Journal of Health Economics 32 (2013) 367–385

Contents lists available at SciVerse ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption☆

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

[a] Waseda Institute for Advanced Study, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan
[b] Faculty of Liberal Arts, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554, Japan
[c] Department of Applied Economics, University of Minnesota, 1994 Buford Avenue 337A, St. Paul 55108, USA

ARTICLE INFO

Article history:
Received 3 November 2011
Received in revised form
21 November 2012
Accepted 22 November 2012
Available online xxx

JEL classification:
D12
I12
Q11

Keywords:
Health information
Hypertension
Diet
Regression discontinuity
China

ABSTRACT

We examine the role of information in understanding the differential effects of income on the demand for health. In the health capital framework of Grossman (*JPE*, 1972), we derive the testable hypotheses that individuals adjust their diet in a healthier direction upon receiving negative health information, and that the effect is greater for richer individuals. Based on unique Chinese longitudinal data and a regression discontinuity design that exploits the exogenous cutoff of systolic blood pressure in the diagnosis of hypertension, we find that, upon receiving hypertension diagnosis, individuals reduce fat intake significantly, and richer individuals reduce more. Our results also indicate that among the rich, hypertension diagnosis is more effective for individuals with lower education.

1. Introduction

Economists have long been interested in the causal relationships between income and health. A large number of empirical studies have found a positive income-health gradient, which suggests that health is a normal good (e.g. Wilkinson, 1986; Pritchett and Summers, 1996; Deaton and Paxson, 2001; Deaton, 2003). Yet, more recently, some studies have also substantiated high prevalence rates of chronic health conditions such as hypertension, cardiovascular diseases and cancer among high-income groups in some developing countries (Case et al., 2004; Van de Poel et al., 2009; Gaziano et al., 2010; Gersh et al., 2010; Koch et al., 2010). This study attempts to reconcile this puzzle: How could chronic health conditions be positively associated with income while the demand for health generally increases with income and wealth?

Chronic diseases are well known for their latency. Individuals need to make long-term and persistent health investments to prevent the development of chronic health conditions, yet have difficulty observing their own health status or the effects of their health investments on it. Given this nature of chronic diseases, higher incomes may not necessarily result in better health *outcomes* under certain economic conditions. On one hand, individuals with higher incomes may demand better health. Yet, on the other hand, these richer individuals can also 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. Hence, individuals would need adequate health information to guide their daily consumption choices in order for an increased demand for health due to higher incomes to result in better health outcomes. In countries undergoing rapid economic transition, incomes may grow faster than the availability and acceptance of reliable health information by consumers, which in turn depends on their innate demand for such information, both at the individual

☆ 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. The authors gratefully acknowledge the helpful comments from the editor and two anonymous referees, Amitabh Chandra, Michael Grossman, Seritsu Ogura, Sayaka Nakamura, Emilia Simeonova, and other seminar participants at the Australian Conference of Economists, the Japanese Health Economics Association meeting, the Japanese Economic Association meeting, and the American Society of Health Economists meeting.

* Corresponding author. Tel.: +81 3 5286 2928; fax: +81 3 5286 2470.
E-mail addresses: zhaox057@umn.edu, konishi-moe@aoni.waseda.jp (M. Zhao).

0167-6296/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhealeco.2012.11.007

and aggregate levels. Thus the key to disentangling the channels by which income affects chronic health outcomes requires an understanding of the role of health information.

This study investigates the causal relationships between incomes, diet, and chronic health outcomes, focusing on the role of health information and carefully accounting for endogeneity in each of these variables. Though a large number of studies have attempted to identify the causal effect of health information on consumers' choices, their results have often met with substantial empirical challenges (see Section 2 for more thorough discussions on this point). First, 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. Second, 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 a regression discontinuity (RD) approach using unique Chinese panel data, which allow us to examine 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 first begin by presenting our theoretical framework that builds upon Grossman's (1972) health capital model. Extending Grossman's framework to account for imperfect information about ones' health status, we derive testable hypotheses that individuals adjust their diet in a healthier direction upon receiving negative health information, and that such dietary adjustments are greater for individuals with higher incomes. These hypotheses are then operationalized in our 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 three separate rounds from 1997 to 2004. We 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 three macronutrients (fat, protein and carbohydrates) and energy, 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 pressure 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 g per day 3–4 years after the blood pressure 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 a good (chronic) health condition 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.

Lastly, our results for the full sample indicate that there is no general differential effect by education level. However, among relatively wealthy individuals the estimated reduction in fat intake is greater for those with primary education than those with higher education. This finding is important, as it helps to identify the sources of the effect of hypertension diagnosis and reveals how income and education may interact with each other. On one hand, one may argue that the differential effects of treatment could be attributed to differences in education rather than in income, because more educated individuals may be more health-conscious or more efficient in adopting healthier dietary habits and because income is positively correlated with education. On the other hand, more educated individuals may be more informed of their health status, which would make hypertension diagnosis more informative to less educated individuals. Most importantly, our findings that dietary adjustments differ by income, but not by education, and that among the rich individuals, the adjustments are greater for those with lower education levels suggest that hypertension diagnosis is most useful to rich individuals with lower educational attainment.

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.

[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 90 mmHg.

[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).

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 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 exploring the unique Chinese longitudinal data, 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; 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 blood pressure readings.[5]

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

[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).

[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.

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 offers 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).[7] We have also excluded all the observations aged less than 18 years old. In 1997, 8688 individuals were surveyed, 72.3% (6283) of whom were re-interviewed in 2000. In 2000, an additional 3516 individuals were included in the survey to maintain the original sample size, resulting in a total sample of 9799 individuals in 2000. Of the individuals in the 2000 sample, approximately 71.1% (6969) 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 (6283 from 1997 and 6969 from 2000) who were interviewed twice in continuous surveys.

3.1. 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 carbohydrates, fat and protein as well as total energy intake from these macronutrients. Ideally, we also would like to examine salt consumption, as salt is also one of the most important risk factors of hypertension. Unfortunately, however, no reliable data on salt consumption are available in the CHNS, because the survey participants had difficulty recalling the exact amount of each seasoning consumed each day (only 1% of the total sample recorded positive amounts of salt consumption).

The food demand patterns in China have changed considerably during the past two decades. Table 1 gives the means of 3-day per capita consumption of 15 major food categories from 1989 to 2004. Consumption of pork has been the major source of meat for Chinese people, while consumption of beef and poultry are relatively small, although they have increased rapidly over that period. Fish consumption has been relatively stable, around 1.2–1.5 g per 3 days.

Table 1
Trends in 3-day food consumption.

	1989	1991	1997	2000	2004
Animal foods (g)					
Pork	2.18	2.44	2.45	3.01	2.94
Beef	0.10	0.13	0.26	0.28	0.30
Poultry	0.26	0.27	0.58	0.61	0.68
Fish	1.19	1.13	1.40	1.46	1.53
Shrimp and crab	0.06	0.07	0.06	0.12	0.21
Egg products	0.79	1.10	1.62	1.77	1.78
Non-animal foods (g)					
Grains	29.55	29.88	26.33	23.97	24.74
Leafy vegetables	11.22	9.10	9.52	8.93	8.15
Root and stems	6.42	4.73	4.72	4.34	4.78
Legumes	2.58	2.31	2.48	2.87	3.28
Fresh beans	1.81	2.09	1.88	2.69	2.94
Nightshades	0.96	1.46	1.82	2.38	2.86
Fruits	1.38	0.78	0.65	1.09	1.95
Melons	0.76	1.34	2.05	1.76	2.68
Mushrooms	0.14	0.14	0.28	0.24	0.23

Source: China Health and Nutrition Survey, 1989–2004.

Consumption of seafood such as shrimp and crabs has also risen rapidly since 1997. For non-animal products, grains, leafy vegetables and roots and stems are the three types of traditional foodstuffs with the largest quantities, and all of them have been declining over time. On the other hand, the consumption of other relatively minor food items such as beans, melons and nightshades (e.g. eggplants, tomatoes, chili peppers, etc.) has increased steadily, as rising incomes allow people to afford more expensive varieties of food. Coinciding with the trend in seafood consumption, the demand for fruits also rose quickly after 1997.

Aggregating food consumption patterns into the three macronutrients as well as the total energy 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 (Fig. 1a). The decline is driven mainly by a sharp drop in protein intake between 1991 and 1997 (Fig. 1c) and a steady decrease in carbohydrate intake (Fig. 1d). 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. Fig. 1b 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 g more of fat per day than do the poor.

3.2. 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 survey. The survey personnel then informed people of the results of their physical examinations. We use the latter since

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

[7] We also have another reason to use only the most recent three rounds of data. The data on individual physical examination are missing for almost half of the sample in 1993, resulting in a sample size of only 4480 individuals. Because we could not rule out the possibility that the data were missing for some systematic reasons, we decided not to include the 1993 data, since using those data might lead to serious sample selection bias.

