WIAS Discussion Paper No.2011-001

Is Chronic Health a Normal Good?
Evidence from the Effect of Hypertension Diagnosis on Food Consumption

April 13, 2011

Meng Zhao[a], Yoshifumi Konishi[b], Paul Glewwe[c]

[a] *Waseda Institute for Advanced Study, Waseda University, Tokyo, Japan*
[b] *Faculty of Liberal Arts, Sophia University, Tokyo, Japan*
[c] *Department of Applied Economics, University of Minnesota, St. Paul 55108, USA*

WIAS 早稲田大学高等研究所
Waseda Institute for Advanced Study

1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan
Tel: +81-3-5286-2460 ; Fax: +81-3-5286-2470

Is Health a Normal Good? Evidence from the Effect of Hypertension Diagnosis on Food Consumption

Meng Zhao*, Yoshifumi Konishi, Paul Glewwe

ABSTRACT

Income and chronic health outcomes often do not exhibit a clear empirical relationship, despite the conventional wisdom that health itself is a normal good. We identify health information as the key to understanding the multiple effects of income on the demand for health. As their incomes rise, richer individuals demand both better health and more health information, yet unhealthy foods such as sweets and fatty and oily foods also become more affordable. Using the health capital framework of Grossman (*JPE*, 1972), this study tests the hypotheses that individuals adjust their diet in a healthier direction upon receiving negative health information, and that the effect is larger for richer individuals. Both measurement and endogeneity of hypertension information present challenges to identifying causal relationships between diet, chronic health conditions and health information. To overcome both of these problems, we adopt a regression discontinuity design approach that exploits the exogenous cutoff of systolic blood pressure in the diagnosis of hypertension. Based on unique Chinese longitudinal data, we find the following: richer individuals are more likely to develop hypertension; the positive income-hypertension gradient disappears once past food consumption is controlled for; upon receiving a diagnosis of hypertension, individuals reduce fat intake significantly; and richer individuals reduce their fat intake more in response to hypertension diagnosis.

JEL codes: D12; I12; Q11

Key words: China, diet, health, health information, health capital model, hypertension, regression discontinuity, nutrient intake

---

* Corresponding author. Tel.: +81 3 5286 2928; fax: +81 3 5286 2470.
E-mail address: konishi-moe@aoni.waseda.jp. §The paper is based on Essay 2 of the corresponding author’s doctoral dissertation, which was in part supported by the Hueg-Harrison Fellowship at University of Minnesota.

1. Introduction

A large number of recent studies have found high prevalence rates of chronic health conditions such as hypertension, cardiovascular diseases and cancer, among high income groups in developing countries (Case, *et al*., 2004; Van de Poel et al., 2009; Gaziano *et al.*, 2010; Gersh *et al.*, 2010; Koch *et al.*, 2010), which stands in contrast to earlier studies that showed positive income-health gradients (Grossman, 1972; Pritchett and Summers, 1996; Deaton, 2003). On one hand, individuals with higher incomes can afford better health care and more nutritious diets and lifestyles. On the other hand, with higher incomes individuals can afford unhealthy diets and lifestyles such as high-calorie foods, cigarettes and alcohol, all of which increase the probability and severity of many chronic diseases. An increased demand for *health* due to higher incomes is more likely to result in better health *outcomes* if individuals have adequate health information to guide their daily lifestyle choices. Yet the availability *and* acceptance of reliable health information by consumers depend on their innate demand for such information, both at the individual and the aggregate level. A correct understanding of the channels by which income affects chronic health outcomes requires that one disentangle consumer demands for health itself *and* for health information.

This study investigates the causal relationships between diet, chronic health outcomes, and health information, carefully accounting for endogeneity in each of these variables. As consumers' knowledge of either private or public health information may correlate with some unobservables that affect their food consumption, estimates of the impact of such knowledge on consumers' choices is likely to suffer from omitted variable bias. Moreover, since the *quality* of health information is important for consumers' decision making, health knowledge is often difficult to quantify. To circumvent both of these problems, we adopt an regression discontinuity (RD) approach, which can mimic a random experiment under certain assumptions (Black, 1999; Angrist and Lavy, 1999; Van der Klaaw, 2002; Lee and Lemieux, 2010).

Using unique Chinese longitudinal data, this paper examines the variation in nutrient intake patterns among individuals with different incomes to new information regarding their true health status, as measured by hypertension status.[1] We use the RD approach by exploiting the fact that hypertension status is determined by one's blood pressure reading relative to a sharply defined cutoff point established by medical experts.[2] Since individuals cannot *precisely* control their blood pressure, among those with blood pressure readings near the cutoff, some randomly are above it while others randomly fall below it, which can be regarded as a random assignment of hypertension status. Because the consumption patterns and other behaviors are likely to be almost identical for the samples right below and right above the cutoff, the difference in the outcomes between these two groups may be used to estimate the treatment effect – i.e. the effect of being informed that one has hypertension.[3]

The data used in this study are from the China Health and Nutrition Survey (CHNS), which was conducted in China in five separate rounds from 1989 to 2004. We

---

[1] Hypertension is an asymptomatic condition that is considered to be one of the most critical risk factors of major chronic diseases (Vasan *et al.*, 2002) and is estimated to affect approximately one third of the world's population (Kearney *et al.*, 2005).

[2] According to the American Heart Association, one is judged to have hypertension if one's systolic blood pressure is above 140 mmHg or if one's diastolic blood pressure is above 90mmHg.

[3] Of course, individuals may respond differently to the notification of hypertension status, depending on their knowledge of hypertension or other health information, either before or after receiving the blood pressure tests. However, there is no a priori reason to believe that individuals' *prior* knowledge differs systematically between the two samples around the cutoff, as the assignment is based on the cutoff point. The estimated treatment effect of being diagnosed as having hypertension may include systematic differences between the two samples in the endogenous accumulation of knowledge on diet and hypertension *after* receiving information regarding one's blood pressure. Moreover, if all individuals in the survey had been perfectly informed of their own hypertension status, the effect would have been weak or null. But this was not the case, as will be seen below. In our sample, three-quarters of those with blood pressure above the cutoff were unaware of their condition. Compare this number with a corresponding rate of 20% for the hypertensive population in United States and Canada (Wolf-Maier *et al.*, 2004).

exploit several features of this data set to control for potential biases. First, the panel nature of the data allows one to condition the outcome variables, intake of four macronutrients (energy, fat, protein and carbohydrates), on a diagnosis of hypertension that occurred as much as 3-4 years earlier. Second, instead of relying on self-reported hypertension status, we draw upon the blood pressure test results from a physical examination conducted for *every* individual surveyed in *each* round of the survey. Finally, since the blood test results are communicated to *all* survey subjects, the data do not suffer from sample selection bias, which is often a problem in identifying the effect of information provision.

Our main results are as follows. First, our results confirm earlier findings in the epidemiological literature that the rich are *more likely to have hypertension in China*, which suggests that good health *outcomes* can be an *inferior* good, and that this effect is likely to occur mainly through an unhealthy diet – the positive income-hypertension gradient disappears once one controls for food consumption and nutrition intake in previous time periods. Second, our non-parametric RD estimates indicate that, on average, individuals who were informed that they have hypertension have reduced their fat intake by about 7.7 grams per day 3-4 years after the blood test. More importantly, estimates by different income groups confirm the theoretical prediction that rich individuals are more responsive to a diagnosis of hypertension than the poor, which implies that good health is indeed a *normal* good, conditional on past food consumption and health information. In fact, the estimates on fat intake are significant only for the rich group.

The rest of the paper is organized as follows. Section 2 reviews the literature on health and income, and the role of information in determining health behaviors.

Section 3 provides background information on China and describes the data. Section 4 presents an organizing framework for empirical analysis, from which our hypothesis is derived. Section 5 discusses the identification and estimation strategies used in the regression-discontinuity design framework. The main results are presented in Section 6, and the last section concludes.

2. Literature Review

In the past several decades, economists have devoted much effort to understand the relationship between health and income. The existing literature has found a relationship that ranges from strongly positive to weakly negative (Fuchs, 2004), with different economists providing different interpretations of this relationship. Whereas some studies claim to have found a positive causal effect of income on health (Pritchett and Summers, 1996; Deaton, 2003), others argue that causality may run in the opposite direction, or through third factors such as education and access to health services (Grossman, 2006; Fuchs, 2004). Moreover, some recent epidemiological studies have found a reverse relationship between income and health in some developing countries.[4] For example, Koch *et al.* (2010) find that income is positively associated with mortality rate in an 8-year cohort study conducted in Chile. Another study, Van de Poel *et al.* (2009), also finds that income is positively associated with the prevalence of obesity and hypertension using data from 1991 to 2004 in China. Our study attempts to

---

[4] Epidemiologists divide epidemiological transition as income rises into four stages. The first stage, at the lowest level of income, is characterized by widespread communicable diseases and malnutrition. In the second stage, as incomes rise chronic health conditions start to emerge. The third stage is defined as the period when burden of chronic diseases exceeds that of infection and malnutrition, accompanied by increased risk factors such as fatty diet, inactivity, and smoking. In this period, the majority of the chronic diseases occur among the privileged ones. In the fourth stage, chronic diseases are the major causes of death and the burden of chronic diseases is mainly born by those in lower socio-economic status. (Gersh *et al.*, 2010).

reconcile these seemingly contradictory findings by examining the role of health information and the competing effects of income on chronic health outcomes.

The study most closely related to ours is that conducted using data from South Africa by Case *et al.* (2004). While their finding that richer people are more likely than the poor to take hypertensive medication is consistent with our result, they also found a puzzling result that observed hypertension does not exhibit a significantly negative relationship with income among those who participated in medical exams. They suspect that this puzzling result may occur because richer individuals have a higher risk of hypertension. Indeed, our study provides empirical support for their conjecture – we find a positive hypertension-income gradient, which comes mainly from individual food consumption. Thus, their result is likely to suffer from endogeneity bias due to the multiple pathways by which income affects health, as discussed above. In addition, Case and her coauthors note that their result is likely to suffer from serious sample selection bias, as only 30% of their sample self-selected to take the medical exams. Lastly, while they focus on medical compliance, hypertension needs a more comprehensive analysis that incorporates long-term changes in lifestyle and diet. Our study overcomes both endogeneity and sample selection bias by employing the regression discontinuity method, and it also focuses on the effect of hypertension diagnosis on consumers' diet patterns.

