adverse selection into drug plans. If sexually risky employees seek out plans with generous ED coverage, one might observe difference in STDs within covering employers, but not other employers, even though ED drugs themselves play no causal role. However, it seems plausible to rule this out, because employers – even those offering multiple drug benefit packages – very rarely offer employees a choice of pharmacy benefit packages. Put differently, individuals can "choose" drug plans only by choosing an employer. Therefore, we make the highly plausible assumption that individuals do not choose their employers based primarily on ED drug coverage in that employer's drug benefit.

STATA version 10 (STATA Corp, College Station, Texas) was used for statistical analyses and the 95% CI reflects .025 in each tail or P .05. All models were clustered either at the individual level or employer level where appropriate.

4. Results

Figure 1 displays the use of ED drugs among men above the age of 40 in our sample. Sildenafil was approved for use in erectile dysfunction by the FDA in March 1998. From 1998 to 2003, sildenafil use among men above 40 increased from 4.38 percent to 6.25 percent in our sample. Vardenafil and tadalafil were approved by the FDA in September 2003 and December 2003, respectively. With the

arrival of these competing drugs, sildenafil use dropped to 3.65 percent in 2006, as vardenafil and tadalafil steadily gained market share. In 2006, the last year of our data set, sildenafil still remained the market leader of erectile dysfunction drugs; 3.65 percent of men above 40 used sildenafil, 1.01 percent of men used vardenafil, and 1.73 percent of men used tadalafil.

Table 1 presents descriptive statistics of users and non-users of erectile dysfunction drugs in our sample of privately insured beneficiaries. Users of ED drugs differed in several important ways from non-users. First, users tended to be older (61.1 years versus 56.6, $p < 0.001$). Second, users had a substantially higher prevalence of STDs (3.44% versus 2.19%, $p < 0.001$). Third, for most diseases, users had a higher prevalence of other co-morbid conditions.

Given the various ways in which users and non-users of ED treatments differ, the unadjusted means presented in Table 1 do not account for how observed factors may confound the relationship between ED drug use and STDs. Figure 2 presents the odds-ratio of STD from a logistic regression in which the outcome measure was whether or not a beneficiary had a claim for an STD in the year of observation.

Figure shows 95% confidence intervals around odds ratios from a logistic regression explaining the presence of STDs in a population of 1,339,382 privately-insured males aged 40 and older (n = 4,292,445 person-years) with employer-provided health insurance. Diseases were classified according to ICD-9 codes available from the authors upon request. Individuals were identified as using an ED drug if they filled one or more claims for either sildenafil, vardenafil, or tadalafil. Data are from 1997–2006.

Adjusting for age and other co-morbidities, Figure 2 shows that ED drug use was associated with a contemporaneous sexually transmitted disease (OR = 1.52, 95% CI 1.48 – 1.56). Older age was

associated with a lower likelihood of STD, while all other co-morbidities except for hypercholesterolemia and stroke were associated with a higher likelihood of STD. When the outcome measure was whether or not a beneficiary had a claim for an STD in the year following utilization of an ED drug, the odds-ratio of sexually transmitted disease was 1.44 (95% CI 1.38 – 1.49, full results not shown). Contemporaneous use of an ED drug was therefore related to both the current and future likelihood of sexually transmitted disease.

To control for selection bias, we compared ED drug utilization and STDs across individuals with high and low clinical likelihood of ED.

Table 2 shows that individuals with diseases predisposing to ED were more likely to utilize ED treatments (874 v. 457 per 10,000 beneficiaries, $p < 0.001$) and had a higher prevalence of STDs compared to individuals with lower disease-based likelihood of ED (236 v. 222 per 10,000 beneficiaries, $p < 0.001$). Comparable logistic regressions explaining ED drug use and STDs by age, year, and the presence of an ED-disposing condition yielded an odds-ratio for ED-disposing conditions of 1.59 ($p < 0.001$, full results not shown) and 1.20 ($p < 0.001$, full results not shown) respectively. Under the hypothesis that individuals with and without diseases predisposing to ED would have nearly identical STD risks in the absence of ED treatments, these estimates imply the availability of these treatments would have increased STD prevalence by 236 – 222 = 14 per 10,000 males above 40.