[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: Ḧas a doctor ever told you that you suffer from high blood pressure?¨

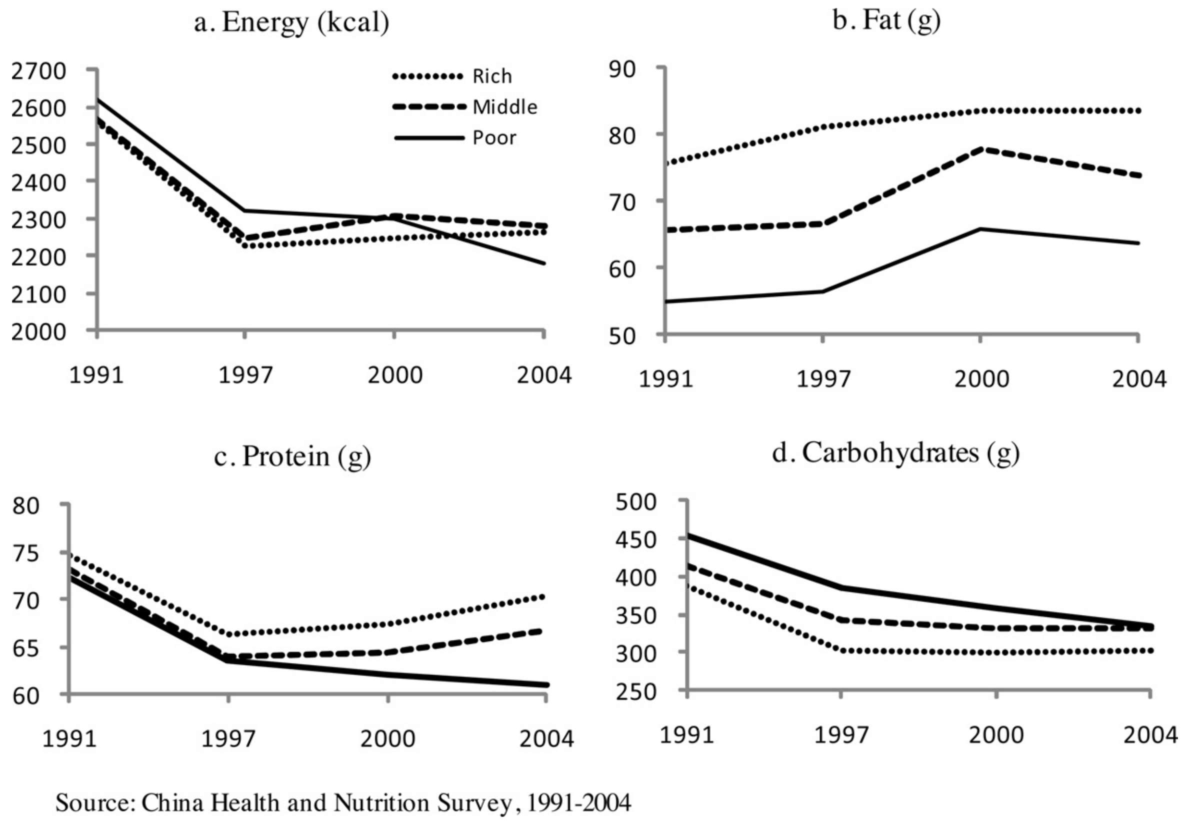

Fig. 1. Trends in daily nutrient intake.

it is likely to be more accurate (see also our discussion on a "fuzzy" RD in Section 6.4).

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 2). The prevalence is higher among the rich than among the poor in all years, and the differences are statistically significant.[10] As shown in column (c), approximately three quarters of survey respondents who were found to have hypertension had never been diagnosed with hypertension by a medical doctor before the survey, and this is the case more for the poor and middle-income groups than for the rich.[11] The percentages of people who reported ever having been diagnosed with hypertension, and currently taking anti-hypertension drugs, regardless whether they were currently hypertensive, are reported in columns (d) and (e). That is, some people included in columns (d) and (e) had normal blood pressure levels in the CHNS survey, though they had been diagnosed with hypertension before. These people might have taken anti-hypertensive measures to keep their blood pressure under control, and the percentages of those falling in the latter case are shown in column (e). It seems that the percentage of those who were currently 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 (refer to column (f)). Though Table 2 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. To identify the *causal* effects of health information and incomes, we rely on the regression discontinuity design, as explained below.

Before we explore the effect of hypertension diagnosis, it is worthwhile examining the possible reasons why incomes could be negatively correlated with chronic health conditions in the observational data. To do this, we estimated both fixed-effects (FE) and random-effects (RE) logit models of the impact of various socioeconomic variables on hypertension status.[12] The first two columns of Table 3 use current hypertension status as the dependent variable for the full sample (the results of the FE logit model are presented in the first column and those of the RE logit model in the second column). We ran the same regressions in the third and fourth columns for the subsample for which hypertension status in the next wave is not available to check for any possible sample selection bias in comparison to the fifth and sixth regressions. The last four columns of Table 3 use hypertension status in the next survey round as the dependent variable, with different sets of covariates, controlling for nutrition intake (the fifth and sixth columns) and food consumption (the seventh and eighth columns) in the base year. In general, the FE estimates and the RE estimates

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

[12] We chose to estimate a logit model instead of a probit model, as we are concerned about the well-documented incidental parameters problem of a fixed-effects probit model.

Table 2
Hypertension status and awareness, age 18+.

Year	Income group	(a) Obs.	(b) Among (a), % of the hypertensive	(c) Among (b), % ever diagnosed with hypertension by a doctor	(d) Among (a), % ever diagnosed with hypertension by a doctor	(e) Among (d), % taking anti-hypertension drugs	(f) Among (e), % keeping condition controlled
1991	Poor	3007	17.5	20.9	4.6	62.9	21.9
	Middle	3039	21.0	31.1	7.2	59.9	10.8
	Rich	2675	18.7	30.4	6.8	60.2	21.9
	Total	8721	19.1	27.7	6.2	60.6	19.3
1997	Poor	2940	23.9	14.4	4.6	54.2	20.3
	Middle	2909	24.0	19.5	5.8	68.8	18.0
	Rich	2814	26.2	25.2	8.1	67.5	19.7
	Total	8663	24.7	19.8	6.1	67.5	19.3
2000	Poor	3244	25.1	23.2	7.1	64.5	16.7
	Middle	3194	22.3	27.2	8.4	67.8	19.5
	Rich	3120	26.4	33.5	12.4	76.2	27.4
	Total	9558	24.6	28.1	9.3	70.7	22.7
2004	Poor	2935	25.4	25.5	8.8	67.9	24.3
	Middle	3104	27.6	32.0	11.9	67.1	22.2
	Rich	3148	27.3	34.6	14.0	84.4	34.4
	Total	9187	26.8	30.9	11.6	74.4	28.4

Source: China Health and Nutrition Survey, 1989–2004.
Note: All the percentages are weighted by age.
Column (a): the number of observations in each subsample. *Column (b)*: the percentage of people in each subsample whose systolic blood pressure is above the cutoff of 140 mmHg. *Column (c)*: the percentage of people who reported to have everbeen diagnosed with hypertension by a medical doctor before the survey interview among those who have high blood pressure as defined in (b). *Column (d)*: the percentage of people in each subsample who reported to have everbeen diagnosed with hypertension by a medical doctor before the survey interview. *Column (e)*: the percentage of people who are currently taking anti-hypertension drugs among those who have been diagnosed with hypertension as defined in (d). *Column (f)*: the percentage of people who preported to be currently taking anti-hypertension drugs but having normal blood pressure among those as defined in (e).

Table 3
Logit regressions of hypertension status on socio-economic variables.

	(1) Current hypertension status		(1) Current hypertension status		(3) Future hypertension status		(4) Future hypertension status	
	FE	RE	FE	RE	FE	RE	FE	RE
Middle income (1 = yes)	0.120**	0.116**	0.158*	0.155**	−0.041	−0.001	−0.103	−0.044
	(0.056)	(0.046)	(0.100)	(0.065)	(0.077)	(0.060)	(0.098)	(0.074)
High income (1 = yes)	0.103*	0.114**	0.230**	0.212***	0.023	0.095	0.095	0.016
	(0.064)	(0.050)	(0.121)	(0.070)	(0.090)	(0.066)	(0.113)	(0.080)
Age	0.118***	0.138***	0.215***	0.142***	0.057*	0.134***	0.078*	0.136***
	(0.008)	(0.008)	(0.048)	(0.013)	(0.033)	(0.012)	(0.043)	(0.016)
Age-squared	−0.001***	−0.001***	−0.001***	−0.001***	−0.001*	−0.001***	−0.001*	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age > 60 (1 = yes)	0.113	0.280***	0.165	0.312***	0.103	0.107	0.196	0.027
	(0.076)	(0.076)	(0.113)	(0.113)	(0.152)	(0.109)	(0.194)	(0.139)
BMI	0.005***	0.001***	0.002***	0.003***	−0.001	0.000	−0.004	0.000
	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)
Sex (1 = male)		0.548***		0.482***		0.495***		0.524***
		(0.047)		(0.064)		(0.062)		(0.074)
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	Yes	Yes	Yes	Yes
Province dummies		Yes		Yes		Yes		Yes
Current nutrition intake					Yes		Yes	
Current food consumption							Yes	
Sample	Full	Full	Partial	Partial	Partial	Partial	Partial	Partial
Number of observation	10,776	39,829	5412	22,411	5412	22,411	3766	15,166

Notes: (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 five 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.