This study is not the first attempt to estimate the impacts of health information on food demand. Several studies (Brown and Schrader, 1990; Chern *et al.*, 1995; Kaabia *et al.*, 2001; Roosen *et al.*, 2009) have investigated how consumers respond to the provision of *public* health information on what constitutes a healthy diet, while others have examined the effects of nutrition labels or social marketing (Martin *et al.*, 1994;

Crutchfield *et al.*, 2001). Many of these studies focus on the short-term impacts of public information on consumer behavior. Yet Grossman's framework posits that consumers build up health capital through long-run health investments, such as sustained efforts to change dietary habits and lifestyles. Viewed in this light, gauging the short-run effect of information is insufficient. In a field experiment conducted in France, for instance, warning of poison in fish modified household fish consumption only slightly, and the impact became insignificant after only three months. Roosen *et al.* (2009) attributed such lack of impact to consumers' weak memory of the information provided. In contrast, our study focuses on the effect of hypertension diagnosis on daily dietary patterns three to four years later.

Furthermore, the findings from the above-mentioned studies should be treated with caution, as they are likely to suffer from endogeneity of health information. For example, Brown and Schrader (1990) created a health information index based on counts of journal articles that found links between cholesterol and heart disease. They found that health information, as measured by the health information index, reduced the per capita demand for eggs by 16% to 25% in United States from 1955 to 1987. Similarly, Kim and Chern (1999) created a cholesterol information index using a modified weighting method, assuming that articles published during specific time periods can have carry-over and decay effects. The study found evidence that health information on fat and cholesterol increased the consumption of fish oil and reduced the use of lard, tallow and palm oil in Japan. These studies rely heavily on the assumption that the numbers of article published are exogenous, which may not be the case as medical research is in fact often driven by public interest and financial support from industry and governments. Our study circumvents the endogeneity problem of health information by adopting the RD approach and exploiting the dynamic features of the longitudinal data from China.

Another problem with some of these previous studies is a potential sample selection bias. For example, Crutchfield *et al.* (2001) analyzed the impact of nutrition labels to estimate the economic benefit of new rules that require the provision of nutrition information for all the raw meat and poultry products. They show that providing these nutrition labels decreases the intake of fat and cholesterol and, therefore, reduces the risks of developing future cases of stroke, cancer and heart disease. However, since those who care more about the potentially harmful effects of food consumption may also look for and read nutritional and other labeling on the products more carefully, the estimated effect of nutrition or social labeling may suffer from sample selection bias, *even if* the introduction of the label itself is exogenous. In other words, the estimation of the effect of nutrition labeling was based on only the sample who had noticed the label, who may be systematically different from those who had not. In contrast, this study is less likely to suffer from such sample selection bias, because the blood pressure test results were communicated to *all* subjects (so everyone at least sees the result) *and* the hypertension diagnosis is based on a well-defined cutoff in systolic blood pressure.[5]

---

[5] Aside from the economic literature, there are also a number of studies in epidemiology and public health examining the effect of health information on lifestyles. For example, Milne *et al.* (1985) show that individuals who are both newly and previously diagnosed as having hypertension are more likely to report a weight loss. Neutel and Campbell (2008) find that individuals who were newly diagnosed as having hypertension in Canada tend to quit smoking and increase physical activities. However, as the hypertension diagnosis examined in these studies is often based on self chosen physical examinations, their estimates are likely to suffer from sample selection bias. Moreover, these studies provide little insight as to how consumers' responses to the diagnosis of hypertension interact with other socio-economic variables, such as income.

3. Background and Data

China's economy has grown rapidly since 1980, with an average real GDP per capita growth rate of 8% during the past three decades. During that same period, the prevalence of chronic diseases in China has also increased. Chronic diseases accounted for only 65% of total deaths in China in 1982, but by 2005 they accounted for about 80% (Bryant, 2003; Wang *et al.*, 2005). The implications of this trend for health care costs are formidable: according to the World Health Organization (WHO, 2005), 560 billion U.S. dollars will be foregone during 2000-2015 due to chronic diseases in China – by far the highest health care costs among all of the countries examined in that study.[6]

This study draws upon a comprehensive household panel survey, the China Health and Nutrition Survey (CHNS), which collected five rounds of data in China from 1989 to 2004 (in 1989, 1991, 1997, 2000, and 2004). The survey data are approximately national representative: sampling with probability proportional to size (PPS) and stratified by income, the CHNS samples are randomly selected from 9 provinces in China. The CHNS data offer two types of data that are particularly important for this study: (1) detailed food and nutrition intake; and (2) blood pressure test results. The CHNS data also provide detailed information on socio-demographic characteristics for each survey subject.

Due to concerns about sample attrition among the survey subjects in early rounds, we use only the data from the three most recent rounds of the CHNS (1997, 2000, and 2004). In 1997, 8,688 individuals were surveyed, 72.3% (6,283) of whom were re-interviewed in 2000. In 2000, an additional 3,516 individuals were included in the survey to maintain the original sample size, resulting in a total sample of 9,799

---

[6] Other countries included in this study are: Brazil, Canada, India, Nigeria, Pakistan, Russia, Tanzania and the United Kingdom.

individuals in 2000. Of the individuals in the 2000 sample, approximately 71.1% (6,969) were re-interviewed in 2004. Individuals who dropped out the survey tend to be young males in rural areas, who are most likely to become migrant workers. Our analysis draws upon the pooled sample of these 13,252 individuals (6,283 from 1997 and 6,969 from 2000) who were interviewed twice in continuous surveys.

A. Food and Nutrition Intake

The CHNS used trained investigators to record each household's food intake over three consecutive days, following a standard procedure. The three consecutive interview days were randomly selected from Monday to Sunday and so were spread throughout a whole week. The investigator interviewed each household member each day, recording detailed information on his or her recalled food consumption. The food consumption data are very detailed; more than 1500 types of food items were recorded in the survey. Using this 3-day food intake data and the Chinese food nutrition table compiled by Yang (2002), the Carolina Population Center calculated each person's daily intake of four macronutrients: carbohydrates, energy, fat and protein.

The quantity and composition of Chinese food demand has changed considerably during the past two decades. Figures 1 and 2 display trends in the means of 3-day per capita consumption of 11 major food categories from 1989 to 2004.[7] Consumption of beef and poultry increased rapidly starting in 1991 (Figure 1). Consumption of shrimp and crab had been low until 1997, but rose dramatically thereafter. For non-animal products, consumption of melons and nightshades (e.g. eggplants, tomatoes, chili peppers, etc.) has increased steadily (Figure 2), as rising incomes allow people to afford

---

[7] All quantities are standardized to the levels in 1989.

more expensive varieties of food. In contrast, consumption of traditional types of vegetables, such as grains, leafy vegetables, roots and stems, has gradually declined. The sole exception, which coincides with the trend in seafood consumption, is that the demand for fruits also rose quickly after 1997.

Aggregating food consumption patterns into the four macronutrients, and comparing across income groups, yields even sharper changes in the structure of the Chinese diet.[8] First, energy intake has declined in all three income groups (Figure 3-a). The decline is driven mainly by a sharp drop in protein intake between 1991 and 1997 (Figure 3-c) and a steady decrease carbohydrate intake (Figure 3-d). In addition, this change has been accompanied by a shift in the source of energy: the percentage of Chinese people who obtain more than 30% of their total energy from fat has increased from 14.7% in 1989 to 44.1% in 2006 (Popkin, 2008), which indicates that the Chinese diet has become increasingly high in fat. Figure 3-b shows an overall increase in fat consumption for all income groups from 1991 to 2004. The trend increases slightly faster for the poor and the middle income groups, although the rich generally consume 20 grams more of fat per day than do the poor.

B. Hypertension and Blood Pressure Test

The CHNS data contain two measures of hypertension: (a) self-reported hypertension status[9] and (b) clinical blood pressure levels from individual physical examinations conducted by professionally trained investigators. The trained examiners measured systolic and diastolic blood pressure three times for each individual in each round of the

---

[8] The sample is divided into three groups, namely rich, middle income and poor, according to adjusted per capita income. Throughout this paper, the rich group is defined as the top third of the income distribution, the middle group is the middle third, and the poor are the bottom third.

[9] This is the answer to a question asked to each adult aged 18 years old or plus: "Has a doctor ever told you that you suffer from high blood pressure?"

survey. The survey personnel then informed people of the results of their physical examinations. We use the latter since it is likely to be more accurate (see also our discussion on a "fuzzy" RD in Subsection 6C).

The age-weighted prevalence of hypertension among the population aged 18 or above in China has been rising steadily over time, from 19.1% in 1991 to 26.8% in 2004 (Column (b) in Table 1), and the prevalence is higher among the rich than among the poor in all years.[10] Approximately three quarters of the survey respondents who were found to have hypertension were unaware of their illness, and the percentage of those who were aware is generally higher for the rich than for the poor or middle income groups.[11] Among those people who reported that they had been diagnosed as having hypertension by a doctor, the percentage of those who were taking anti-hypertension drugs increased from 60.6% in 1991 to 74.4% in 2004. However, these drugs do not appear to be very effective – only 28.4% of those who took anti-hypertension drugs were keeping their condition under control in 2004, though the rate for keeping hypertension under control is approximately one and a half times the rate in 1991. Though Table 1 does seem consistent with the idea that the rich appear more informed of their hypertension status and are more engaged in anti-hypertension activity, it is only suggestive at this point – it may be that the more health concerned among the rich are diagnosed as having hypertension and therefore are more responsive to the diagnosis. To identify the *causal* effects of health information and incomes, we rely on the regression

---

[10] The common rule used to determine hypertension, e.g. by China's Department of Health, American Heart Association, World Health Organization, is: one is hypertensive if his or her systolic blood pressure is above 140mmHg and/or if his or her diastolic blood pressure is above 90mmHg. The hypertension rates are weighted by age to net out the effect of population aging.

[11] The table is also suggestive of interesting relationships between diet, health, and health information. Between 1991 and 1997, the prevalence of hypertension increased rapidly, by 5.6 percentage points, presumably due to the structural change in China's diet. Yet, consumer awareness of it did not keep up with this rapid pace, leading to the sharp decline in the rate of awareness in 1997.

discontinuity design, as explained below.

Furthermore, simple random-effect (RE) logit regression analysis also indicates that the rich generally are more likely to develop hypertension than the poor, as shown in the first and second columns of Table 2. However, when nutrient intake and food consumption are included, as shown in the third and the fourth columns, the correlation between income and hypertension status disappears.

4. Empirical Framework

We present an organizing framework for our empirical analysis, building on the health capital framework of Grossman (1972). According to recent clinical findings, daily diet is one of the most important determinants of health capital. We thus posit that consumer's health capital in the next period, H_{t+1}, depends on current intake of fat, F_t, current intake of other nutrients, N_t, and current health capital, H_t.

$$H_{t+1} = I(F_t, N_t) + (1 - \delta_t)H_t, \tag{1}$$

where I is the health investment function, with $I_F < 0$ and $I_N > 0$, and $\delta_t \in [0,1]$ is the depreciation rate of health capital.[12] Except for the investment function I, all other essential components of Grossman's model are maintained. One important prediction of the model is that, as in Grossman, an increase in the wage rate raises the marginal benefit of health capital since it increases the opportunity cost of the working time lost due to sickness and results in higher demand for both health investment and health capital.