Table 3 further refines the comparison in Table 2 by decomposing our results into employers that do and do not cover ED drugs.

Table 3 shows how employer coverage of ED drugs affected both the use of ED drugs and the prevalence of STDs among beneficiaries with low and high likelihoods of ED based on diseases

known to predispose towards erectile dysfunction. Among employers that did not cover ED drugs, there were virtually no claims for ED drugs among groups with either low or high likelihood of ED; the small percent of claims present were due to special exceptions (third column, top panel). Among employers that did cover ED drugs, those with diseases predisposing to ED were substantially more likely to use ED drugs than those without predisposing conditions (1,027 versus 565 per 10,000 men, $p < 0.001$).

In employers that did cover pharmacologic ED treatments, STD prevalence was significantly higher among those with diseases predisposing to ED (246 versus 233 per 10,000 men, $p < 0.001$). At the same time, STD prevalence per 10,000 men was nearly identical between male beneficiaries of low and high likelihood of ED in employers that did not cover ED drugs (180 versus 179, $p = 0.83$). This result goes against an alternative explanation of our findings, namely that underlying sexual propensity might be correlated with disease status. Therefore, using the group without ED-associated diseases as a control, employer offers of ED drug coverage were associated with an increase in the prevalence of STDs of $(246 – 233) – (179 – 180) = 14$ per 10,000 male beneficiaries.

Finally, logistic regression recovers the estimated impact of employer coverage of ED drugs on STD prevalence in terms of an odds-ratio, adjusting for age of the beneficiary and year. When users without diseases known to predispose to ED served as a control group, the odds-ratio of STD was 1.12 (95% CI, 1.03 – 1.24) among employers that covered ED drugs compared to those that did not.

5. Discussion

Since the introduction of Viagra (sildenafil) in 1998, pharmacologic treatments for erectile dysfunction have gained increased popularity among middle-aged and older males. By increasing levels of sexual activity, these treatments may, however, have had the unintended consequence of raising levels of sexually transmitted disease. This is particularly important given that conversations with physicians about safe sexual practices likely occur more often with younger adults than with older ones.

We investigated the relationship between STDs and ED drug use in a comprehensive, large sample of privately insured, middle-age and older adult male beneficiaries. Generally, we found that users of ED drugs had substantially higher rates of sexually transmitted disease. Because selection bias may threaten this comparison, we analyzed how STD rates differed across individuals with diseases that predispose them towards ED. This analysis was further broken down by employers that do and do not cover ED drugs. We found that ED-disposed individuals had higher rates of STDs in employers that covered ED drugs, but not in employers without such coverage. Using beneficiaries at low risk of utilizing ED treatments as a control group, we found that employer coverage of ED treatments was associated with higher rates of sexually transmitted disease (OR = 1.12).

Our analysis raises several important issues and possibilities. First, the simple fact that STD rates are higher among users of ED drugs suggests a particular subset of middle-aged and older adults for whom renewed physician conversations about safe sexual practices may be warranted. This is particularly relevant, since most primary care physicians rarely or never discuss sexual risk factor reduction with their middle-age and older patients (Skiest and Keiser 1997), and only 9% of adults aged 40-80 years report that a doctor asked them about their sexual health during a routine doctor

visit in the past three years (Moreira *et al.* 2005). Put another way, use of ED drugs by middle-aged and older patients may serve as a simple screening tool for physicians to use in identifying those patients who may benefit from reminders about safe sexual practice. This finding coincides with other researchers' recommendations that physicians should include discussions about sexual health in conversations with older patients (Kuehn 2008). Second, if the availability or increased insurance coverage of ED treatments leads to increases in STDs, a natural question is what steps, if any, should be taken to ensure responsible utilization of these treatments. For example, health plans may consider increasing co-pays for these treatments, or manufacturers may consider including information about safe sexual practices as part of package inserts or direct to consumer advertising. Physicians may also remind their older adult patients about responsible utilization of these drugs.