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

are very similar, with the latter being more efficient due to a larger sample size. The results in columns (1)–(4) confirm that the rich generally are more likely to be hypertensive than the poor. Yet, after nutrition intake and food consumption in the base year were controlled for, as shown in the columns (5)–(8), the estimated effects of income dummies turn completely insignificant. This finding is consistent with our explanation about the observed puzzle – i.e. the negative gradient between income and hypertension exists, but is mainly intermediated by food consumption and/or nutrition intake.

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

$$\tilde{H}_0 = H_0 + \varepsilon_0. \tag{2}$$

where $\tilde{H}_0$ is perceived health status. Note that some individuals may have accurate knowledge of their initial health status, so that $\varepsilon_0 = 0$, while others either overestimate or underestimate their initial health status ($\varepsilon_0 > 0$ or $\varepsilon_0 < 0$). Assume that the consumer makes health investments over time, observing her initial health capital with errors but knowing the health production function (1).[14] One important implication of this modeling strategy is that, because her true and perceived health capitals evolve according to (1), she consistently misperceives her health capital over time, with an error evolving according to[15]:

$$\varepsilon_{t+1} = (1 - \delta_t)\varepsilon_t. \tag{3}$$

[13] 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

[14] 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.

[15] To see this, observe that, for any time period t, $\tilde{H}_{t+1} = I(F_t, N_t) + (1 - \delta_t)\tilde{H}_t = I(F_t, N_t) + (1 - \delta_t)(H_t + \varepsilon_t) = H_{t+1} + (1 - \delta_t)\varepsilon_t$.
We posit that the error in health perception is most closely related to innate personal characters (e.g. optimistic vs. pessimistic persons) so that the consumer receives this error in period 0 and carries over her lifetime according to (3), unless she experiences exogenous information shocks such as a chronic health condition or a medical checkup.

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 respect to investment in health capital.

Importantly, if a consumer realizes in some period τ that her true health capital is lower than she thought it was (due to $\tilde{H}^*_\tau$, being higher than H_τ), she must further increase her health investment to raise her future health capital to the new optimum given her actual health capital today. That is, because she had chosen $(\tilde{F}^*_t, \tilde{N}^*_t)$ up to period $\tau > t$ rationally expecting their impact on her future health capital $\tilde{H}^*_t$ for $t \geq \tau$, and because her health capital evolves according to (1), she must raise I_τ to achieve the *new* optimal path $H^*_{\tau+1}$ conditional on observing $H_\tau < \tilde{H}^*_\tau$.[16] Provided that $I_F < 0$ and $I_N > 0$, it follows that: $F^*_t \leq \tilde{F}^*_t$ and $N^*_t \geq \tilde{N}^*_t$ for $t \geq \tau$, iff $H_\tau < \tilde{H}^*_\tau$, where $(\tilde{F}^*, \tilde{N}^*)$ and (F^*, N^*) indicate the consumer's optimal consumption path along her perceived health and her true health conditional on observing H_τ, 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 $h_i = 1$ 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: $\tilde{b}_i = b_i + \varepsilon_i$.

An individual whose belief $\tilde{b}_i$ lies below c but whose true blood pressure b_i lies above c gets a negative information shock: $\tilde{h}_i - h_i < 0$. Analogously, there are two other cases: $\tilde{h}_i - h_i > 0$ and $\tilde{h}_i - h_i = 0$. Therefore, any individual whose true blood pressure b_i lies above c would get either no information surprise (her $\tilde{b}_i$ also lies above c) or a negative information shock. Assuming that individuals consider hypertension status h (not the blood pressure level 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 $\tilde{h}_i - h_i < 0$. It follows then that we should observe $E[F^* - \tilde{F}^* \leq 0 | b > c]$ and $E[F^* - \tilde{F}^* \geq 0 | b < c]$ where the expectation operator is taken over all i. Hence, we should observe that:

$$\begin{aligned}
&\lim_{b \to c^+} E[F^*(b) - \tilde{F}^*(b)] - \lim_{b \to c^-} E[F^*(b) - \tilde{F}^*(b)] \\
&\quad = \lim_{b \to c^+} E[F^*(b)] - \lim_{b \to c^-} E[F^*(b)] \leq 0.
\end{aligned}$$

[16] More precisely, a sufficient condition for the optimal health investment I^*_τ conditional on H_τ to be higher than that conditional on $\tilde{H}^*_\tau$ when $H_\tau < \tilde{H}^*_\tau$, is that the demand for health capital is inelastic.

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.

A final question is: Can we derive testable hypotheses on consumers' dietary adjustments in terms of education rather than by income? Our answer is "No". As Grossman (1972) points out, more educated consumers may be more productive in health production. His argument also seems valid in this context, because more educated consumers may have better knowledge of what constitutes a healthy diet or may simply be better at (self-)controlling daily diet. Following this line of argument, it is tempting to derive a hypothesis that dietary adjustments upon receiving a health information shock are greater for consumers with higher education levels than for those with lower education levels, conditional on the same true health status. However, such a hypothesis does not account for the possibility that more educated consumers may be more informed about their own health status. In other words, hypertension diagnosis may be more informative to less educated consumers than more educated ones. In the empirical framework described below, we have no means to disentangle these two competing effects of education, and thus we do not derive any testable hypotheses regarding education levels. Indeed, our results indicate that within the wealthiest group, dietary adjustments are greater for the less educated consumers than for the more educated consumers (see Section 6.2).

5. Identification and estimation

5.1. The regression discontinuity design

To operationalize the organizing framework in Section 4, the study exploits the facts that all survey subjects were informed of their blood pressure test results in each survey round and 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).[17] Thus, for simplicity this study focuses only on SBP.

This assignment of hypertension status in the CHNS lends itself well to estimating causal impacts of health information using a 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 ith 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:

$$Y_i = \alpha + \beta T_i + u_i \tag{4}$$

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 $E[u|T] \neq 0$, 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.

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.[18] 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 Eq. (4):

$$Y_i = \alpha + \beta T_i + f(b_i) + \mu_i \tag{5}$$

where $\mu_i = u_i - E[u_i|T_i, b_i]$ with $E[\mu_i|T_i] = 0$. If Eq. (5) is linear in T_i and $f(.)$ can be correctly specified, the parameter b 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 b is consistent yet inefficient; if under-specified, the estimate will be efficient but inconsistent. Quartic polynomials are preferred for their flexibility in practice (Lee, 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:

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

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

$$(\delta_{Yr}, \theta_{Yr}) \equiv arg \min_{\delta,\theta} \sum_{i:b_i \geq c} [Y_i - \delta - \theta(b_i - c)^2]\lambda_i,$$

$$(\delta_{Yl}, \theta_{Yl}) \equiv arg \min_{\delta,\theta} \sum_{i:b_i < c} [Y_i - \delta - \theta(b_i - c)^2]\lambda_i,$$

where $\lambda_i = K[(b_i - c)/h]$ 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).

5.2. 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

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

[18] 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 Section 6.4.

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 Fig. A.1, 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 Fig. A.2. 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.

Table 4
The Effects of being informed of hypertension status on daily nutrient intake.

	All	Poor	Middle	Rich
Fat (g)				
Nonparametric estimates				
Optimal bandwidth = 13	−7.7***	0.6	−4.4	−10.2***
	(3.3)	(5.3)	(5.9)	(5.3)
Bandwidth = 6.5	−7.7**	1.1	−5.1	−12.7**
	(4.4)	(7.7)	(7.4)	(7.4)
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	−6.9**
	(2.3)	(3.6)	(4.2)	(4.1)
Protein (g)				
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 (g)				
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)
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)

Notes: (1) Robust standard errors are reported in parenthesis. (2) The optimal bandwidth is actually sightly different for different subsamples, as it is data-dependent. For a clearer expression, we only report the averages here. (3) Quartic polynomials of systole blood pressure are included in all parametric estimation specifications. (4) Additional controls include age, sex, education, urban residence, province and year dummies.

* Significant at 15% level.
** Significant at 10% level.
*** Significant at 5% level.

Table 5
The effects of being informed of hypertension status on the use of anti-hypertension drugs (%).