We now introduce information imperfections into the model. First, assume that the

---

[12] The health effect of fat can be positive when under consumed, e.g. in the case of malnutrition. For simplicity, we assume it is bad for health as the health outcome we are interested in is hypertension, for which fat intake is generally considered to have a negative impact.

consumer observes her true initial health status (H_0) with an *i.i.d.* error:

$$\widetilde{H}_0 = H_0 + \epsilon_0 \,. \tag{2}$$

where $\widetilde{H}_0$ is observed health status. Note that some individuals may have accurate knowledge of their initial health status, so that $\epsilon_0 = 0$, while others either overestimate or underestimate their initial health status ($\epsilon_0 > 0$ or $\epsilon_0 < 0$). Assume that the consumer makes health investments over time, observing her initial health capital with errors but knowing the health production function (1).[13] One important implication of this modeling strategy is that, because her true and observed health capitals evolve according to (1), she consistently misperceives her health capital over time, with an error evolving according to:[14]

$$\epsilon_{t+1} = (1 - \delta_t)\epsilon_t. \tag{3}$$

This equation implies that consumers who overestimated their initial health status continue to overestimate their health, but such over-confidence decreases over time as they get older.

Once in a while, however, the consumer experiences a chronic health condition or has a medical checkup, either of which allows her to observe her true health status H_t. Grossman (1972) does recognize the importance of medical checkups (e.g. p.227), but focuses on the time-investment aspect rather than the informational aspect of such activities. Assuming perfect information, Grossman's model fails to explain how information from medical checkups impacts consumers' decision making processes with

---

[13] Obviously, there are a number of other ways consumers may misperceive their health conditions. For example, consumers may have inaccurate knowledge on the effect of health investment *I* or depreciation rate of health capital *δ*. The extent of such misperception may also interact with consumers' education levels. We refrain from these complications in order to obtain clear predictions from the model.

[14] To see this, observe that, for any time period *t*,

$$\begin{aligned}\widetilde{H}_{t+1} &= I(F_t, N_t) + (1 - \delta_t)\widetilde{H}_t \\ &= I(F_t, N_t) + (1 - \delta_t)(H_t + \epsilon_t) \\ &= H_{t+1} + (1 - \delta_t)\epsilon_t\end{aligned}$$

respect to investment in health capital.

Importantly, if a consumer realizes that her true health capital is lower than she thought it was (due to $\widetilde{H}_{\mathrm{t}}^{*}$, being higher than H_{t}^{*}), she must further increase her health investment to raise her actual health capital to the optimal level that she thought she previously had. That is,

$$\widetilde{H}_{\mathrm{t+1}}^{*} = I(\widetilde{F}_{\mathrm{t}}^{*}, \widetilde{N}_{\mathrm{t}}^{*}) + (1 - \delta_{\mathrm{t}})\widetilde{H}_{\mathrm{t}}^{*} > I(\widetilde{F}_{\mathrm{t}}^{*}, \widetilde{N}_{\mathrm{t}}^{*}) + (1 - \delta_{\mathrm{t}})H_{\mathrm{t}}^{*}.$$

Provided that $I_F < 0$ and $I_N > 0$, it follows that: $F_t^{*} < \tilde{F}_t^{*}$ and $N_t^{*} > \widetilde{N}_t^{*}$ iff $H_{\mathrm{t}}^{*} < \widetilde{H}_{\mathrm{t}}^{*}$, where $(\tilde{F}^{*}, \widetilde{N}^{*})$ and (F^{*}, N^{*}) indicate the consumer's optimal consumption path along her perceived health and her true health, respectively. Moreover, since individuals with higher wage incomes demand higher health capital than those with lower wage incomes, *ceteris paribus*, the former must adjust their diet more than the latter, *conditional on* observing the *same* true health capital.

Hypothesis: *Consumers adjust their diet toward less/more fat intake and more/less intake of other nutrients, upon receiving a negative/positive health information shock. Moreover, dietary adjustments are greater for consumers with higher wage incomes than those with lower wage incomes, conditional on the same true health status.*

Our approach to testing this hypothesis is to use a regression-discontinuity estimator, making use of the exogenous cutoff in blood pressure readings for diagnosing hypertension. Hypertension status is one of the most important measures of one's health capital. Let $h_{\mathrm{i}} = 1$ if individual i has hypertension (we suppress t henceforth). An individual is diagnosed as having hypertension if her blood pressure level b_{i} is above a cutoff level c. The individual observes b_{i} with an error ϵ_i:

$\tilde{b}_{\mathrm{i}} = b_{\mathrm{i}} + \epsilon_{\mathrm{i}}$.

An individual whose belief $\tilde{b}_{\mathrm{i}}$ lies below c but whose true blood pressure b_{i} lies above c gets a negative information shock: $\tilde{h}_{\mathrm{i}} - h_{\mathrm{i}} < 0$. Analogously, there are two other cases: $\tilde{h}_{\mathrm{i}} - h_{\mathrm{i}} > 0$ and $\tilde{h}_{\mathrm{i}} - h_{\mathrm{i}} = 0$. Therefore, any individual whose true blood pressure b_{i} lies above c would get either no information surprise (her $\tilde{b}_{\mathrm{i}}$ also lies above c) or a negative information shock. Assuming that individuals consider hypertension status h (not the value of blood pressure per se) as an important part of their health capital, we posit that $F_i^* < \tilde{F}_i^*$ and $N_i^* > \tilde{N}_i^*$ iff $\tilde{h}_{\mathrm{i}} - h_{\mathrm{i}} < 0$. It follows then that we should observe $E[F^* - \tilde{F}^* \leq 0 | b > c]$ and $E[F^* - \tilde{F}^* \geq 0 \big| b < c]$ where the expectation operator is taken over all i. Hence, we should observe that:

$$\lim_{b \to c^+} E[F^*(b) - \tilde{F}^*(b)] - \lim_{b \to c^-} E[F^*(b) - \tilde{F}^*(b)]$$

$$= \lim_{b \to c^+} E[F^*(b)] - \lim_{b \to c^-} E[F^*(b)] \leq 0.$$

Note that $\lim_{b \to c^+} E[\tilde{F}^*(b)]$ and $\lim_{b \to c^-} E[\tilde{F}^*(b)]$ are the same because, before knowing their true hypertension status, people whose true blood pressure is close to the cutoff behave similarly and demand the same amount of $\tilde{F}^*$. An analogous expression exists for other nutrients N. According to our model, this difference should also be larger for richer individuals.

5. Identification and Estimation

A. The Regression Discontinuity Design

A regression discontinuity (RD) design is used to test our hypotheses by identifying the causal effect of hypertension diagnosis on nutrition intake.[15] The study exploits the fact

---

[15] RD was first introduced by Thistlethwaite and Campbell (1960) and has gained popularity in empirical research in economics since the late 1990s because it mimics experimental design under certain assumptions. See Lee and Lemieux (2010) for a review of papers that have adopted the RDD approach.

that hypertension status is a *deterministic* function of continuous blood pressure measures – i.e. an individual is judged to have hypertension if either her systolic blood pressure (SBP) is above 140 mmHg and/or her diastolic blood pressure (DBP) is above 90 mmHg. Though DBP is also an important indicator of hypertension, patients and physicians often pay closer attention to SBP (Kannel, 2000).[16] Thus, for simplicity this study focuses only on SBP. Importantly, all survey subjects were informed of their blood pressures and whether or not they had hypertension, using these cutoffs, in each survey round.

This assignment of hypertension status in the CHNS lends itself well to estimating causal impacts of health information using an RD estimation method. Consider a random sample of individuals with data on the outcome measure, Y_i, and the treatment indicator T_i, where the subscript *i* indicates the *i*th individual. T_i equals one if *i* receives the treatment and equals zero otherwise. The standard parametric econometric specification to evaluate the treatment effect is:

$$Y_\mathrm{i} = \alpha + \beta T_\mathrm{i} + u_\mathrm{i} \tag{4}$$

where β measures the treatment effect and u_i is the unexplained variation in Y_i. If the assignment of the treatment is random, then β can be consistently estimated by Ordinary Least Square (OLS). However, if the treatment is not randomly assigned, it is likely that $\mathrm{E}[u|T] \neq 0$, in which case the OLS estimate will be biased. In this study, the treatment of interest is the notification of hypertension status and the outcomes of interest are nutrition intakes. Individuals develop hypertension due to a variety of unobservable factors such as diet, lifestyle and genetic inheritance. Therefore, the assignment of hypertension status is often endogenous in a non-experimental setup.

---

[16] Moreover, hypertension due to high DBP only is often treated differently from that due to high SBP in medical practice.

The RD design can circumvent this problem by exploiting the cutoff point for blood pressure that determines the diagnosis of hypertension. Consider the individuals who are within a small interval in the neighborhood of the cutoff point. Because these individuals have essentially the same blood pressure, and since they cannot *precisely* control their blood pressure on a particular day of the blood pressure test, some may fall slightly below, and some may fall slightly above, the cutoff point. As the average characteristics of the two samples slightly below and slightly above the cutoff are likely to be the same (the validity of this assumption is discussed below), the average outcomes for the two samples should be the same *in the absence of treatment*. Thus, in the small neighborhood of the cutoff point, our RD design mimics a randomized experiment.

If consumers are well informed of their hypertension status based on the cutoffs, we could use a "sharp" RD design.[17] In a "sharp" RD, the assignment of treatment T_i is based on a deterministic function of the cutoff: i.e. $T_i = 1(b_i \geq c)$. Because the cutoff is fixed, the error term u_i is uncorrelated with T_i conditional on b_i, so that $E[u_i|T_i, b_i] = E[u_i|b_i] = f(b_i)$ where $f(.)$ is a flexible continuous function of b_i. We can thus rewrite equation (4) as:

$$Y_i = \alpha + \beta T_i + f(b_i) + \mu_i \qquad (5)$$

where $\mu_i = u_i - E[u_i|T_i, b_i]$ with $E[\mu_i|T_i] = 0$. If equation (5) is linear in T_i and $f(.)$ can be correctly specified, the treatment parameter β can be estimated using OLS. The primary issue in OLS estimation is the choice of the functional form for f. If over-specified, the estimate of β is consistent yet inefficient; if under-specified, the

---

[17] If consumers are informed of their SBP readings but not of hypertension status *and* if consumers are not well informed of the cutoffs, a "fuzzy" RD may be more appropriate. We will return to this discussion in Subsection 6C.

estimate will be efficient but inconsistent. Quadratic polynomials are preferred for their flexibility in practice (Lee *et al.*, 2004).