Our analysis has several limitations. First, we did not conduct a randomized control trial, which would be ideal in assessing the direct impact of ED drug use on STD incidence. We addressed this limitation in several ways, by looking at predisposing conditions to ED and differences in employer coverage of ED drugs. Second, we identified users of pharmacologic ED treatments from insurance claims data. Measured ED drug utilization may not capture prescriptions purchased outside of a patient's health plan. Despite this limitation, however, we still find that documented users of ED drugs have higher rates of STD than documented non-users. A related concern is that individuals in plans that do not cover ED drugs may of course purchase these drugs without prescription. For example, one study showed that sildenafil is readily available over the internet without the need for a physician visit (Armstrong et al. 1999). As a result, the impact of employer coverage on STD rates might understate the actual impact of ED drugs' availability on STD rates. In this sense, the unintended negative consequences of these therapies might be even larger than they appear.

Figure III-1 Use of erectile dysfunction drugs for men aged 40 and older with employer-provided health insurance, 1997 – 2006

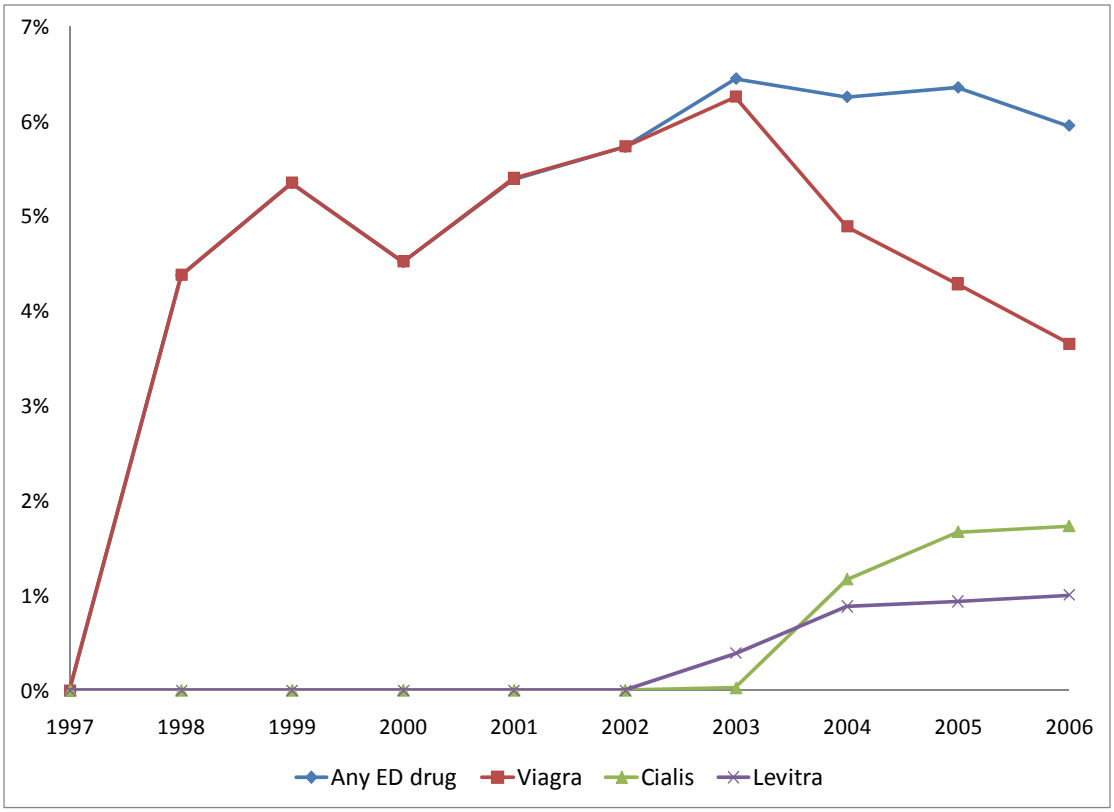

Figure III-2 Determinants of STD prevalence among privately-insured men aged 40 and older, Odds-ratios