	All	Poor	Middle	Rich
(a) Full sample				
Nonparametric estimates				
Optimal bandwidth = 5	1.8	5.7***	2.9***	1.7***
	(0.2)	(0.5)	(0.3)	(0.3)
Parametric estimates				
With additional controls	5.3***	5.7***	5.3***	4.30***
	(0.3)	(0.6)	(0.5)	(0.5)
Number of observations	12,985	4466	4335	4184
(b) Excluding self-reported hypertensive individuals in base year[a]				
Nonparametric estimates				
Optimal bandwidth = 5	4.9***	4.3***	5.1***	5.9***
	(0.1)	(0.2)	(0.9)	(0.3)
Parametric estimates				
With additional controls	4.0***	4.2***	4.8***	2.44***
	(0.3)	(0.6)	(0.5)	(0.7)
Number of observations	12,255	4272	4129	3854
(c) Excluding anti-hypertension drug users in base year[b]				
Nonparametric estimates				
Optimal bandwidth = 5	6.1***	4.3***	2.6***	7.6***
	(0.2)	(1.2)	(0.5)	(0.5)
Parametric estimates				
With additional controls	4.1***	4.3***	4.5***	3.10***
	(0.3)	(0.6)	(0.5)	(0.6)
Number of observations	12,476	4342	4196	3938

Notes: (1) Robust standard 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.

[a] Self-reported hypertensive individuals are defined as those who reported to have ever been diagnosed of hypertension by a medical doctor before the survey interview.
[b] Anti-hypertension drug users are defined as those who reported to be currently using anti-hypertension drug.
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

6. Results

Local linear regressions were estimated, with optimal bandwidths, for the four outcome measures of the nutrient intake in Table 4. Though not our primary focus, we also report our results on the use of anti-hypertension drugs in Table 5. 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 6 for the summary statistics of these covariates).

Table 6
Summary statistics for additional control variables.

	Observations	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

Source: China Health and Nutrition Survey, 1997–2004.

6.1. Main results

We begin with a graphical representation of the observed relationship between blood pressure levels and daily nutrition intake. Figs. 2–6 present averages (circles) and local linear smoothers (solid

[19] We have also considered two alternative sets of dependent variables: (a) consumption of specific food groups such as meats, vegetables, fruits, etc.; (b) the share of a component nutrient intake (e.g. fat) in the total energy intake. These results are available upon request. We do not report these results for the following reasons. For the former, estimation with aggregated food categories will not capture subtle changes in food choices as well as the substitution between food items in the same category, which may mask what consumers are attempting to do to change their diet toward a healthier direction. For example, consumers may reduce fats by reducing fat-cut meats or substituting different types of meats without changing the overall quantity of meats. Because the CHNS survey records nutrient contents of different cuts as well as different types of meats, we think that nutrition intake can give us a better and clearer picture of the changes in overall diet patterns. For the second set of dependent variables, it turns out that when a major nutrient intake such as fat increases, the total calorie intake also increases. Thus the relative measure gives us substantially less variation than the absolute measure. For instance, in our data the proportion of calories coming from fat does not differ significantly between hypertensive and non-hypertensive people (29.3% and 28.9%, respectively).

[20] As will be discussed in Section 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.

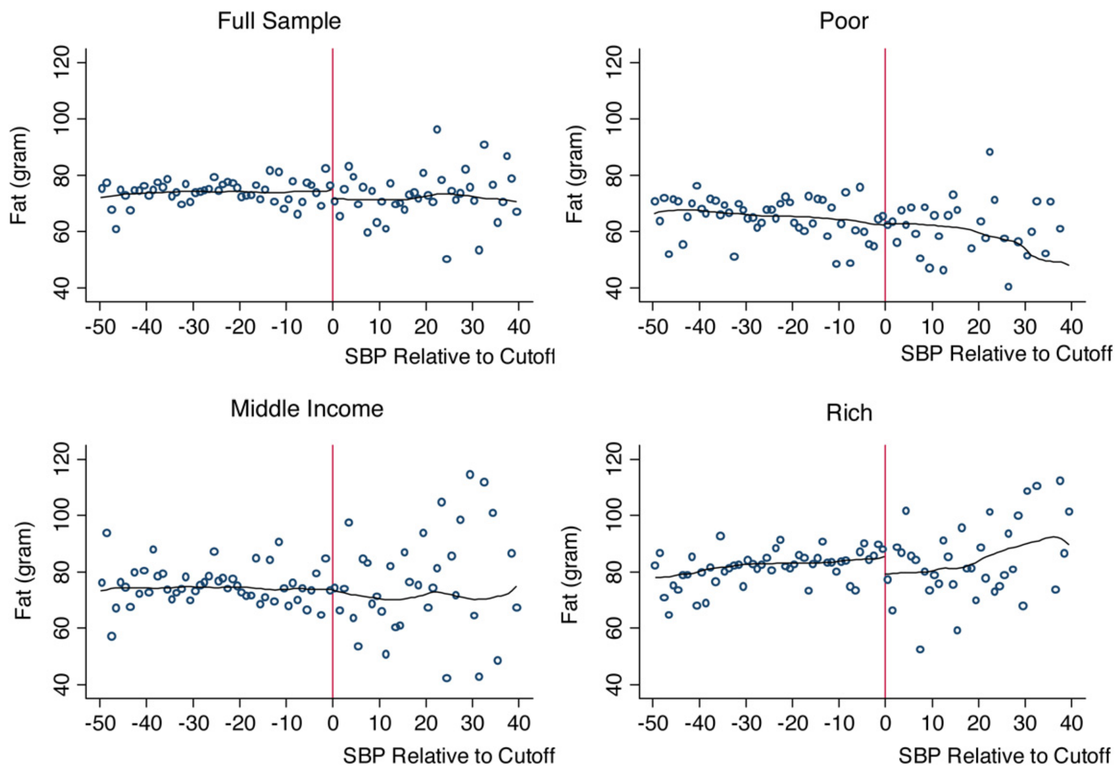

Fig. 2. Blood pressure and daily fat intake by income level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

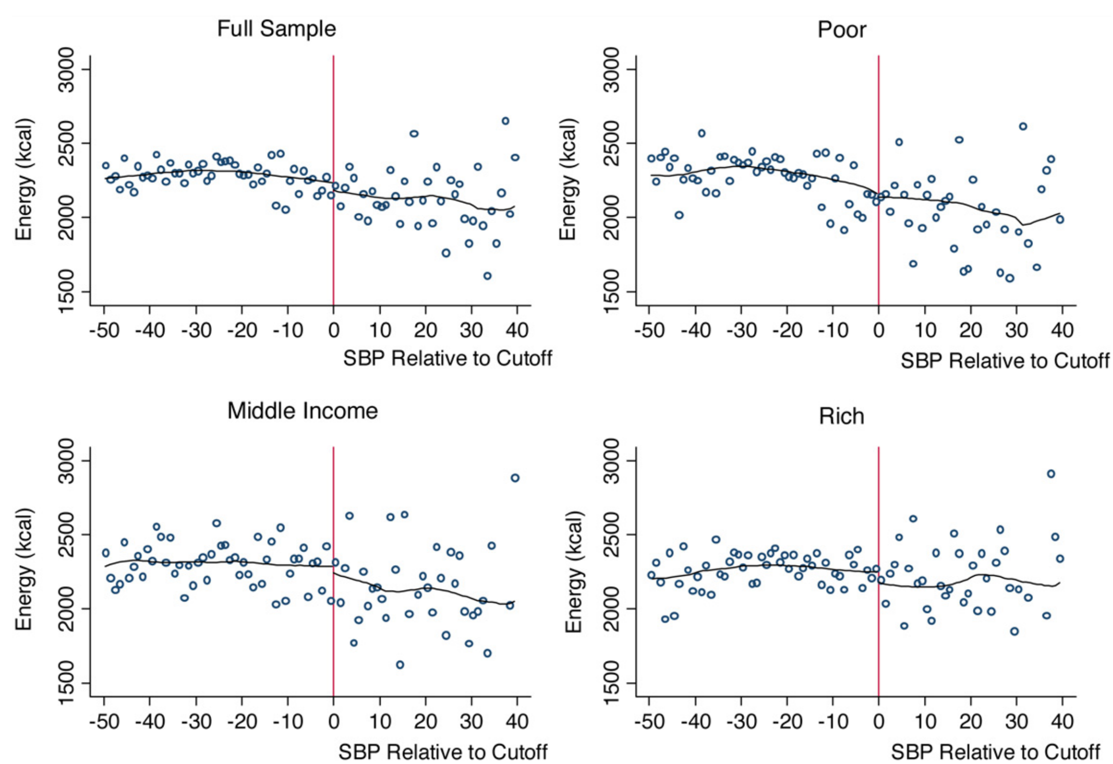

Fig. 3. Blood pressure and daily energy intake by income level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