However, since there is no *a priori* reason to believe that the underlying model is linear, estimation is often done non-parametrically (Lee and Lemieux, 2010). Following Hahn *et al.* (2001), we estimate a local linear regression (LLR), using only the data close to the cutoff point. Thus the treatment effect is:

$$\beta = \mathrm{Y}^{+} - \mathrm{Y}^{-} \tag{6}$$

where $Y^{+} = \lim_{b \to \mathrm{c}^{+}} \mathrm{E}[Y_{\mathrm{i}} | b_{\mathrm{i}}]$ and $Y^{-} = \lim_{b \to \mathrm{c}^{-}} \mathrm{E}[Y_{\mathrm{i}} | b_{\mathrm{i}}]$. The LLR estimators for Y^{+} and Y^{-} are given by δ_{Yr} and δ_{Yl} from the following optimization:

$$(\delta_{\mathrm{Yr}}, \theta_{\mathrm{Yr}}) \equiv \mathrm{argmin}_{\delta,\theta} \sum_{\mathrm{i:b_i \geq c}} [Y_{\mathrm{i}} - \delta - \theta (b_{\mathrm{i}} - c)^2] \lambda_{\mathrm{i}},$$

$$(\delta_{\mathrm{Yl}}, \theta_{\mathrm{Yl}}) \equiv \mathrm{argmin}_{\delta,\theta} \sum_{\mathrm{i:b_i} < c} [Y_{\mathrm{i}} - \delta - \theta (b_{\mathrm{i}} - c)^2] \lambda_{\mathrm{i}},$$

where $\lambda_{\mathrm{i}} = K(\frac{b_{\mathrm{i}} - c}{h})$ is a kernel function. Following the literature, a triangular kernel is used in this study, since the choice of the kernel function "typically has little impact in practice" (Lee and Lemieux, 2010).

B. Preliminary Checks on the Regression Discontinuity Design

A regression discontinuity design is appropriate only if two assumptions are satisfied (Hahn *et al.*, 2001). First, the individuals being studied cannot precisely control the value of the assignment variable (i.e. systolic blood pressure in our case). Second, the observed or unobserved characteristics of individuals whose SBP readings are right above or below the cutoff point do not differ systematically. The first assumption appears to trivially hold in this setting, as one cannot precisely control his or her systolic blood pressure at a particular time of the day. Though one can surely influence blood pressure by taking some measures such as anti-hypertension drugs, one cannot control the effectiveness of such measures.

To check if the second assumption holds, we first examine the distribution of several observable socio-economic factors by blood pressure in a manner similar to an experimental design. As shown in Figure 4, these variables are distributed continuously around the cutoff point of 140 mmHg. To check for unobservables, Lee and Lemieux (2010) suggest examining the distribution of the assignment variable itself. This is done in Figure 5. The kernel density of systolic blood pressure shows that SBP is approximately normally distributed, without a notable change in its distribution around the cutoff point.

Since there is no systematic difference between the two samples to the left and to the right of the cutoff point, if we observe changes in the outcomes variables of interest, they are likely to be due to the treatment, i.e. notification of hypertension status. To see this, the distributions of four measures of daily nutrition intake (fats, energy, protein and carbohydrates) in the next round (3-4 years later) are plotted against the systolic blood pressure in the previous round in Figures 6-9. The distributions are smoothed by the polynomials of degree two to allow for a flexible relationship between the outcomes and blood pressure.[18] The smoothing process is done separately for the samples to the left and to the right of the cutoff points. Moreover, this is done on the whole sample ("full sample") as well as separately for the three income groups (poor, middle income, rich). We observe a clear discontinuity in fat and energy intake, especially for the rich group, which suggests that the RD design is appropriate.

---

[18] Different orders of polynomials were tried, and the results are similar.

6. Results

Local linear regressions were estimated, with optimal bandwidths, for the four measures of the nutrient intake: fats, energy, protein and carbohydrates. As discussed in Section 5, we use individual systolic blood pressure (SBP) in the 1997 and 2000 CHNS surveys as our assignment variable, and daily nutrient intake from the next survey round (i.e. 3-4 years later), as recorded in the 2000 and 2004 surveys, as our outcome variable.[19] Following Imbens and Kalyanaraman (2009), we first estimate the consistent optimal bandwidth around the cutoff by minimizing the mean integrated squared error. We then use the sample within the interval to estimate the upper and lower limits of outcomes at the cutoff point. Lastly, the standard errors are calculated for statistical inference. For robustness, the estimation is also done using a bandwidth that is half of the calculated optimal bandwidth.[20] In addition, we estimate the parametric specification with quartic polynomials with and without other covariates (see Table 4 for the summary statistics of these covariates).

A. Main Results

Table 3 summarizes our main results. For each type of nutrient, we present both non-parametric (top two rows) and parametric (bottom two rows) estimates, for both the whole sample (the first column) as well as for the subsamples corresponding to the different income groups (the second, third and fourth columns). The non-parametric

---

[19] We have also considered the share of a component nutrient intake (e.g. fat) in the total energy intake as an alternative dependent variable. We dismissed this idea because when a major nutrient intake such as fat increases, the total calorie intake also increases so that the share does not change as sharply. It is certainly true that by reducing the fat share of the total calorie intake, consumers may build better chronic health over time. The variation in such a variable is, however, much smaller than one would like. For instance, in our data the proportion of calories coming from fat does not differ significantly even between hypertensive and non-hypertensive people (29.3% and 28.9%, respectively).

[20] As will be discussed in subsection 5.1, we have also estimated the local linear regression with different bandwidths. Due to space limitations, we report only the results from these two bandwidths in Table 3. The results with other bandwidths are available upon request.

estimate using the optimal bandwidth indicates that a diagnosis of hypertension induces individuals to reduce their fat intake by 7.7 grams per day, which is approximately 10% of average total fat intake. This estimate is significant at the 5% level, and it is also quite robust to the choice of bandwidth: reducing the bandwidth by half does not change the magnitude of the estimate, although it seems to reduce its efficiency, which is what one would expect given that the number of observations is smaller. However, the parametric estimates of the treatment effect are generally smaller and less significant.

The results for each income group (poor, middle and rich) reveal that the effect of hypertension diagnosis on fat intake is significant only for the rich group, with the magnitude of the estimate increasing to 10.2 grams (nonparametric estimate with optimal bandwidth). Indeed, the estimate for the poor is slightly positive, though insignificant. We thus conclude that the estimated effect on the whole sample is driven mainly by the impact on the rich group. This result is essentially the same in the non-parametric estimation that uses half of the optimal bandwidth, and for the parametric estimates, both with and without additional controls. Thus these results provide strong support for our hypothesis in Section 4.

Our results for other nutrients, energy, protein, and carbohydrates, are less clear-cut. The estimates for energy intake for the whole sample and for the rich group are generally negative, indicating that a diagnosis of hypertension induces people to reduce their calorie intake. However, these effects are statistically insignificant. The estimated effects for both protein and carbohydrates vary in sign, depending on groups and estimation methods, and all are completely insignificant.

We suspect that these less clear-cut results for other nutrients may reflect the general advice to people who have hypertension. The most common advice regarding diet for a hypertensive individual includes: (a) consuming less fats and meats; (b) reducing the consumption of salt; (c) cutting back calorie intake if overweight or obese; (d) maintaining a moderate amount of protein intake; and (e) consuming more fruits, vegetables and whole grains (Appel et al., 2006). The advice on the consumption of fats, red meats and salt is clear and relatively easy to follow, while the advice on intakes of energy and protein is less clear and hard to implement. According to the World Health Organization (2009), the prevalence of being overweight or obese is still very low in China, below 15% and 5% respectively, so consumers may be less responsive to the advice on energy intake.

B. Discussion on Robustness

The above results rely heavily on local linear estimation, so it is critical to examine the sensitivity of the estimates in Table 3 to different bandwidths and specifications. First, our local linear regression results on fat intake seem robust to the choice of different bandwidths, as discussed in Section 6A. Either an increase or a decrease in the bandwidth does not lead to significantly different results.

Second, we have also estimated parametric regressions with different polynomials, as recommended by Imbens and Lemieux (2008). As the order of polynomials increases, the estimated effect of hypertension diagnosis on fat intake rises and becomes more significant, suggesting that a more flexible functional form may be preferable. Including additional controls tends to reduce the magnitude of the estimates, which suggests that some of these controls may be correlated with the treatment variable, causing a downward bias in the estimates. This is a concern for the parametric results. Since the parametric estimation is based on *all* the data, including observations “far away” from the cutoff point, a poor approximation of the function f in equation (5) may result in the violation of the RD assumption and, consequently, lead to biased estimates.

Though our parametric estimates are generally smaller in magnitude than our non-parametric estimates, it is generally hard to compare these two methods. Parametric estimates may be biased if a polynomial is a poor approximation of the function f, while the non- parametric local linear regression estimates may also be biased if the model is non-linear even within a close neighborhood of the cutoff point. However, the bias resulting from the linearity assumption in the local linear regression decreases as smaller bandwidths are used. The robustness of the local linear regression estimates to smaller bandwidths and the sensitivity of the parametric estimates to inclusion of covariates thus seem to suggest that the non-parametric estimates are more likely to be consistent than the parametric estimates.

Lastly, we have also estimated the parametric specifications in which the income tertiles are interacted with “being informed of hypertension status” (Table 5). To avoid inappropriate approximation of function f in equation (5), these regressions use only data points “close to” the cutoff, based on the optimal bandwidths estimated by the local linear regressions in Table 3. After controlling for the quartic polynomials of SBP readings and other major socio-economic status, we find the results are fairly consistent with our non-parametric results: only the rich group responds to the notification of hypertension and cut their fat intake by 9.5 grams significantly. Moreover, as indicated by the interaction of “being informed” and education, people with higher education levels are more responsive in increasing their intake of protein and carbohydrates. There seems to be no significant difference in the reaction to hypertension notification between men and women, although men tend to consume more of all these nutrients.

C. “Sharp” vs. “Fuzzy” RD

A “sharp” RD is appropriate if consumers are perfectly informed of their hypertension status based on their SBP readings relative to the cutoff. Although the interviewers in the CHNS survey communicated the blood pressure test results to all sample subjects, there is still a concern that the interviewers might have communicated *only* their SBP readings in some cases and *not* their hypertension status. If this occurs, whether individuals can self-diagnose hypertension based on her SBP readings is critical for the consistency of the “sharp” RD estimates. We note that the bias may not be severe, because the cutoffs used for hypertension diagnosis are often regarded as common knowledge in China, as individuals often receive physical checkup reports from hospitals, which carry the information on the blood pressure cutoffs. If, however, it is indeed the case that some subjects are only informed of the blood pressure readings *and* are not well informed of the cutoffs, our “sharp” RD estimates would be biased downward and a “fuzzy” RD may be more appropriate.

