RIETI Discussion Paper Series 23-E-010

Make Behavioral Changes for a Healthier Liver? Evidence from a liver function test in Japan

ZHAO, Meng
Gakushuin University

YIN, Ting
RIETI

SEKIZAWA, Yoichi
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 23-E-010

February 2023

Make behavioral changes for a healthier liver? Evidence from a liver function test in Japan*

Meng Zhao[a], Ting Yin[bc], Yoichi Sekizawa[c]

[a] Gakushuin University, Tokyo, Japan

[b] Hitotsubashi University, Tokyo, Japan

[c] Research Institute of Economy, Trade and Industry, Tokyo, Japan

Abstract

Health screening has gained increasing attention for promoting early detection of chronic diseases and lifestyle improvements. This study investigates the impacts of risk information obtained from a regular liver function test on individuals' health behaviors. More specifically, we focus on a biomarker called aspartate aminotransferase (AST) reported from the test. Using rich longitudinal data on health screening and health insurance claims in Japan, we adopt a regression discontinuity design (RDD) approach to examine how individuals respond to the notification that their AST level was abnormally higher than the normal reference range. Our results suggest that, upon receiving an abnormal liver function test result, individuals tend to reduce the likelihood and amount of alcohol use. Furthermore, knowing one's AST value has crossed 40U/I, the higher upper limit used for normal reference range, leads to a significant increase in the expenses on follow-up care such as abdominal ultrasound tests and a decrease in body mass index (BMI) and triglycerides. Some impacts of health signal appear greater among people with high metabolic syndrome risks.

Keywords: health behaviors, liver disease, health checkup, information, Japan
JEL classification: C14; C26; I12; J26

* This study is conducted as a part of the project titled "Basic Research for Exploring the Ideal Medical Intervention after the Advent of the New Coronavirus" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the data provided by the JMDC data center, with the funding support from the Research Institute of Economy, Trade and Industry (RIETI). The authors are grateful for helpful comments and suggestions by Shujiro Urata (RIETI) and Discussion Paper seminar participants at RIETI.

1. Introduction

Chronic diseases are the leading causes of death, accounting for approximately 70% of annual global death, and are often preventable by improving lifestyles (WHO, 2009; 2017). In the past decades, health information has gained rising attention as a policy instrument for health promotion, under the assumption that individuals will choose a healthier lifestyle if they are better informed. For example, a shock in the knowledge about one's health status may lead an individual to update her perceptions on the benefits of a health investment and, as a result, increase the optimal amount of such investment.

One of the important approaches to address the lack of information on health status is providing regular health screening. In most developed countries, public health screenings are often recommended, or even mandated. For example, free regular health checkups are provided to those aged 40-75 in England by the National Health Service since 2011 (Dalton and Soljak, 2012). All private health plans and Medicare are required to provide free preventive care, including regular health screenings, to adult enrollees by the Affordable Care Act in the United States (KFF, 2013). In Japan, employers are mandated to provide free annual health checkups to their employees under the Safety and Health Act of 1972. Furthermore, the government passed a law in 2007 that requires every insurer provide free annual health checkups to those aged 40-74 and implement counselling for those with elevated risks of metabolic syndromes (Tsushita et al., 2018).

However, some recent studies have reported little or limited response in health behaviors after individuals received risk information (Almond et al., 2016; Cook, 2019; Dai et al., 2022; Iizuka et al., 2021; Kim et al., 2022; Kim et al., 2019; Rodríguez-Lesmes, 2021). These studies have generally addressed the endogeneity issue of risk information by applying a regression discontinuity design (RDD), so their findings call for more

careful analysis to understand the impacts of health information comprehensively. Possible reasons for the lack of response are threefold. Firstly, individuals may be already well informed and making the optimal choices regarding health behaviors at the stage of diagnosis (Oster, 2018). Secondly, some health behaviors are more likely to be modified than the others, determined based on their cost-effectiveness. Missing important outcomes may lead to a conclusion of the lack of response. Thirdly, a number of previous studies assumed that the same threshold is applied to create risk information. It will lead to a downward bias in the estimates from a sharp RDD if medical institutions actually use different normal reference ranges.

To address these issues, using rich longitudinal data from Japan and applying a RDD, this study examines the impacts of a regular liver function test on individuals' health behaviors. To the best of our knowledge, there is almost no previous research on health behavior effects of health signals of liver diseases so far. Liver diseases, accounting for 2 million deaths worldwide, are well-known for being "silent" and often deteriorate without noticeable symptoms for years (Asrani et al., 2019). Moreover, lethal liver diseases, such as liver cancer and cirrhosis, are increasing caused by chronic liver conditions affected by health behaviors, making liver diseases more relevant to preventive care.