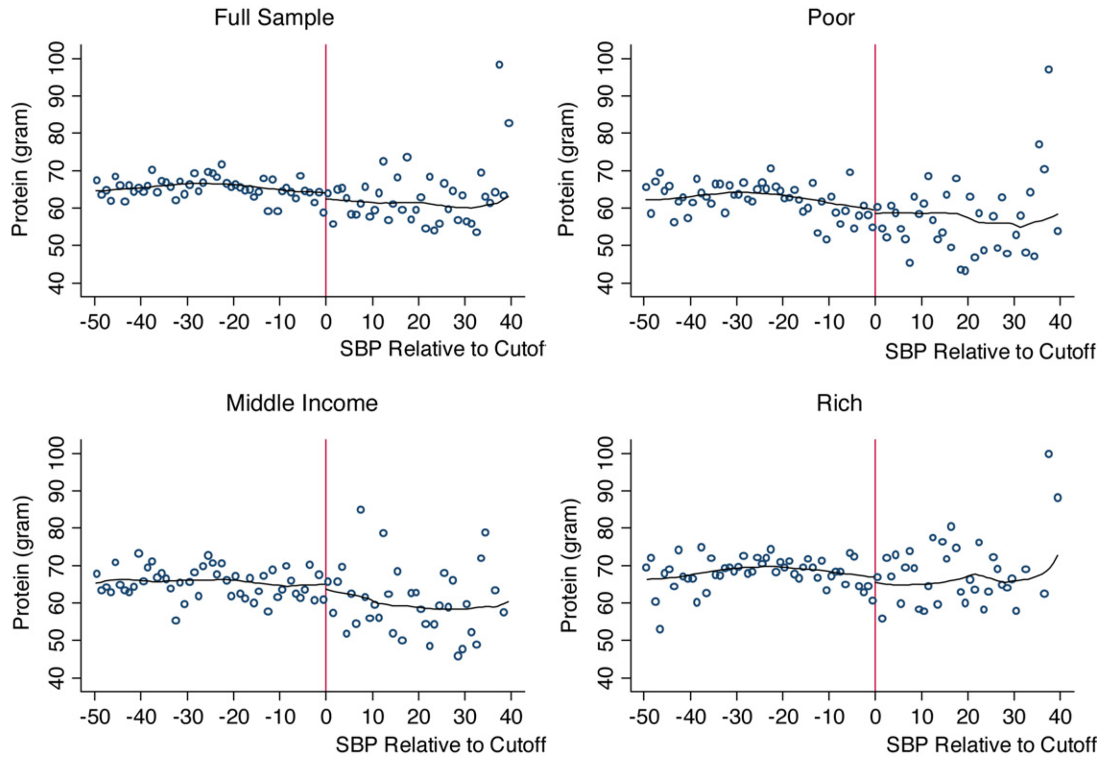

Fig. 4. Blood pressure and daily protein intake by income level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

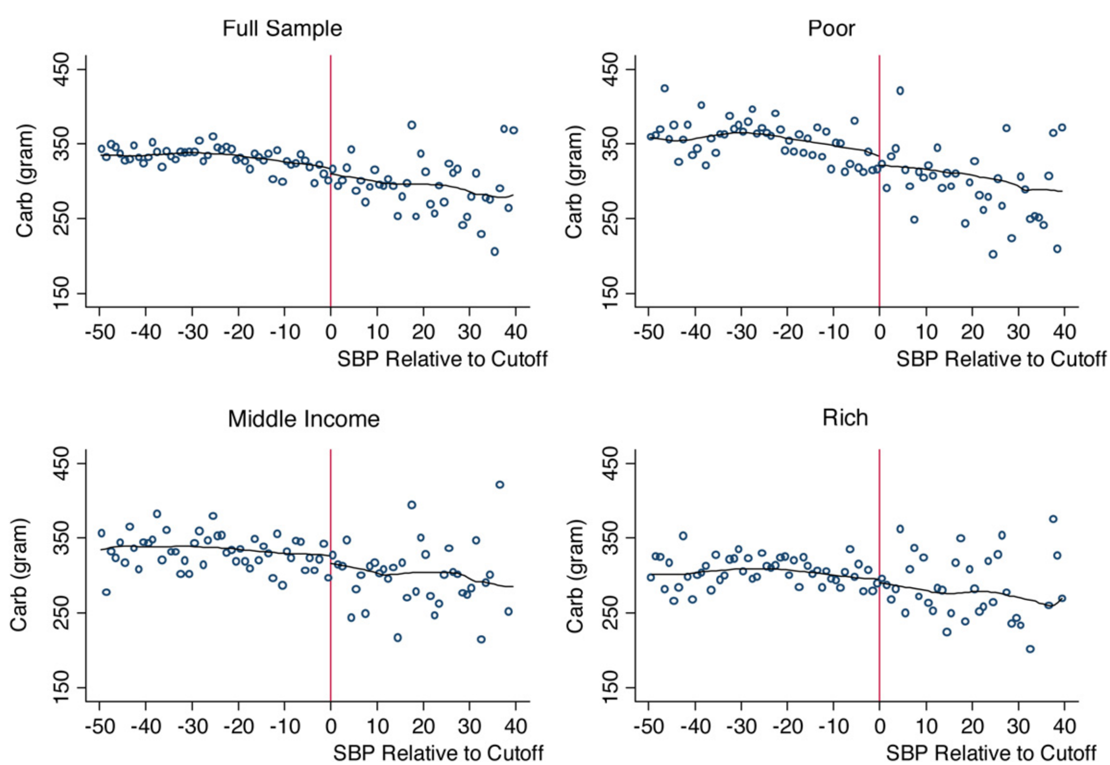

Fig. 5. Blood pressure and daily carbohydrate intake by income level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

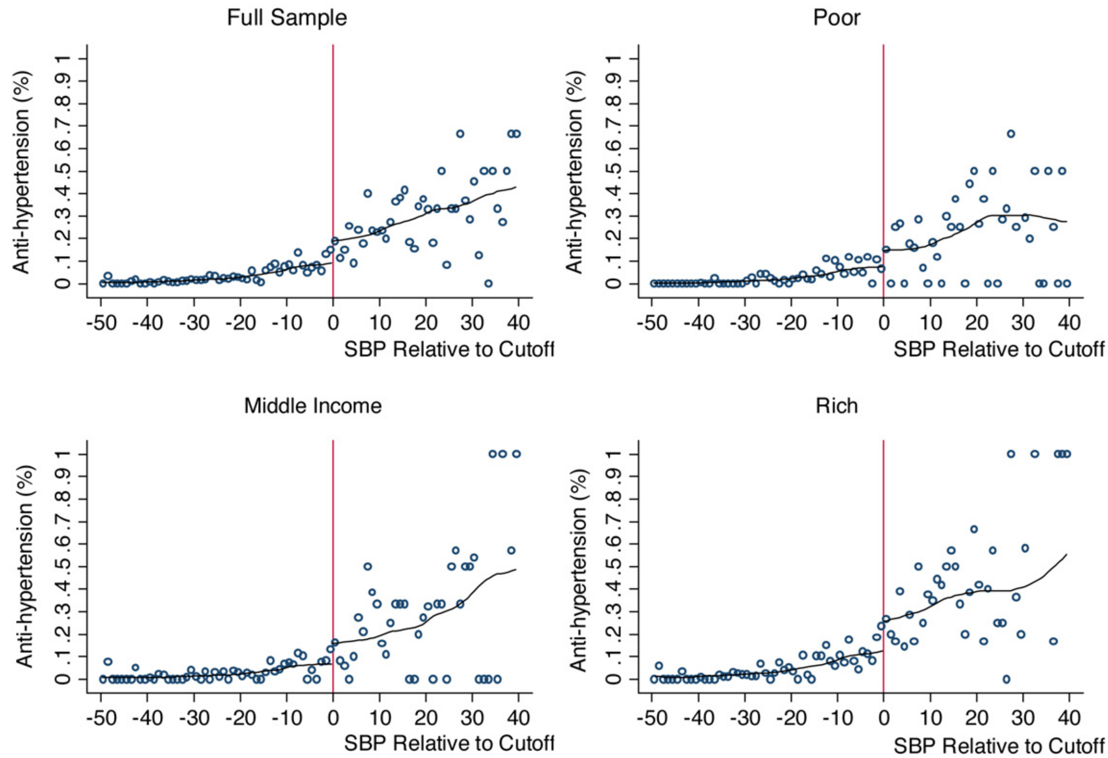

Fig. 6. Blood pressure and use of anti-hypertension drugs by income level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

curves) of the daily intake of four macronutrients (fats, energy, protein and carbohydrates) and the use of anti-hypertension drugs in the next round (3–4 years later), plotted against systolic blood pressure. These figures are shown for the whole sample as well as separately for the three income groups (poor, middle income, rich).