Unfortunately, due to the data limitation, we were not able to credibly estimate the “fuzzy” RD. In order to implement the “fuzzy” RD, we would need to have a treatment variable that indicates whether individuals have ever been diagnosed to have hypertension, regardless whether it was diagnosed by a doctor or a survey interviewer.. In the CHNS, however, we only have data on self-reporting of a doctor’s diagnosis on hypertension status (see Subsection 3B). For the “fuzzy” RD to be valid, the probability of receiving doctor’s diagnosis has to increase discontinuously at the cutoff. As seen from Figure 10, which reports the proportion of people who reported to have been diagnosed of hypertension by a doctor, no significant “jump” can be identified at the cutoff of SBP readings, suggesting that the self-reported diagnosis of a doctor is a poor measure of the actual hypertension status and thus cannot serve as the treatment variable in our setup. Therefore, the “fuzzy” RD appears inappropriate given the data we have. Indeed, we have tried the “fuzzy” RD with the self-reported doctor’s diagnosis of hypertension as the treatment variable. However, since the discontinuity in the probability of being treated at the cutoff point is small, it resulted in very large standard errors of the estimates of the treatment effect. Therefore, we only report the results from the “sharp” RD, with a caveat that our estimated treatment effects may be underestimated if some individuals are not perfectly aware of the cutoff and, therefore, of their hypertension status.

7. Conclusions

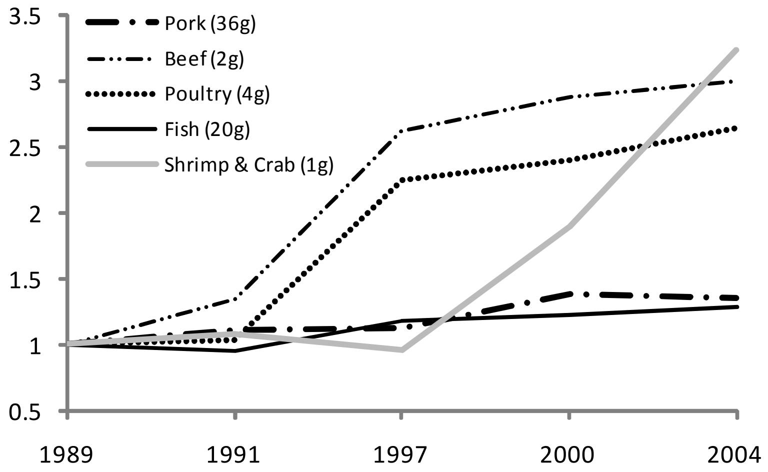
This study has investigated empirical relationships between diet, chronic health outcomes and health information, using rich longitudinal data from a series of national surveys on Chinese households. Building upon the health capital framework of Grossman (1972), a regression discontinuity approach was used to help disentangle the competing effects of income on the demand for health. The study provides strong empirical support for the hypotheses that consumers adjust their lifestyles toward healthier ones upon receiving negative health information, and that richer individuals respond much more to such information. Hypertension diagnosis induces Chinese individuals to reduce their daily fat intake by 7.7 grams three to four years after their diagnosis, although this effect is significant only for the richest third of the population.

Our findings are in a sharp contrast with those of Case *et al.* (2004). They found no significant relationship between observed hypertension and income among individuals who participated in medical exams and were diagnosed as having hypertension. Their result was puzzling, as one would expect to see a negative relationship if richer individuals are more likely to adhere to medical protocols, as our model predicts, and if all individuals face the same risk of hypertension. As the authors themselves noted, their study seems to suffer from both sample selection and endogeneity problems – richer individuals may face higher risk of hypertension because their diet and lifestyles are less healthy. Indeed, in Section 3 we find a significantly positive income-hypertension gradient, which mainly reflects food consumption patterns. Our results are sensible and suggest that these biases were reduced, as our results suggest that, after hypertension was diagnosed, only rich individuals significantly reduced fat intake. The disparity in the response to health information across income groups may increase as more health information becomes available and, eventually, lead to a positive relationship between observed health outcomes and income as commonly observed in developed countries.

Lastly, our result may be interpreted as evidence that consumers are imperfectly informed of their own health status. In our regression discontinuity design, the difference between the right and left one-sided limits of expected fat intake can occur only if individuals adjust their diet upon being diagnosed as having hypertension. In public health, preventive care and medical examinations are considered to be important policy interventions. Our study provides empirical support for the idea that consumers are imperfectly informed of their health status and thus need to be informed of their true health status through regular medical exams. In particular, our finding that individuals adopt a less fatty diet upon hypertension diagnosis supports the provision of medical exams as an effective public health intervention.

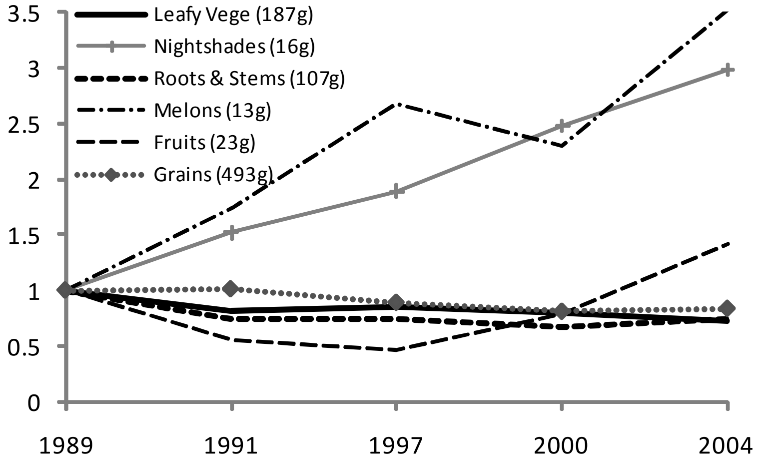
Figures and Tables:

Figure 1: Trends in 3-Day Meat and Seafood Consumption (1989 Value = 1)



Note: 1989 values in parentheses, unit is g/3 days.
Source: China Health and Nutrition Survey, 1989-2004

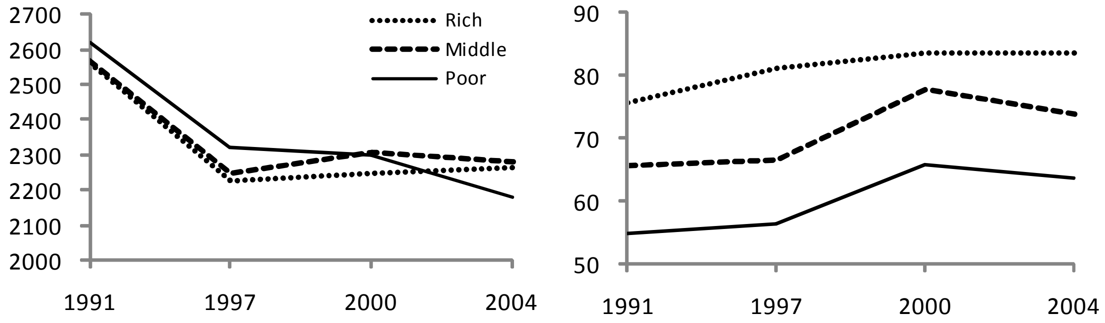
Figure 2: Trends in 3-Day Non-animal Food Consumption (1989 Value = 1)



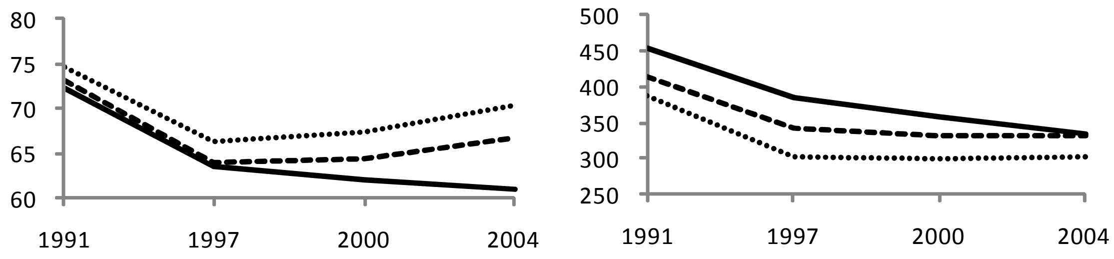
Note: 1989 values in parentheses, unit is g/3 days.
Source: China Health and Nutrition Survey, 1989-2004

Figure 3: Trends in Daily Nutrient Intake

a. Energy (kcal) b. Fat (g)



c. Protein (g) d. Carbohydrates (g)



Source: China Health and Nutrition Survey, 1991-2004

Figure 4: Local Mean Smoothing of Major Socio-economic Variables by Systolic Blood Pressure (Bandwidth = 3 mmHg)