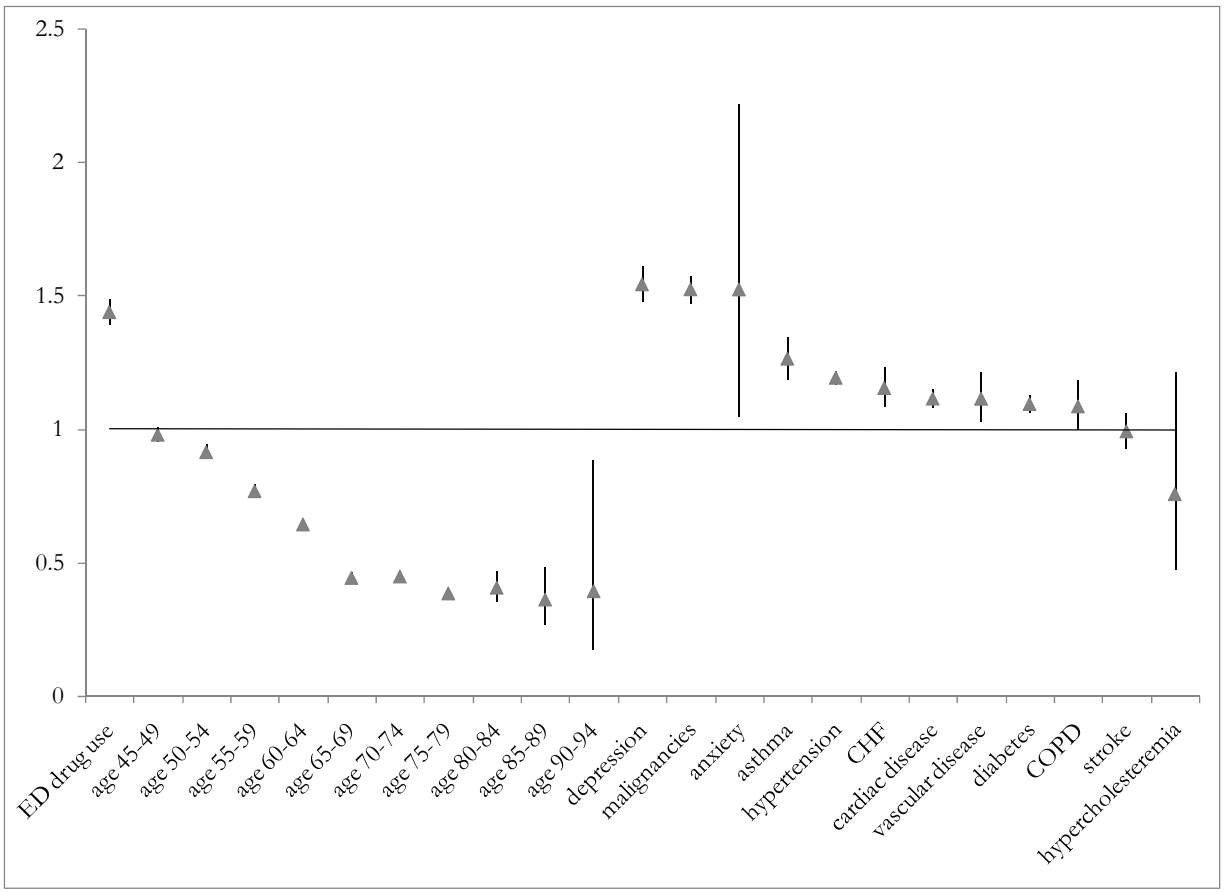

Table III-1 Summary statistics for users and non-users of erectile dysfunction drugs among men aged 40 and older with employer-provided health insurance

	Non-users	Users	p-value
Average age	57.59	61.08	< 0.001
STD (%)	2.19	3.44	< 0.001
Anxiety (%)	0.02	0.03	< 0.001
Asthma (%)	0.98	1.43	< 0.001
Cancer (%)	4.62	7.98	< 0.001
Cardiac disease (%)	8.75	11.10	< 0.001
CHF (%)	1.53	1.39	< 0.001
COPD (%)	0.73	0.96	< 0.001
Depression (%)	1.91	3.24	< 0.001
Diabetes	7.12	12.68	< 0.001
Hypercholesterolemia (%)	6.27	10.52	< 0.001
Hypertension (%)	14.14	24.41	< 0.001
Stroke (%)	1.32	1.65	< 0.001
Vascular disease (%)	0.79	0.96	< 0.001
No. of person years	4,050,254	242,191	

Notes: Users and non-users were defined annually, according to whether they filled one or more claims for either sildenafil, vardenafil, or tadalafil. Diseases were classified according to ICD-9 codes available from the authors upon request. Data come from 1997-2006. p-values are from a t-test comparison of means.