Four features of the relationships shown in Figs. 2–6 are particularly noteworthy. First, there is a clear discontinuity in fat intake for the wealthy group, but not for the poor and middle-income groups. Second, the variance in daily nutritional intake gradually increases to the right of the cutoff. As the cutoff is close to the top tail of the distribution of systolic blood pressure, one often observes increased variance to the right side of the cutoff.[21] Third, there are no apparent marked discontinuities in intakes of the other nutrients, though high variance is still observed to the right of the cutoff. Lastly, there is a sharp discontinuity in anti-hypertension drug use for all income groups. However, the RD estimates on the effect of diagnosis on consumers' use of anti-hypertension drugs may be severely biased. As anti-hypertension drugs are mainly prescribed by doctors in China, in addition to receiving the same diagnosis information, those to the right of the SBP cutoff are likely to have more access to anti-hypertension drugs than those to the left of the cutoff. To the extent that limited access precluded some consumers from using anti-hypertension medication more often to the left than to the right of the cutoff, the identified effect of hypertension diagnosis on the use of anti-hypertension drugs is likely to be biased upward.

Turn now to the discontinuity point estimates presented in Table 4. For each type of nutrient, both nonparametric (top two rows) and parametric estimates (bottom two rows) are presented, for both the whole sample (the first column) as well as for the subsamples of different income groups (the second, third and fourth columns). The nonparametric estimate with the optimal bandwidth indicates that the diagnosis of hypertension induces individuals to reduce their fat intake by 7.7 g 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. As one might expect, reducing the bandwidth reduces efficiency, because the number of observations gets smaller. The discontinuity point estimates across different income groups reveal that the effect of hypertension diagnosis on fat intake is significant only for the rich individuals. Based on the nonparametric estimate with optimal bandwidth, the discontinuity point estimate increases to 10.2 g for wealthy individuals while those for the poor and the middle income groups are smaller and statistically insignificant. The estimated effect of treatment on the whole sample is thus driven mainly by the impact on the wealthy group. These results still hold for nonparametric estimates that use half of the optimal bandwidth and for parametric estimates with or without additional controls, though significance levels slightly decrease. However, the nonparametric estimates are our preferred estimates, as parametric estimates are often biased (Lee and Lemieux, 2010). Thus, these empirical results provide strong support for our hypotheses.

Our results for other nutrients, protein and carbohydrates, and total energy 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

[21] The increased variance would not bias our estimates as long as the conditional variance is continuous at the cutoff (Hahn et al., 2001). Moreover, we use White heteroskedasticity robust standard errors for our statistical inference to account for the variation in the conditional variance.

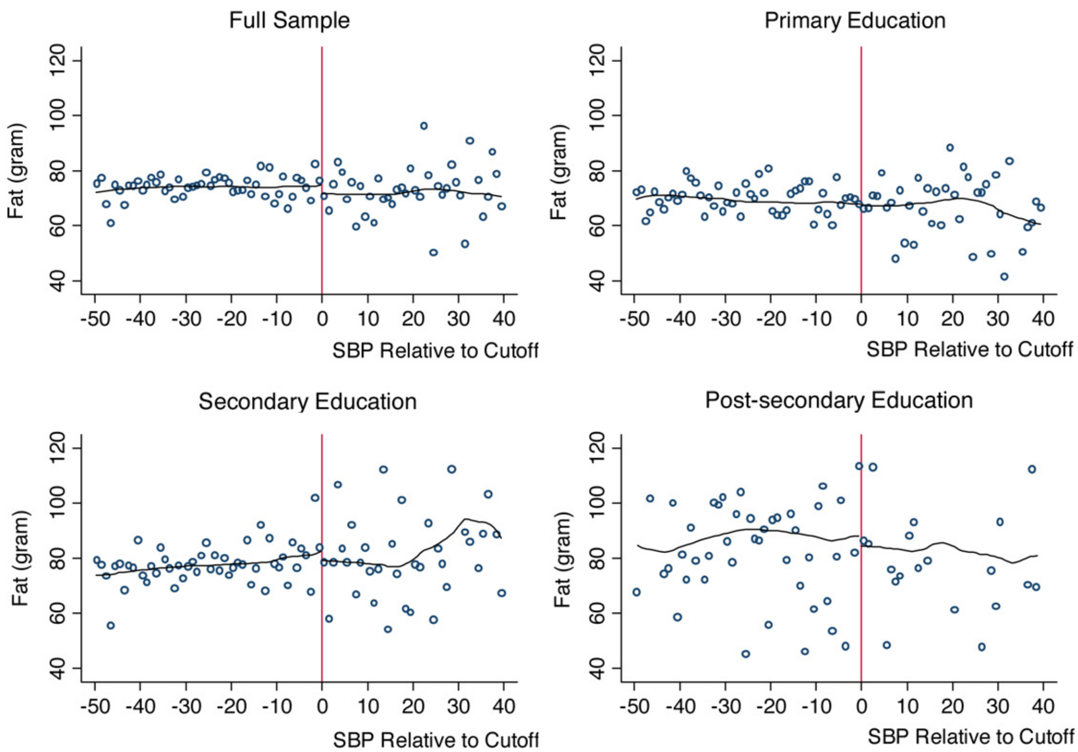

Fig. 7. Blood pressure and daily fat intake by education level. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother (h = 5). Estimates based on individuals aged above 18.

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 (2003), 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.

Table 5 presents discontinuity point estimates for current use of anti-hypertension drugs. Both nonparametric and parametric estimates are significantly positive, indicating that consumers increase anti-hypertension drug use after being informed of their hypertension status. Contrary to the findings on fat intake, the estimated effects in the top panel of Table 5 are greater for the poor. This may occur because a larger fraction of the hypertensive wealthy people may have already been using anti-hypertension drugs, as indicated in Table 2.[22] Another possible reason is that the use of anti-hypertension drugs and the reduction in fat intake may be substitutes for controlling blood pressure, i.e. the wealthy may be choosing to cut fat over taking medications while the poor do the opposite. However, we do not observe any significant relationship between these two, even after controlling for individuals' socioeconomic status and the need for anti-hypertension activities.

To filter out the possible downward bias resulting from the possibility that some of those were already using anti-hypertension drugs in the base years, we have also estimated the effects of hypertension diagnosis using two partial samples: (i) a sample that excludes those who reported that they had been diagnosed with hypertension in the base years, and (ii) partial sample, excluding those who had been using anti-hypertension drugs in the base years. The results are shown, respectively, in panels (b) and (c) of Table 5. Changes in the nonparametric estimates are consistent with our expectation. The nonparametric estimate for the rich increased from 1.7 for the full sample, to 5.9 for the first partial sample, and to 7.6 for the second partial sample. On the other hand, the nonparametric estimate for the poor decreased from 5.7 for the full sample, to 4.3 for the first partial sample, and to 4.3 for the second partial sample. The parametric results are less sensitive to the sample choices, somewhat contrary to our expectation, and are still showing a smaller effect for the rich. This inconsistency might be because parametric estimates are more susceptible to misspecification of the functional relationship than are non-parametric estimates – the former uses observations far away from the cutoff whereas the latter uses only observations close enough to the cutoff (as determined by the optimal bandwidth).

[22] In the data we used for regression analysis, a total of 514 people (approximately 3.8% of the total sample) were found to be taking anti-hypertension drugs in the base years (1997 and 2000). In general, the rich were twice as likely to be taking anti-hypertension drugs as people in other two income groups in the base years. In the next wave, the total number of people who used anti-hypertension drugs increased by 61% to 720:429 as new users of anti-hypertension drugs and 291 as continuing users. Again, the share of drug users among wealthy people is twice as large as those in the other two income groups.