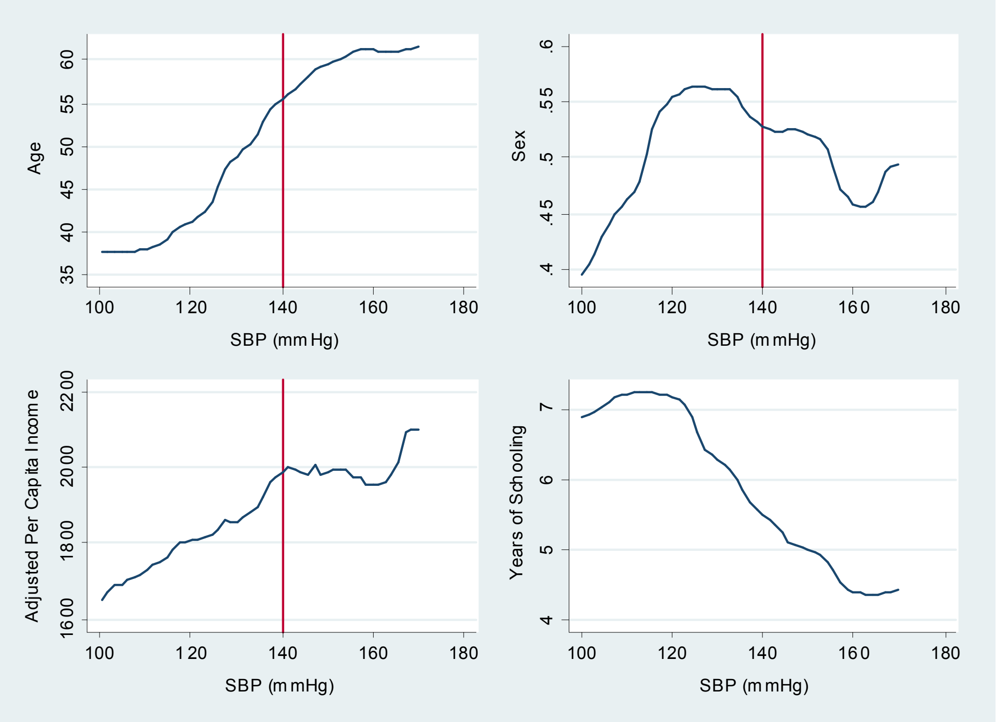


Figure 5: Distribution of Systolic Blood Pressure

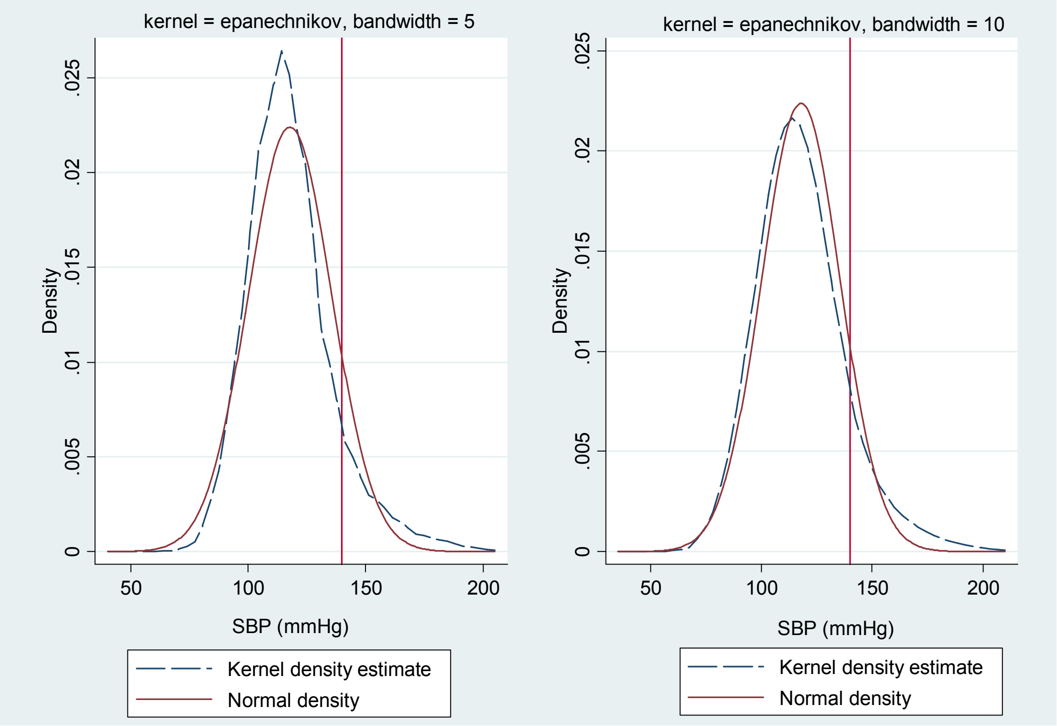
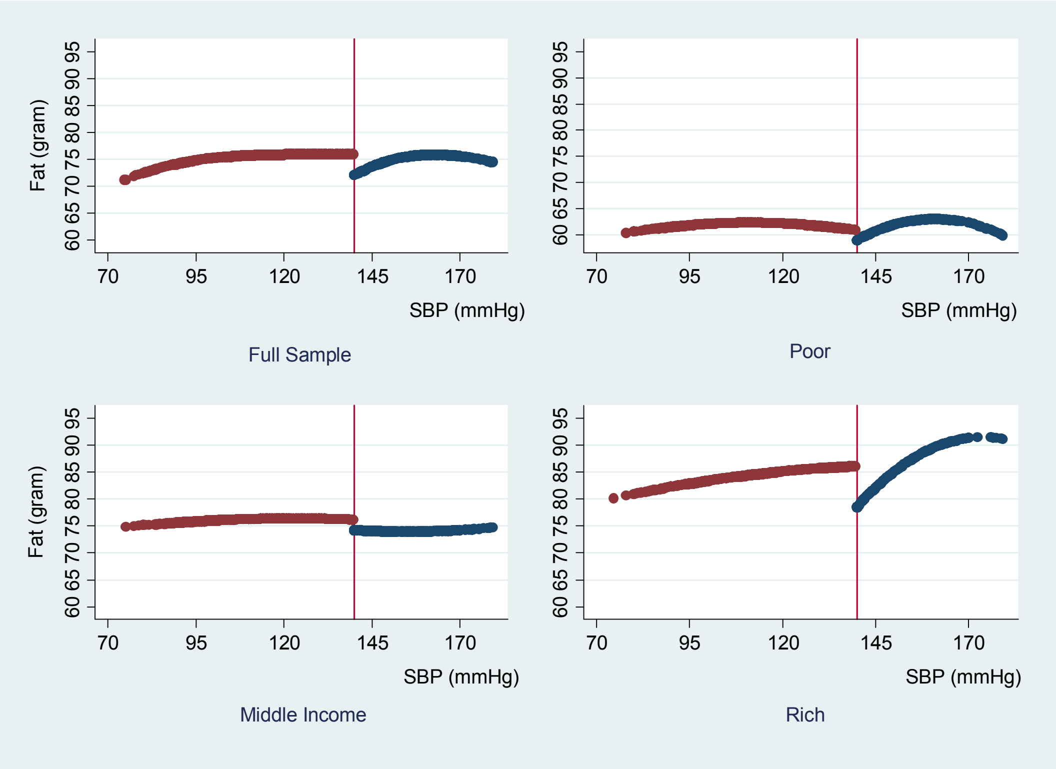
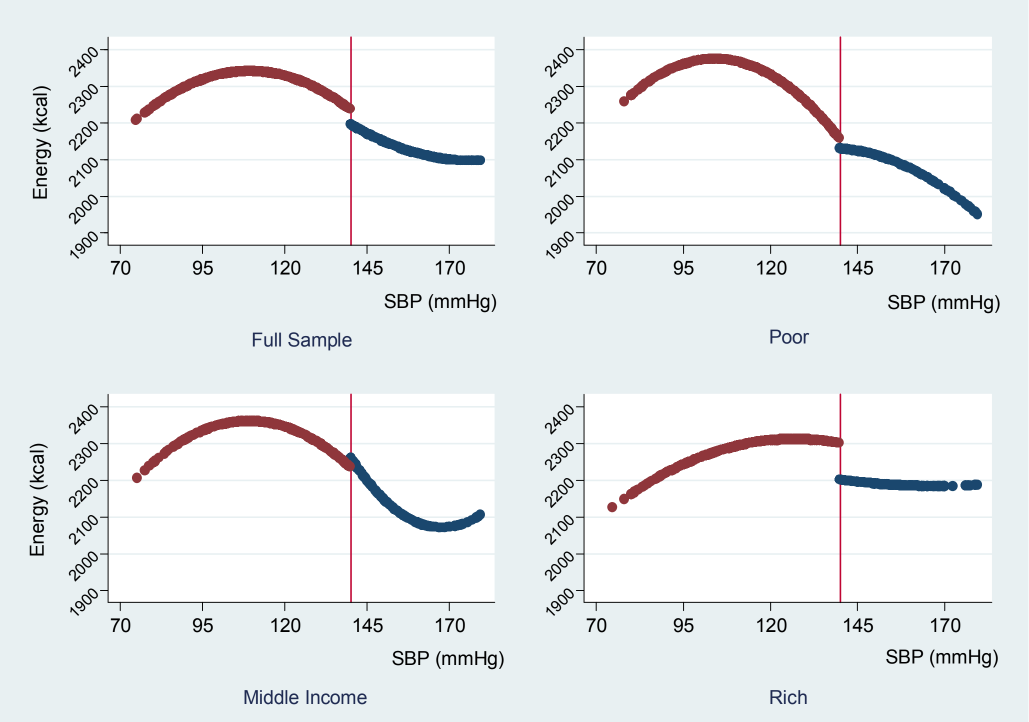


Figure 6: Predicated Daily Fat Intake from Polynomial Regression by Systolic Blood Pressure



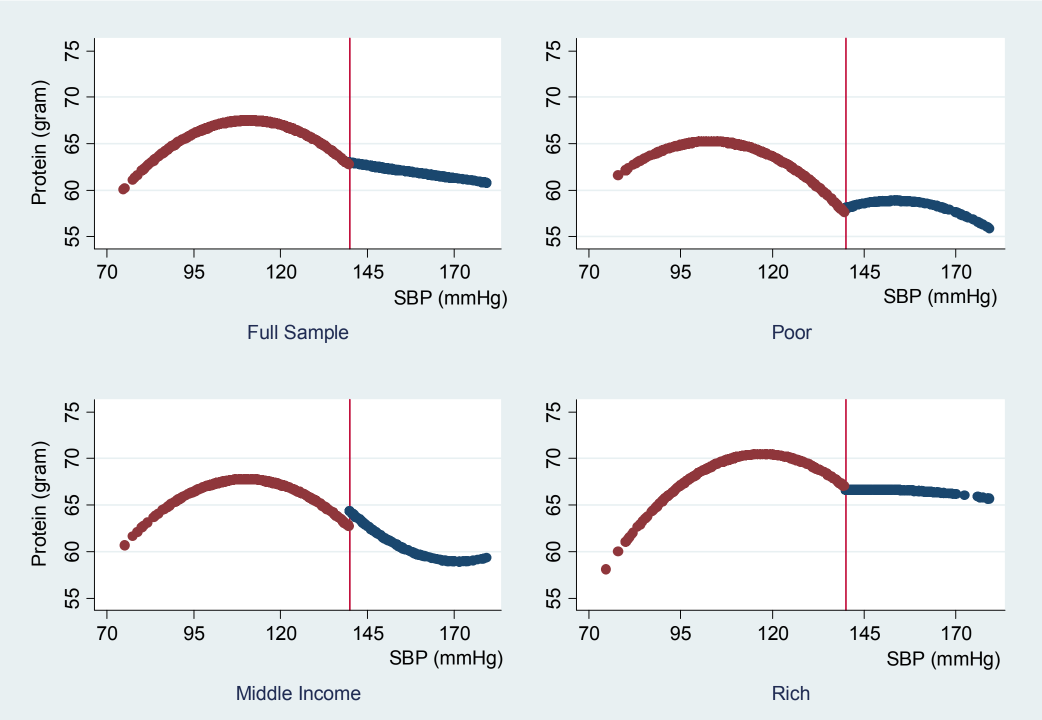
Note: The SBP was measured 3-4 years ago. The vertical line indicates the cutoff point of 140 mmHg.