Table III-2 ED drug use and STDs among men older than 40, by disease predisposing to erectile dysfunction

	N	Use of ED drugs per 10,000 men**	Prevalence of STDs per 10,000 men
No disease predisposing to ED	3,190,717	457	222
Disease predisposing to ED	1,101,728	874	236
p-value*		< 0.001	< 0.001

Notes: Table shows mean annual use of ED drugs and mean annual prevalence of STDs per 10,000 men, by presence disease predisposing towards erectile dysfunction. Diseases were classified according to ICD-9 codes available from the authors upon request. Individuals were identified as using an ED drug if they filled one or more claims for either sildenafil, vardenafil, or tadalafil. Data are from 1997–2006. N is sample size.

***p-values test for equivalence of mean use, as well as mean STD prevalence, between individuals with diseases that do and do not predispose to erectile dysfunction.**

Table III-3 ED drug use and STDs among men older than 40, by employer coverage and disease predisposing to erectile dysfunction

	N	Use of ED drugs per 10,000 men**	Prevalence of STDs per 10,000 men
ED drugs not covered by employer			
No disease predisposing to ED	609,259	0.30	180
Disease predisposing to ED	163,642	0.49	179
p-value*		0.23	0.83
ED drugs covered by employer			
No disease predisposing to ED	2,581,458	565	233
Disease predisposing to ED	938,086	1,027	246
p-value*		< 0.001	< 0.001

Notes: Table shows mean annual use of ED drugs and mean annual prevalence of STDs per 10,000 men, by employer coverage of ED drugs and the presence of one of several diseases predisposing towards erectile dysfunction. Diseases were classified according to ICD-9 codes available from the authors upon request. Individuals were identified as using an ED drug if they filled one or more claims for either sildenafil, vardenafil, or tadalafil. Data are from 1997–2006. N is sample size.

***p-values test for equivalence of mean use, as well as mean STD prevalence, between individuals with diseases that do and do not predispose to erectile dysfunction.**

****Several plans that did not broadly cover erectile dysfunction drugs still had a trivial number claims for these drugs.**

References

Araujo, A. B., B. A. Mohr and J. B. McKinlay (2004). "Changes in Sexual Function in Middle-Aged and Older Men: Longitudinal Data from the Massachusetts Male Aging Study." J Am Geriatr Soc **52**(9): 1502-1509.

Armstrong, K., J. S. Schwartz and D. A. Asch (1999). "Direct Sale of Sildenafil (Viagra) to Consumers over the Internet." N Engl J Med **341**(18): 1389-1392.

Bacon, C. G., M. A. Mittleman, I. Kawachi, E. Giovannucci, D. B. Glasser and E. B. Rimm (2003). "Sexual Function in Men Older Than 50 Years of Age: Results from the Health Professionals Follow-up Study." Ann Intern Med **139**(3): 161-168.

Baum, C. F., M. E. Schaffer and S. Stillman (2007). "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing." Stata Journal **7**(4): 465-506.

Bertrand, M., E. Duflo and S. Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics **119**(1): 249-275.

Bodley-Tickell, A. T., B. Olowokure, S. Bhaduri, D. J. White, D. Ward, J. D. Ross, G. Smith, H. V. Duggal and P. Goold (2008). "Trends in Sexually Transmitted Infections (Other Than Hiv) in Older People: Analysis of Data from an Enhanced Surveillance System." Sex Transm Infect **84**(4): 312-317.

Burkhauser, R. V. and J. Cawley (2008). "Beyond Bmi: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research." Journal of Health Economics.