We believe this study could contribute to the literature in at least three ways. Firstly, the study focuses on a biomarker called aspartate aminotransferase (AST) that has gained little attention in the literature. AST is an enzyme found mainly in the liver, but also in the heart and other tissues. A high level of AST in blood indicates possible damages in liver or heart. Previous studies mainly studied biomarkers such as blood sugar, blood pressure, triglyceride, cholesterol and BMI (see Section 2 for a complete review). AST

test provides information that could be used to directly monitor the health of liver, one of the largest and most important organs in human body. Individuals may take the risk of liver diseases more seriously, hence respond differently as they would to the risks of metabolic syndromes. Moreover, because liver diseases usually progress without noticeable symptoms, regular AST test result may be an important source of information for individuals to adjust their health behaviors before a late-stage diagnosis.

Secondly, we have examined a wide range of outcomes, including alcohol drinking which is considered as critically important for the prevention of liver diseases. The liver basically removes toxins of alcohol and filter waste products from a person's blood. Alcohol consumption generally increases the burden of the liver and has been widely known for its detrimental impacts on liver health (Becker, 2005). Being able to focus on this key outcome turns out quite important, as our results show that other health behaviors are hardly modified, suggesting the reluctancy to adjust lifestyles in general.[2]

Thirdly, since the risk of liver diseases tends to be higher among those with unhealthy lifestyles, we follow the literature and apply a RDD to address this endogeneity issue. More specifically, we use AST as the running variable and exploit the discontinuity at the cutoff that clinically determines the abnormal range. Without making a strong assumption on uniform criteria across medical institutes, we explicitly consider two commonly used cutoffs: 41U/I and 31U/I. Furthermore, a recent movement among medical institutes that increasingly adopt a tighter criterion serves as a double-edged sword: it complicates our analysis while providing an opportunity to investigate the

[2] Although not reported in the paper, we have also examined a wide range of health behaviors including smoking, regular exercising, daily walking, sleeping and breakfast skipping, for which we find little effects of risk information of liver diseases. Results are available upon request.

interacting effects of such changing background.[3]

Based on rich health checkup and health insurance claims data from 1.9 million Japanese employees during 2008-2018, we find that risk information of liver diseases may reduce the frequency and amount of drinking. Our results also suggest that, knowing one's AST value exceeding 40U/I, the higher upper limit used for normal reference range, leads to an increase in the expenses on follow-up care such as abdominal ultrasound test significantly and a decrease in body mass index (BMI) and triglyceride in the long run. Heterogenous effects are observed by gender and the impacts on drinking and triglyceride reduction appear greater among people with high metabolic syndrome risks. We generally find less effects at the lower cutoff, 31U/I, and for the more recent years when lower cutoff became more prevalent, indicating that a stricter judgement criterion may not necessarily serve its purpose.

The remainder of the paper is organized as follows. Section 2 briefly reviews the previous literature. Important background information on health checkup programs in Japan and liver diseases is provided in Section 3, followed by a description of the data and sample generation process in Section 4. The empirical estimation strategy is explained in Section 5 and results are discussed in Section 6. Lastly, Section 7 concludes.

[3] During the past several decades, 40U/I has often been used as the upper limit of the normal reference range for AST (Wroblewski, 1959). Yet, as more recent clinical studies found that AST of healthy population is getting lower because of the improved public health, the upper limit should be redefined at a lower level. Therefore, the cutoff has been increasingly reduced from 40U/I to 30U/I in recent years.

2. Literature review

Traditionally, economists have studied intensively the effects of public health information such as food nutrition labels and advertisements (Folland et al., 2017). As rapid progress of digit information makes general health information more accessible, especially in developed countries, needs increase for more personal information on health status which may play an important role in decision making regarding health investments, -i.e., updating depreciation rate of health in one's health production function (Grossman, 1972). This trend motivates a growing number of studies focus on information on health status, such as disease diagnosis or warning signals received from health screenings, over the past decade.

Our study is closely related to the strand of literature that apply a RDD to analyze the impacts on health behaviors of providing risk information created based on one's health screening results relative to a clinical threshold. For example, past studies have examined the effects of risk information of hypertension (Chen et al., 2019; Dai et al., 2022; Rodríguez-Lesmes, 2021; Zhao et al., 2013), diabetes (Gaggero, 2020; Gaggero et al., 2022; Iizuka et al., 2021; Kim et al., 2019), hyperlipidemia (Kim, et al., 2019), and obesity (Almond et al., 2016; Cook, 2019; Kim et al., 2022; Kim et al., 2019).

In general, the literature report mixed findings on the health behavior effects of health information. Some studies found positive effects of information provision (Chen et al., 2019; Gaggero, 2020; Gaggero et al., 2022; Zhao et al., 2013), while others found little or no effect (Almond et al., 2016; Rodríguez-Lesmes, 2021), or limited effects for those with higher health risks (Cook, 2019; Dai et al., 2022; Iizuka et al., 2021; Kim et al., 2022; Kim et al., 2019).