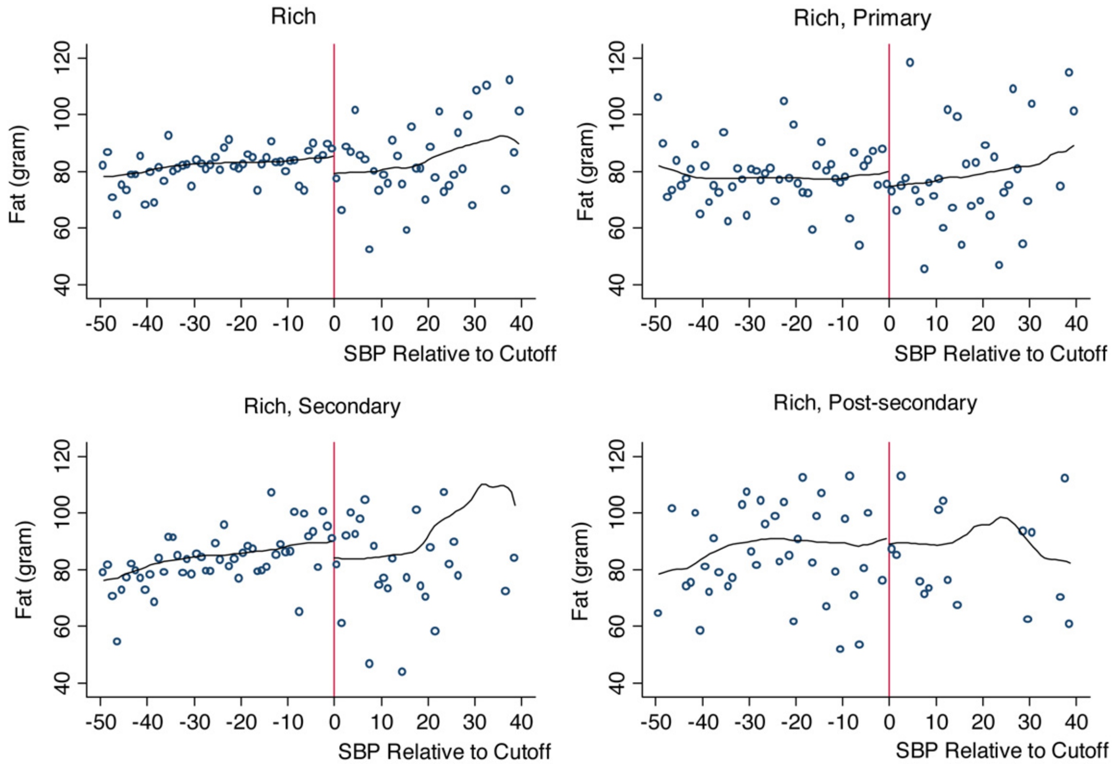

Fig. 8. Blood pressure and daily fat intake by education level among rich individuals. *Note*: open circles are unconditional averages. Solid curve is a local linear smoother ($h = 5$). Estimates based on individuals aged above 18.

6.2. Income vs. education

It is possible that the differential effects of hypertension diagnosis on fat intake are due to education rather than income. More educated consumers may be more health conscious, better at controlling their daily diet, and/or more knowledgeable about hypertension risks. Indeed, Grossman (1972) argues that more educated individuals are more efficient in health production. Because education is also positively associated with income, the observed differential effects by income may be due to education rather than income. To investigate this possibility, graphical representations of the relationship between fat intake and blood pressure are provided by education level (primary, secondary, and higher education) for the whole sample in Fig. 7 and for the wealthy group in Fig. 8.

First, as already shown in Fig. 2, for the full sample there is no marked discontinuity at the cutoff. Moreover, the discontinuity point estimates are insignificant for all education groups. For the higher-education group, there is high variance in fat intake both to the right and to the left of the cutoff. This is mainly due to the relatively small sample of people with high education, who are less than 20% of the total sample. The large variance also suggests the possibility that a subpopulation may exist within the higher-education group of people who are generally health conscious and have a low-fat diet in order to keep their blood pressure low.

Second, within the wealthy group, there is a clear discontinuity at the cutoff for those with primary but not for those with higher education. At first glance, this finding appears in contrast to the theoretical prediction of Grossman's original model that more educated consumers invest more in health capital. We conjecture that this result occurs because less educated consumers are less informed about their own health status, and therefore, the blood pressure information was more informative to those with lower education levels than to those with higher education levels.

Table 7 presents nonparametric and parametric point estimates for rich individuals of different education groups. Based on nonparametric estimation with optimal bandwidth, the point estimates are significant only for those with primary education. Parametric estimates show that these results are generally robust. One concern is the sensitivity of the nonparametric estimates to the choice of bandwidth; the point estimate is insignificant when half of the optimal bandwidth is used. However, this occurs mainly due to the small sample size – because the sample size is already small for estimates restricted to wealthy individuals, choosing half of the optimal bandwidth further reduces the sample size and increases the variance.

Lastly, one caveat with our finding is that our RD approach does not allow us to control for potential endogeneity of income, other than by increasing the number of covariates. If individuals' incomes are correlated with their unobserved characteristics such as risk attitudes and carefulness, the differential effects of hypertension diagnosis by income groups may be biased. In other words, we cannot completely eliminate the possibility that richer individuals respond more to hypertension diagnosis, not because they are richer, but because they are systematically different from other income groups. This concern does not pose an issue if unobservable characteristics are caused by income or uncorrelated with income. To address this concern, we used income information 3–4 years before, which should at least prevent the reverse causality. We also conducted similar nonparametric analysis by controlling for a number of covariates such as sex, age, and urban–rural residence. We found that the estimated effect of hypertension diagnosis on fat intake is robust and does not vary by these additional controls.

Table 7
The effects of being informed of hypertension status on fat intake among rich individuals.

	All	Primary	Secondary	Post-secondary
Nonparametric estimates				
Optimal	−10.2***	−11.5**	−3.0	−10.0
bandwidth = 13	(5.3)	(7.1)	(11.4)	(9.9)
Bandwidth = 6.5	−12.7**	−9.1	−10.9	−14.9
	(7.4)	(9.7)	(15.0)	(11.9)
Parametric estimates				
Without additional	−7.6**	−10.0**	−5.6	2.3
controls	(4.1)	(5.4)	(10.6)	(7.4)
With additional	−6.9**	−9.3**	−4.5	2.3
controls	(4.1)	(5.3)	(10.6)	(7.3)
Number of observations	4149	1581	1258	1279

Notes: (1) Robust standard 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.
* Significant at 15% level.
** Significant at 10% level.
*** Significant at 5% level.

6.3. Discussion on robustness

The above results rely heavily on local linear regression estimation, so it is critical to examine the sensitivity of the estimates in Table 4 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 6.1. Either an increase or a decrease in the bandwidth does not lead to significantly different results.

Secondly, 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 Eq. (5) may result in the violation of the RD assumptions 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 8). To avoid inappropriate approximation of function f in Eq. (5), these regressions use only data points "close to" the cutoff, based on the estimated optimal bandwidths in Table 4. 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 g (which is the only result that is statistically significant). Moreover, as indicated by the interaction of "being informed" and education, people with higher education levels are more responsive in the sense that they increase 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.

6.4. "Sharp" vs. "fuzzy" RD

A "sharp" RD is appropriate if two conditions are satisfied: (a) consumers are perfectly informed of their hypertension status based on their SBP readings relative to the cutoff; and (b) the hypertension diagnosis is given in a discontinuous manner. As for the former, although the interviewers in the CHNS survey were required to communicate the blood pressure test results to all sample subjects, there is a concern that the interviewers might have communicated *only* their SBP readings in some cases but *not* their hypertension status. If this occurs, whether individuals can self-diagnose hypertension based on their SBP readings is critical for the consistency of the "sharp" RD estimates. In our opinion, any possible bias due to subjects not being explicitly told that their SBP readings indicate that they have hypertension is unlikely to be severe, because the cutoffs used for hypertension diagnosis are likely to be common knowledge for much of the adult population in China, as individuals may have obtained such general knowledge via TV, newspapers, the internet, social media, talking to people, or occasional physical checkup reports from hospitals.[23] As for the second condition, it is possible that hypertension diagnosis is communicated to the individuals in a more continuous manner.[24] For example, one is diagnosed as hypertensive, pre-hypertensive and normal, rather than hypertensive or normal. To check on this, we experimented with different cutoffs (120, 130, 135, 145, 150, 160, etc.) to see if there were any impacts at these different cutoffs. If people's hypertensive status changes more continuously over different blood pressure levels, we should see impacts of these cutoffs in a manner similar to the 140 cutoff. Yet we do not find any significant results at these other cutoffs. However, if it is indeed the case, the RD estimate will capture only the effect of hypertension diagnosis relative to a diagnosis of pre-hypertension and our estimated treatment effects are likely to underestimate the full effect of hypertension diagnosis. If either of the concerns mentioned above

[23] In China, a majority of consumers do not receive medical checkups frequently. However, they may still have acquired the general knowledge about the cutoff from mass media and their old checkup reports.
[24] We thank an anonymous referee for pointing this out.

Table 8
Parametric estimates of the effects of being informed of hypertension status on daily nutrient intake in full specifications.