Calle, E. E., M. J. Thun, J. M. Petrelli, C. Rodriguez and C. W. Heath (1999). "Body-Mass Index and Mortality in a Prospective Cohort of U.S. Adults." New England Journal of Medicine **341**(15): 1097-1105.

Chou, S.-Y., M. Grossman and H. Saffer (2004). "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." Journal of Health Economics **23**(3): 565-587.

Chu, P. L., W. McFarland, S. Gibson, D. Weide, J. Henne, P. Miller, T. Partridge and S. Schwarcz (2003). "Viagra Use in a Community-Recruited Sample of Men Who Have Sex with Men, San Francisco." J Acquir Immune Defic Syndr **33**(2): 191-193.

Delate, T., V. A. Simmons and B. R. Motheral (2004). "Patterns of Use of Sildenafil among Commercially Insured Adults in the United States: 1998-2002." Int J Impot Res **16**(4): 313-318.

Deurenberg-Yap, M., S. K. Chew and P. Deurenberg (2002). "Elevated Body Fat Percentage and Cardiovascular Risks at Low Body Mass Index Levels among Singaporean Chinese, Malays and Indians." Obesity Reviews **3**(3): 209-215.

Deurenberg, P., M. Yap and W. A. v. Staveren (1998). "Body Mass Index and Percent Body Fat: A Meta Analysis among Different Ethnic Groups." International Journal of Obesity **22**(12): 1164-1171.

Drewnowski, A., C. L. Kurth and J. E. Rahaim (1991). "Taste Preferences in Human Obesity: Environmental and Familial Factors." American Journal of Clinical Nutrition **54**: 635.

Durnin, J. and J. Womersley (1974). "Body Fat Assessed from Total Body Density and Its Estimation from Skinfold Thickness: Measurements on 481 Men and Women Aged from 16 to 72 Years." British Journal of Nutrition **32**(01): 77-97.

Enzlin, P., R. Mak, F. Kittel and K. Demyttenaere (2004). "Sexual Functioning in a Population-Based Study of Men Aged 40-69 Years: The Good News." Int J Impot Res **16**(6): 512-520.

Fang, C. and J. C. Beghin (2002). "Urban Demand for Edible Oils and Fats in China: Evidence from Household Survey Data." Journal of Comparative Economics **30**(4): 732-753.

Frijters, P. and J. Baron (2009). Do the Obese Really Die Younger or Do Health Expenditures Buy Them Extra Years? IZA Discussion Paper No. 4149.

Gallagher, D., M. Visser, D. Sepulveda, R. N. Pierson, T. Harris and S. B. Heymsfield (1996). "How Useful Is Body Mass Index for Comparison of Body Fatness across Age, Sex, and Ethnic Groups?" American Journal of Epidemiology **143**(3): 228-239.

Garrison, R. J., W. B. Kannel, J. Stokes, 3rd and W. P. Castelli (1987). "Incidence and Precursors of Hypertension in Young Adults: The Framingham Offspring Study." Prev Med **16**(2): 235-251.

Goldman, D. P., G. F. Joyce, J. J. Escarce, J. E. Pace, M. D. Solomon, M. Laouri, P. B. Landsman and S. M. Teutsch (2004). "Pharmacy Benefits and the Use of Drugs by the Chronically Ill." Jama **291**(19): 2344-2350.

Goldman, D. P., G. F. Joyce and P. Karaca-Mandic (2006). "Varying Pharmacy Benefits with Clinical Status: The Case of Cholesterol-Lowering Therapy." American Journal of Managed Care **12**(1): 21-28.

Goldman, D. P., G. F. Joyce, G. Lawless, W. H. Crown and V. Willey (2006). "Benefit Design and Specialty Drug Use." Health Affairs **25**(5): 1319-1331.

Hall, J. E. (2000). "Pathophysiology of Obesity Hypertension." Curr Hypertens Rep **2**(2): 139-147.

Hall, J. E. (2003). "The Kidney, Hypertension, and Obesity." Hypertension **41**(3): 625-633.