Figure 7: Predicated Daily Energy Intake from Polynomial Regression by Systolic Blood Pressure



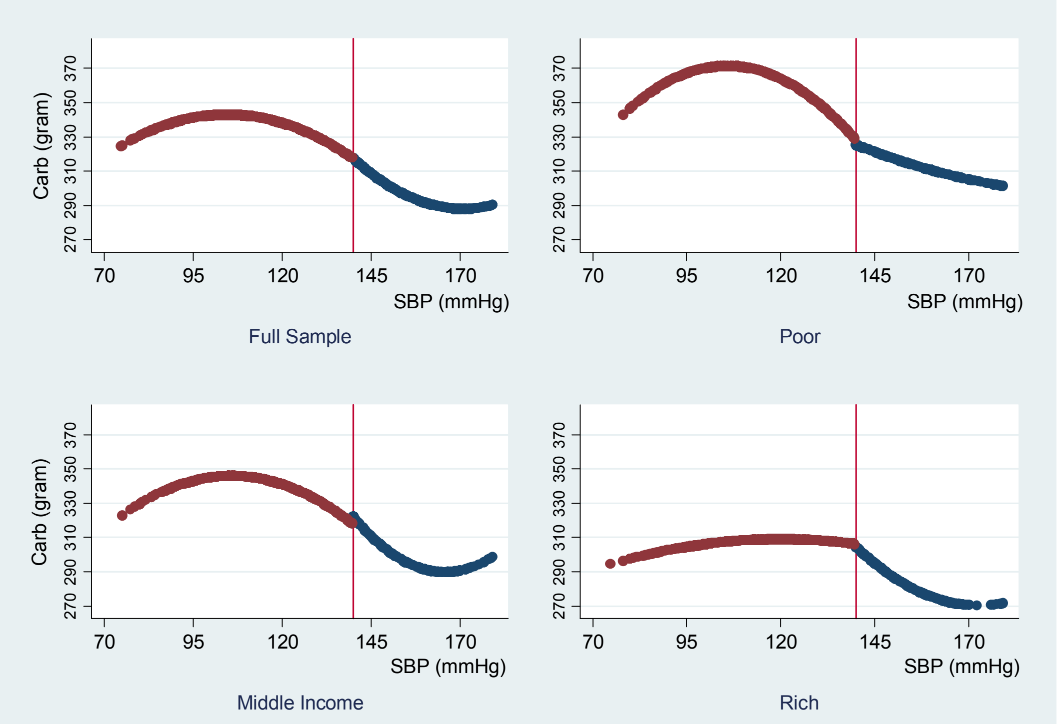
Note: The SBP was measured 3-4 years ago. The vertical line indicates the cutoff point of 140 mmHg.

Figure 8: Predicated Daily Protein Intake from Polynomial Regression by Systolic Blood Pressure



Note: The SBP was measured 3-4 years ago. The vertical line indicates the cutoff point of 140 mmHg.

Figure 9: D Predicated Daily Carbohydrate Intake from Polynomial Regression by Systolic Blood Pressure



Note: The SBP was measured 3-4 years ago. The vertical line indicates the cutoff point of 140 mmHg.

Figure 10: Percentage of Self-reported Hypertension Diagnosed by a Doctor, Bandwidth = 5 mmHg

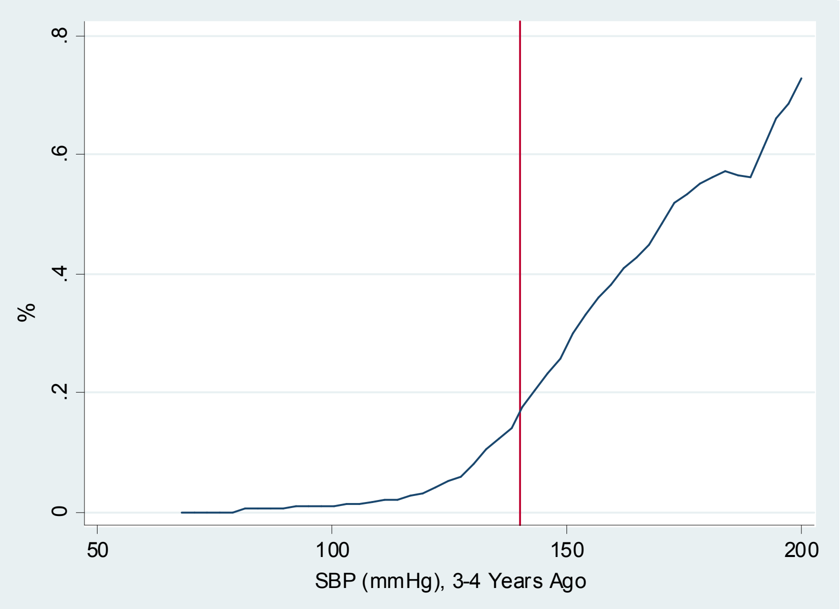


Table 1: Hypertension Status and Awareness, Age 18+

Year	Income Group	(a) Obs.	(b) Among (a), % of the Hypertensive	(c) Among (b), % of People Aware of the Condition	(d) Among (a), % Diagnosed of Hypertension	(e) Among (d), % Taking Anti-Hypertension Drug	(f) Among (e), % Keeping Condition Controlled
1991	Poor	3,007	17.5	20.9	4.6	62.9	21.9
	Middle	3,039	21.0	31.1	7.2	59.9	10.8
	Rich	2,675	18.7	30.4	6.8	60.2	21.9
	Total	**8,721**	**19.1**	**27.7**	**6.2**	**60.6**	**19.3**
1997	Poor	2,940	23.9	14.4	4.6	54.2	20.3
	Middle	2,909	24.0	19.5	5.8	68.8	18.0
	Rich	2,814	26.2	25.2	8.1	67.5	19.7
	Total	**8,663**	**24.7**	**19.8**	**6.1**	**67.5**	**19.3**
2000	Poor	3,244	25.1	23.2	7.1	64.5	16.7
	Middle	3,194	22.3	27.2	8.4	67.8	19.5
	Rich	3,120	26.4	33.5	12.4	76.2	27.4
	Total	**9,558**	**24.6**	**28.1**	**9.3**	**70.7**	**22.7**
2004	Poor	2,935	25.4	25.5	8.8	67.9	24.3
	Middle	3,104	27.6	32.0	11.9	67.1	22.2
	Rich	3,148	27.3	34.6	14.0	84.4	34.4
	Total	**9,187**	**26.8**	**30.9**	**11.6**	**74.4**	**28.4**

Note: All the percentages are weighted by age.

Source: China Health and Nutrition Survey, 1991-2004

Table 2: Logit Regressions of Hypertension Status on Socio-economic Variables

	(1) Current Hypertension Status, Full Sample, Logit RE	(2) Current Hypertension Status, Partial Sample, Logit RE	(3) Future Hypertension Status, Logit RE	(4) Future Hypertension Status, Logit RE
Middle Income (1=yes)	0.116 **	0.155 **	-0.001	-0.044
	(0.046)	(0.065)	(0.060)	(0.074)
High Income (1=yes)	0.114 **	0.212 ***	0.095	0.016
	(0.050)	(0.070)	(0.066)	(0.080)
Age	0.138 ***	0.142 ***	0.134 ***	0.136 ***
	(0.008)	(0.013)	(0.012)	(0.016)
Age-squared	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***
	(0.000)	(0.000)	(0.000)	(0.000)
Age>60 (1=yes)	0.280 ***	0.312 ***	0.107	0.027
	(0.076)	(0.113)	(0.109)	(0.139)
Sex (1=male)	0.548 ***	0.482 ***	0.495 ***	0.524 ***
	(0.047)	(0.064)	(0.062)	(0.074)
BMI	0.001 ***	0.003 ***	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Urban (1=yes)	0.291 ***	0.300 ***	0.220 ***	0.249 ***
	(0.051)	(0.070)	(0.068)	(0.081)
Years of Schooling	-0.023 **	-0.019 **	-0.015 **	-0.022 **
	(0.006)	(0.009)	(0.008)	(0.010)
Time Dummies	Yes	Yes	Yes	Yes
Province Dummies	Yes	Yes	Yes	Yes
Current Nutrition Intake			Yes	Yes
Current Food Consumption				Yes
No. of Observation	39,829	22,411	22,411	15,166

Note: * significant at 10% level, ** significant at 5% level, *** significant at 1% level

1. The dependent variable equals one for hypertensive individuals, and equals zero otherwise.
2. The dependent variable in the first and the second regressions is the current hypertension status, while that in the second and the third regressions is the hypertension status in the next time period.
3. Data come from six rounds of CHNS (1989, 1991, 1997, 2000 and 2004).
4. With the concern of sample attrition bias, for comparison, we have estimated the first and the second regressions for both the full sample and the subsample for which the hypertension information in the next time period is available.
5. Nutrition intake is measured by the intake of energy, fat, protein and carbohydrate.
6. Consumption of the following food items are controlled: pork, beef, mutton, poultry, fish, shrimp/crab, eggs, leafy vegetables, beans, nightshades, legumes, roots/stems, nuts/seeds, mushrooms, melons, grains, and fruits.

Table 3: The Effects of Being Informed of Hypertension Status on Daily Nutrient Intake

	All	Poor	Middle	Rich
	Fat (gram)			
Nonparametric Estimates				
Optimal bandwidth=10	-7.7 ***	0.6	-4.4	-10.2 ***
	(3.3)	(5.3)	(5.9)	(5.3)
Bandwidth=5	-7.6 *	-4.0	-1.2	-13.6 *
	(5.3)	(9.2)	(8.0)	(9.3)
Parametric Estimates				
Without additional controls	-3.5 *	1.2	-2.4	-7.6 **
	(2.2)	(3.7)	(4.3)	(4.1)
With additional controls	-2.9	1.2	-1.5	-6.7 **
	(2.3)	(3.6)	(4.2)	(4.1)
	Energy (kcal)			
Nonparametric Estimates				
Optimal bandwidth=20	-40.0	8.6	-4.2	-90.3
	(45.3)	(72.6)	(84.5)	(86.7)
Bandwidth=10	-43.0	57.7	5.3	-121.9
	(79.8)	(117.9)	(134.5)	(150.4)
Parametric Estimates				
Without additional controls	-57.3	-43.7	-2.1	-114.9 *
	(43.2)	(70.8)	(74.7)	(76.3)
With additional controls	-30.1	1.4	-9.5	-75.2
	(41.8)	(68.7)	(72.5)	(73.8)
	Protein (gram)			
Nonparametric Estimates				
Optimal bandwidth=10	0.6	3.7	-1.1	2.6
	(2.8)	(3.1)	(3.6)	(5.2)
Bandwidth=5	0.8	1.8	3.4	-0.3
	(3.8)	(4.7)	(5.2)	(8.5)
Parametric Estimates				
Without additional controls	-0.9	-0.6	0.5	-1.8
	(1.4)	(2.1)	(2.3)	(2.6)
With additional controls	0.3	0.7	1.4	-0.9
	(1.3)	(2.1)	(2.3)	(2.5)
	Carbohydrates (gram)			
Nonparametric Estimates				
Optimal bandwidth=15	7.1	19.5	2.9	-1.3
	(11.1)	(15.5)	(16.7)	(19.9)
Bandwidth=7.5	6.3	13.1	7.8	-3.8
	(16.5)	(21.8)	(23.9)	(35.2)
Parametric Estimates				
Without additional controls	-1.0	-3.4	-7.8	-1.2
	(2.5)	(7.7)	(12.1)	(13.1)
With additional controls	0.4	1.3	-6.1	2.9
	(7.5)	(11.7)	(12.6)	(13.9)

Note: * significant at 15% level, ** significant at 10% level, *** significant at 5% level

[1] Robust stand errors are reported in parenthesis.