	Fat (g)[a,b]	Protein (g)	Carb. (g)	Energy (kcal)
Being informed × low income	−5.040 (5.114)	−0.451 (3.026)	−4.035 (15.22)	3.429 (74.76)
Being informed × middle income	−6.274 (5.590)	−0.313 (3.030)	−5.859 (16.25)	7.067 (78.41)
Being informed × high income	−9.508** (5.669)	−1.944 (3.728)	4.191 (20.06)	27.95 (94.92)
Being informed × education	0.125 (0.473)	0.514** (0.309)	1.880** (1.064)	4.119 (6.367)
Being informed × sex	0.352 (4.117)	1.236 (1.949)	−7.209 (9.316)	−9.489 (50.56)
Age	0.014 (0.108)	−0.270*** (0.053)	−2.278*** (0.270)	−9.623*** (1.163)
Sex (1 = male)	6.519*** (2.070)	7.457*** (1.231)	51.87*** (4.848)	340.4*** (22.52)
Years of schooling	0.965*** (0.366)	0.266 (0.224)	−2.698*** (0.891)	−0.262 (4.116)
Income per capita (in 1989 Yuans)	0.004*** (0.002)	0.002* (0.001)	−0.007** (0.004)	0.018 (0.017)
Middle income (1 = yes)	5.006* (3.109)	1.892 (1.898)	−4.318 (7.919)	30.19 (41.00)
High income (1 = yes)	9.264*** (4.647)	2.482 (3.065)	−11.81 (13.94)	−1.739 (60.26)
Urban (1 = yes)	3.478 (3.511)	2.433 (1.802)	−23.68*** (9.153)	−80.60** (48.73)
Year dummy: 2000	−3.076 (2.529)	1.327 (1.675)	−2.908 (7.004)	−46.46 (39.04)
Constant	59.33*** (7.581)	66.74*** (4.149)	456.9*** (21.22)	2562.2*** (91.64)
Number of observation[c]	2477	2477	3351	5866
R-squared	0.06	0.1	0.11	0.09

[a] Quartic polynomials of systolic blood pressure are included in all specifications.
[b] Robust stand errors are reported in parenthesis.
[c] Sample size for each regression is determined by the optimal bandwidth estimated by LLR.
* Significant at 15% level.
** Significant at 10% level.
*** Significant at 5% level.

occurs, our "sharp" RD estimates would be biased downward and a "fuzzy" RD would be more appropriate.

Unfortunately, due to data limitations, we were unable to directly estimate the "fuzzy" RD because doing so requires a treatment variable that indicates whether individuals have ever been diagnosed as having hypertension, regardless of whether it was diagnosed by a doctor or a survey interviewer. In the CHNS, however, we have data only on subject's self-reports concerning a *doctor's* diagnosis on hypertension status (see Section 3.2). For the "fuzzy" RD to be valid, the probability of receiving a doctor's diagnosis has to increase discontinuously at the cutoff in the SBP reading measured in the previous round (i.e. 3–4 years ago). Unfortunately, we do not observe a significant "jump" in the proportion of people who reported to have been diagnosed with hypertension by a doctor at the cutoff, suggesting that the self-reported diagnosis *by a doctor* is a poor measure of being told one's actual hypertension status by any health practitioner and thus cannot serve as the treatment variable in our setup.[25]

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 health.

[25] It is still possible that a person might have been diagnosed as hypertensive even before the base year and may have been keeping her SBP under control by then. Therefore, we also excluded individuals who kept their SBP under control by taking anti-hypertension drugs at the base year, which accounted for about 0.8% of the total sample. This exclusion did not make much difference in the findings.

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 g three to four years after their diagnosis, although this effect is significant only for the richest third of the population.

Our findings complement 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 is 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. Our results can help explain their puzzling findings. 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. In Section 3, we find a significantly positive income-hypertension gradient, which mainly reflects food consumption patterns. On the other hand, our results in Section 6 also suggest that after hypertension was diagnosed, only rich individuals significantly reduced fat intake. Together these two effects may result in no significant income-hypertension gradient.

Lastly, our results 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 often imperfectly informed of their health status, which may result in a loss of consumer welfare, and thus need to be informed of their true health status through regular medical checkups.

Appendix A.

See Appendix Fig. A.1
See Appendix Fig. A.2

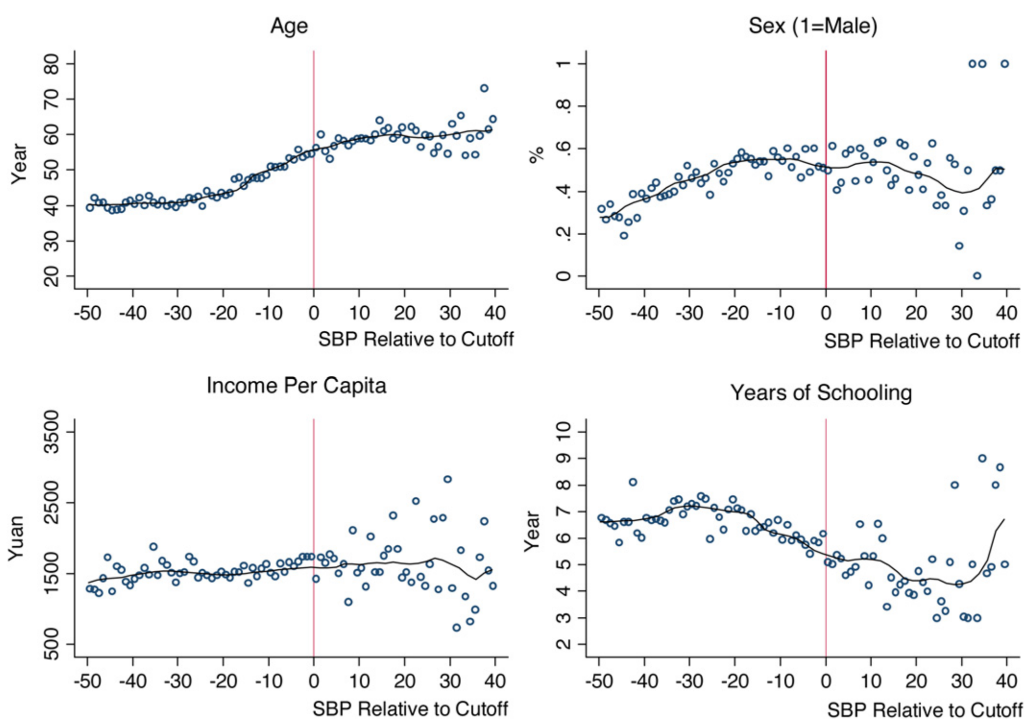

Fig. A.1. Local mean smoothing of major socio-economic variables by systolic blood pressure (bandwidth = 3 mmHg).

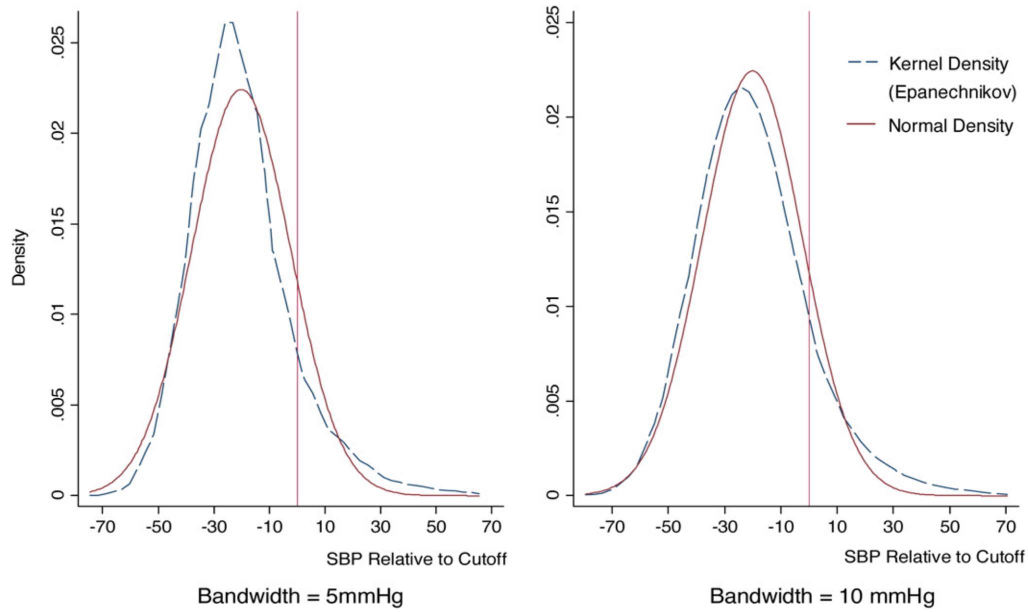

Fig. A.2. Distribution of systolic blood pressure.

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.

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, Thousand Oaks, CA.

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.

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., Paxson, 2001. Mortality, education, income, and inequality among American cohorts. In: Wise, D.A. (Ed.), Themes in the Economics of Aging. University of Chicago Press for the National Bureau of Economic Research, Chicago, pp. 129–165.

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 Economics 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, vol. 2. North-Holland, Amsterdam.

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

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.

Kannel, W.B., 2000. Elevated systolic blood pressure as a cardiovascular risk factor. American Journal of Cardiology 85 (2), 251–255.

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., Manríquez, L., Paredes, M., Román, 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.

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.

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

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 growth? Health Affairs 27 (4), 1064–1076.

Pritchett, L., Summers, L., 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.

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, 200–216.

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.

Wilkinson, R.G., 1986. Class and Health: Research and Longitudinal Data. Tavistock Press, London and New York.

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, Canada and the United States. Hypertension 43, 10–17.

World Health Organization, 2003. Obesity and Overweight: Global Strategy on Diet, Physical Activity and Health. WHO Report.

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

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