He, M., K. C. B. Tan, E. T. S. Li and A. W. C. Kung (2001). "Body Fat Determination by Dual Energy X-Ray Absorptiometry and Its Relation to Body Mass Index and Waist Circumference in Hong Kong Chinese." International Journal of Obesity **25**: 748-752.

Hossain, P., B. Kawar and M. El Nahas (2007). "Obesity and Diabetes in the Developing World -- a Growing Challenge." N Engl J Med **356**(3): 213-215.

Johannes, C. B., A. B. Araujo, H. A. Feldman, C. A. Derby, K. P. Kleinman and J. B. McKinlay (2000). "Incidence of Erectile Dysfunction in Men 40 to 69 Years Old: Longitudinal Results from the Massachusetts Male Aging Study." J Urol **163**(2): 460-463.

Karlovsky, M., B. Lebed and J. H. Mydlo (2004). "Increasing Incidence and Importance of Hiv/Aids and Gonorrhea among Men Aged >/= 50 Years in the Us in the Era of Erectile Dysfunction Therapy." Scand J Urol Nephrol **38**(3): 247-252.

Kim, A. A., C. K. Kent and J. D. Klausner (2002). "Increased Risk of Hiv and Sexually Transmitted Disease Transmission among Gay or Bisexual Men Who Use Viagra, San Francisco 2000-2001." AIDS **16**(10): 1425-1428.

Kuehn, B. M. (2008). "Time For "The Talk"--Again: Seniors Need Information on Sexual Health." Jama **300**(11): 1285-1287.

Lakdawalla, D. and T. Philipson (2002). "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination." National Bureau of Economic Research Working Paper Series **No. 8946**.

Laumann, E. O., A. Nicolosi, D. B. Glasser, A. Paik, C. Gingell, E. Moreira and T. Wang (2005). "Sexual Problems among Women and Men Aged 40-80 Y: Prevalence and Correlates Identified in the Global Study of Sexual Attitudes and Behaviors." Int J Impot Res **17**(1): 39-57.

Laumann, E. O., A. Paik and R. C. Rosen (1999). "Sexual Dysfunction in the United States: Prevalence and Predictors." Jama **281**(6): 537-544.

Lean, M. E., T. S. Han and P. Deurenberg (1996). "Predicting Body Composition by Densitometry from Simple Anthropometric Measurements." American Journal of Clinical Nutrition **63**(1): 4-14.

Lindau, S. T., S. A. Leitsch, K. L. Lundberg and J. Jerome (2006). "Older Women's Attitudes, Behavior, and Communication About Sex and Hiv: A Community-Based Study." J Womens Health (Larchmt) **15**(6): 747-753.

Lindau, S. T., L. P. Schumm, E. O. Laumann, W. Levinson, C. A. O'Muircheartaigh and L. J. Waite (2007). "A Study of Sexuality and Health among Older Adults in the United States." N Engl J Med **357**(8): 762-774.

Mokdad, A. H., E. S. Ford, B. A. Bowman, W. H. Dietz, F. Vinicor, V. S. Bales and J. S. Marks (2003). "Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001." JAMA **289**(1): 76-79.

Mokdad, A. H., J. S. Marks, D. F. Stroup and J. L. Gerberding (2004). "Actual Causes of Death in the United States, 2000." JAMA **291**(10): 1238-1245.

Moreira, E. D., Jr., G. Brock, D. B. Glasser, A. Nicolosi, E. O. Laumann, A. Paik, T. Wang and C. Gingell (2005). "Help-Seeking Behaviour for Sexual Problems: The Global Study of Sexual Attitudes and Behaviors." Int J Clin Pract **59**(1): 6-16.

Must, A., G. E. Dallal and W. H. Dietz (1991). "Reference Data for Obesity: 85th and 95th Percentiles of Body Mass Index (Wt/Ht2) and Triceps Skinfold Thickness [Published Erratum Appears in Am J Clin Nutr 1991 Nov;54(5):773]." American Journal of Clinical Nutrition **53**(4): 839-846.

National Institutes of Health (1998). "Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report " NIH Publication **No. 98-4083**.