[2] The optimal bandwidth is actually slightly different for different subsamples, as it is data-dependent. For a clearer expression, we only report the averages here.

[3] Quartic polynomials of systolic blood pressure are included in all parametric estimation specifications.

[4] Additional controls include age, sex, education, urban residence, province and year dummies.

Table 4: Summary Statistics of Additional Controls

	Obs.	Mean	S.D.	Min	Max
Age	13,229	45.40	14.50	17	118
Years of Schooling	13,116	6.55	4.17	0	18
Sex (1=Male)	13,252	0.48	0.50	0	1
Urban Residence (1=Yes)	13,252	0.29	0.45	0	1
Rich Tertile (1=Yes)	13,252	0.32	0.47	0	1
Middle Tertile (1=Yes)	13,252	0.34	0.47	0	1
Poor Tertile (1=Yes)	13,252	0.34	0.48	0	1

Table 5: Parametric Estimates of the Effects of Being Informed of Hypertension Status on Daily Nutrient Intake in Full Specifications

	Fat (gram)[1,2]	Energy (kcal)	Protein (gram)	Carb. (gram)
Being Informed * Low Income	-5.040	3.429	-0.451	-4.035
	(5.114)	(74.76)	(3.026)	(15.22)
Being Informed * Middle Income	-6.274	7.067	-0.313	-5.859
	(5.590)	(78.41)	(3.030)	(16.25)
Being Informed * High Income	-9.508 **	27.95	-1.944	4.191
	(5.669)	(94.92)	(3.728)	(20.06)
Being Informed * Education	0.125	4.119	0.514 **	1.880 **
	(0.473)	(6.367)	(0.309)	(1.064)
Being Informed * Sex	0.352	-9.489	1.236	-7.209
	(4.117)	(50.56)	(1.949)	(9.316)
Age	0.014	-9.623 ***	-0.270 ***	-2.278 ***
	(0.108)	(1.163)	(0.053)	(0.270)
Sex (1=male)	6.519 ***	340.4 ***	7.457 ***	51.87 ***
	(2.070)	(22.52)	(1.231)	(4.848)
Years of Schooling	0.965 ***	-0.262	0.266	-2.698 ***
	(0.366)	(4.116)	(0.224)	(0.891)
Income per capita (in 1989 yuans)	0.004 ***	0.018	0.002 *	-0.007 **
	(0.002)	(0.017)	(0.001)	(0.004)
Middle Income (1=yes)	5.006 *	30.19	1.892	-4.318
	(3.109)	(41.00)	(1.898)	(7.919)
High Income (1=yes)	9.264 ***	-1.739	2.482	-11.81
	(4.647)	(60.26)	(3.065)	(13.94)
Urban (1=yes)	3.478	-80.60 **	2.433	-23.68 ***
	(3.511)	(48.73)	(1.802)	(9.153)
Year Dummy: 2000	-3.076	-46.46	1.327	-2.908
	(2.529)	(39.04)	(1.675)	(7.004)
Constant	59.33 ***	2562.2 ***	66.74 ***	456.9 ***
	(7.581)	(91.64)	(4.149)	(21.22)
No. of Observation[3]	2477	5866	2477	3351
R-squared	0.06	0.09	0.1	0.11

Note: * significant at 15% level, ** significant at 10% level, *** significant at 5% level

[1] Quartic polynomials of systolic blood pressure are included in all specifications.

[2] Robust stand errors are reported in parenthesis.

[3] Sample size for each regression is determined by the optimal bandwidth estimated by LLR.

References:

Appel, L.J., Brands, M.W., Daniels, S.R., Karanja, N., Elmer, P.J., Sacks, F.M., 2006. Dietary Approaches to Prevent and Treat Hypertension A Scientific Statement From the American Heart Association. *Hypertension* 47, 296-308.

Angrist, J., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114(2), 533-537.

Barger, S.D., Muldoon, M.F., 2006. Hypertension labelling was associated with poorer self-rated health in the Third US National Health and Nutrition Examination Survey. *Journal of Human Hypertension* 20, 117-123.

Black, S., 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114(2), 577-599.

Brown, D.J., Schrader, L.F., 1990. Cholesterol information and shell eggs consumption. *American Journal of Agricultural Economics* 72, 548-555.

Bryant, C.D., 2003. *Handbook of Death and Dying*, the Sage Publication.

Case, A., Roux, I.L., Menendez, A., 2004. Medical compliance and income-health gradients. *American Economic Review*, 94(2): 331–335.

Chern, W.S., Loehman, E.T., Yen, S.T., 1995. Information, health risk beliefs, and the demand for fats and oils. *The Review of Economics and Statistics* (77), 555-564.

Chern, W.S., Rickertsen, K., 2003. *Health, Nutrition and Food Demand*, MA: CABI Publishing.

Crutchfield, S., Kuchler, F., Variyam, J.N., 2001. The economic benefits of nutrition labeling: A case study for fresh meat and poultry products. *Journal of Consumer Policy* 24(2), 183-207.

Deaton, A., 2003. Health, inequality, and economic development. *Journal of Economic Literature* 41(1), 113-158.

Fuchs, V., 2004. Reflections on the socio-economic correlates of Health. *Journal of Health* 23, 653-661.

Gaziano, T.A., Bitton, A., Anand, S., Abrahams-Gessel, S., Murphy, A., 2010. Growing epidemic of coronary heart disease in low- and middle-income countries. *Current Problems in Cardiology* 35, 72-115.

Gersh, B.J., Sliwa, K., Mayosi, B.M., Yusuf, S., 2010. The epidemic of cardiovascular disease in the developing world: global implications. *European Heart Journal* 31(6), 642-648.

Grossman, M., 1972. On the concept of health capital and the demand for health. *Journal of Political Economy* 80(2), 223-255.

Grossman, M., 2006. Education and nonmarket outcomes. In: Hanushek, E., Welch, F. (Eds.), *Handbook of the Economics of Education* 2. North-Holland, Amsterdam.

Hahn, J., Todd, P., Klaauw, W. V., 2001. Regression discontinuity. *Econometrica* 69, 201-209.

Haynes, R.B., Sackett, D.L., Taylor, D.W., Gibson, E.S., Johnson, A.L., 1978. Increased absenteeism from work after detection and labeling of hypertensive patients. *New England Journal of Medicine* 299, 741-744.

Imbens, G., Kalyanaraman, K., 2009. Optimal bandwidth choice for the regression discontinuity estimator. NBER Working Paper 14726.

Imbens, G., Lemieux, T., 2008. Regression discontinuity designs : A guide to practice. *Journal of Econometrics* 142(2), 615-635.

Kaabia, B.M., Angulo, A.M., Gil, J.M., 2001. Health information and the demand for meat in Spain. *European Review of Agricultural Economics* 28(4), 499-517.

Kannel, W.B., 2000. Elevated systolic blood pressure as a cardiovascular risk factor. *American Journal of Cardiology* 85(2), 251-5.

Kearney, M.P., Whelton, M., Reynolds, K., Muntner, P., Whelton, K.P., He, J., 2005. Global burden of hypertension: analysis of worldwide data. *Lancet* 365(9455), 217-223.

Kim, S.W., Chern, S., 1999. Alternative measures of health information and demand for fats and oils in Japan. *Journal of Consumer Affair* 33, 92-109.

Koch, E., Romero, T., Romero, C.X., Akel, C., Manrı´quez, L., Paredes, M., Roma´n, C., Taylor, A., Vargas, M., Kirschbaum, A, 2010. Impact of education, income and chronic disease risk factors on mortality of adults: Does ‘a pauper-rich paradox’ exist in Latin American societies? *Public Health* 124, 39-48.

Lawrence, W.F., Fryback, D.G., Martin, P.A., Klein, R., Klein, B.E., 1996. Health status and hypertension: a population based study. *Journal of Clinical Epidemiology* 49, 1239-1245.

Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281-355.

Lee, D.S., 2004. Do voters affect or elect policies? Evidence from the U.S. House. *Quarterly Journal of Economics* 119, 807-859.

Lugwig, J., Miller, D.L., 2007. Does Head Start improve children’s life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics* 122, 159-208.

Martin, H.D., Mader, T.L., Pedersen, M., 1994. Influencing diet and health through project LEAN. *Journal of Nutrition Education* 26(4), 191-194.

Melamed, S., Froom, P., Green, M.S., 1997. Hypertension and sickness absence: the role of perceived symptoms. *Journal of Behavior Medicine* 20, 473-487.

Milne, B., Logan, G.A., Flanagan, P., 1985. Alterations in health perception and lifestyle in treated hypertensives. *Journal of Chronic Diseases* 38(1), 37-45.

Neutel, C.I., Campbell, N.R.C., 2008. Changes in lifestyle after hypertension diagnosis in Canada. *Canadian Journal of Cardiology* 24(3),199-204.

Popkin, B., 2008. Will China’s nutrition transition overwhelm its health care system and slow economic growh? *Health Affairs* 27(4), 1064-1076.

Pritchett, L., Summers, L.H., 1996. Wealthier is Healthier. *Journal of Human Resources* 31(4), 841-868.

Roosen, J., Marette, S., Blanchemanche, S., Verger, P., 2009. Does health information matter for modifying consumption? A field experiment measuring the impact of risk information on fish consumption. *Review of Agricultural Economics* 31(1), 2-20.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-Discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* 51, 309-317.

Van de Poel, E., O’Donnell, O., Van Doorslaer, E., 2009. Urbanization and the spread of diseases of affluence in China. *Economics and Human Biology* 7, 2000-216.

Van der Klaaw, W., 2002. Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* 43(4), 1249-1287.

Vasan, R.S., Beiser, A., Seshadri, S., et al., 2002. Residual lifetime risk for developing hypertension in middle-aged women and men: The Framingham Heart Study. *Journal of American Medical Association* 287, 1003-1010.

Wang, L., Kong, L., Wu, F., Bai, Y., Burton, R., 2005. Preventing chronic diseases in China. *Lancet* 366(9499), 1821-1824.

Wolf-Maier, K., Cooper, R.S., Kramer, H., Banegas, J.R., Giampaoli, S., Joffres, M.R., Poulter, N., Primatesta, P., Stegmayr, B., Thamm, M., 2004. Hypertension treatment and control in five European countries, Cananda and the United States. *Hypertension* 43, 10-17.

World Health Organization, 2005. *Preventing Chronic Diseases: a Vital Investment: WHO Global Report*.

Yang, Y., 2002. *Food Nutrient Table in China*, Beijing: Peking University Medical Press.