Ng, S. W., F. Zhai and B. M. Popkin (2008). "Impacts of China's Edible Oil Pricing Policy on Nutrition." Social Science & Medicine **66**(2): 414-426.

Nicolosi, A., E. O. Laumann, D. B. Glasser, E. D. Moreira, Jr., A. Paik and C. Gingell (2004). "Sexual Behavior and Sexual Dysfunctions after Age 40: The Global Study of Sexual Attitudes and Behaviors." Urology **64**(5): 991-997.

Ogden, C. L., M. D. Carroll, L. R. Curtin, M. A. McDowell, C. J. Tabak and K. M. Flegal (2006). "Prevalence of Overweight and Obesity in the United States, 1999-2004." Journal of the American Medical Association **295**(13): 1549-1555.

Ory, M., Mack, K. (1998). "Middle-Aged and Older People with Aids: Trends in National Surveillance Rates, Transmission Routes, and Risk Factors." Research on Aging **20**(6): 11.

Patel, D., B. Gillespie and B. Foxman (2003). "Sexual Behavior of Older Women: Results of a Random-Digit-Dialing Survey of 2,000 Women in the United States." Sex Transm Dis **30**(3): 216-220.

Paul, J. P., L. Pollack, D. Osmond and J. A. Catania (2005). "Viagra (Sildenafil) Use in a Population-Based Sample of U.S. Men Who Have Sex with Men." Sex Transm Dis **32**(9): 531-533.

Piers, L., M. Soares, S. Frandsen and K. O'Dea (2000). "Indirect Estimates of Body Composition Are Useful for Groups but Unreliable in Individuals." International Journal of Obesity **24**(9): 1145-1152.

Rolland-Cachera, M. F., P. Brambilla, P. Manzoni, M. Akrout, S. Sironi, A. Del Maschio and G. Chiumello (1997). "Body Composition Assessed on the Basis of Arm Circumference and

Triceps Skinfold Thickness: A New Index Validated in Children by Magnetic Resonance Imaging." American Journal of Clinical Nutrition **65**(6): 1709-1713.
Romero-Corral, A., V. M. Montori, V. K. Somers, J. Korinek, R. J. Thomas, T. G. Allison, F. Mookadam and F. Lopez-Jimenez (2006). "Association of Bodyweight with Total Mortality and with Cardiovascular Events in Coronary Artery Disease: A Systematic Review of Cohort Studies." The Lancet **368**(9536): 666-678.
Schmid, G. (2009). "The Unexplained Story of Hiv and Ageing." WHO Bulletin **87**(3): 1.
Sharples, T. (2008). "More Midlife (and Older) Stds." Time Magazine.
Skiest, D. J. and P. Keiser (1997). "Human Immunodeficiency Virus Infection in Patients Older Than 50 Years. A Survey of Primary Care Physicians' Beliefs, Practices, and Knowledge." Arch Fam Med **6**(3): 289-294.
Stall, R. and J. Catania (1994). "Aids Risk Behaviors among Late Middle-Aged and Elderly Americans. The National Aids Behavioral Surveys." Arch Intern Med **154**(1): 57-63.
Swearingen, S. G. and J. D. Klausner (2005). "Sildenafil Use, Sexual Risk Behavior, and Risk for Sexually Transmitted Diseases, Including Hiv Infection." Am J Med **118**(6): 571-577.
Wang, Y., J. Mi, X. Y. Shan, Q. J. Wang and K. y. Ge (2007). "Is China Facing an Obesity Epidemic and the Consequences? The Trends in Obesity and Chronic Disease in China." International Journal of Obesity **31**(1): 177-188.
Williams, D. P., S. B. Going, T. G. Lohman, M. J. Hewitt and A. E. Haber (1992). "Estimation of Body Fat from Skinfold Thicknesses in Middle-Aged and Older Men and Women: A Multiple Component Approach." American Journal of Human Biology **4**: 595-605.
World Health Organization (2000). "Obesity: Preventing and Managing the Global Epidemic." WHO Technical Report Series **894**.
World Health Organization (2004). "Appropriate Body-Mass Index for Asian Populations and Its Implications for Policy and Intervention Strategies." The Lancet **363**(9403): 157-163.