Regarding the effects of health information obtained from annual health checkups in

Japan, several studies based on a RDD have been published in recent years (Fukuma et al., 2020; Iizuka et al., 2021; Narisada et al., 2022). In 2008, Japan introduced a nationwide health screening program called the specific health checkup (SHC), and started to provide specific health guidance (SHG) to people who are identified to have elevated risks of metabolic syndromes in the SHC (see Subsection 3.1 for more details). Fukuma et al. (2020) explored the effects of being eligible for SHG using waist circumference as the running variable and found small effects on body weight and waist circumference reduction, but not for cardiovascular risk factors. Narisada et al. (2022) also applied a RDD, using age as the running variable, and showed that being eligible for SHG at age 40 led to lower incidence of diabetes and a reduction in BMI and waist circumference.

Probably the most relevant previous study is Iizuka et al. (2021) which used the same data, but for a different time period (2005-2014), to analyze the impact of being notified of abnormal fasting blood sugar level after a regular health checkup in Japan. Although they found some significant positive effects for the high-risk population, their results generally suggest that the health signal of prediabetes condition has little effect on health behaviors such as exercising, drinking and smoking, yet tends to increase the use of follow-up medical care mainly for a glucose tolerance test used for the diagnosis of diabetes.

There are two possible reasons why Iizuka and his colleagues found little effects of health risk information. Firstly, food consumption is probably one of the most important factors of diabetes, yet this outcome is missing in their analysis due to the lack of detailed data. In fact, Oster (2018) reported a significant reduction in the purchase of unhealthy foods after the diagnosis of diabetes in the United States. Examining detailed data on

nutrient intake, Zhao et al. (2013) also reported a reduction in fat intake after individuals were notified of the risks of hypertension in China. The second possible reason is that the threshold used to determine prediabetes condition actually varies by medical institutes in Japan, thus the estimates are likely to be under-estimated so that marginal effects may be overlooked. This study complements the previous research by focusing on the effects of a liver function test which is medically important but has gained little attention so far. Furthermore, we examined a wide range of health behaviors including alcohol use, by far the most important factor of liver diseases, while explicitly considering the heterogenous cutoff issue.

3. Background

3.1 Health checkup programs in Japan

Japan has two major programs of health screenings: general health checkup (GHC) and specific health checkup (SHC). Employers are mandated to provide annual GHC to their employees under the Industrial Safety and Health Act of 1972 with penalties for noncompliance. Employees are obliged to take GHC, although there are no legal penalties for noncompliance. Many organizations including private companies oblige their employees to take GHC every year under their labor regulations. Hence, approximately 80% of employees take the GHC in Japan and the ratio reaches as high as 87.8% for larger organizations (Ministry of Health, Labor and Welfare, 2012, 2013b).

In April of 2008, the government of Japan initiated a National Health Screening and Intervention Program (NHSIP), which consists of two components: (a) specific health checkup (SHC) and (b) specific health guidance (SHG). This program covers the general public aged 40-74 and intends to identify people with obesity and cardiovascular disease

risks (known as metabolic syndrome) through an annual SHC. Those who are identified as having high risks, based on the criteria set by the guidelines of the Ministry of Health, Labor and Welfare (MHLW), are provided with SHG to improve their lifestyles (MHLW, 2013a).[4] Besides the typical health checkup, SHC participants usually need to complete a questionnaire on their health-related lifestyles and existing health conditions. See the sample of a standard questionnaire in the Appendix (MHLW, 2013a). The SHC and SHG programs have been implemented through three terms: (a) fiscal years of 2008-2012; (b) fiscal years of 2013-2017; and (c) fiscal years of 2018-2023.[5]

3.2 Liver diseases and liver function test

Liver diseases have long been a serious threat to global public health. In general, major chronic liver diseases include hepatitis, alcoholic hepatitis, nonalcoholic steatohepatitis (NASH) in early stages, and cirrhosis and liver cancer in late stages. The major risk factors of liver diseases include hepatitis virus (type B and type C), alcohol consumption and liver fat. According to the International Agency of Research on Cancer (IARC) at the World Health Organization (WHO), in 2020, a total of 905,700 people were diagnosed of liver cancer, among whom 830,200 (91.7%) died, making liver cancer one of the top 3 causes of cancer death in 46 countries (Rumgay et al., 2022). Meanwhile, cirrhosis was the 11th leading cause of global death, accounting for 1.32 million total deaths worldwide in 2017 (Cheemerla and Balakrishnan, 2021).

Thanks to the improvement of prevention and treatment of viral hepatitis in the past

[4] Although the GHC and SHC are established by different laws, those who take the GHC are treated as taking the SHC under the rule set by the MHLW. Therefore, employees are generally more likely to take the SHC than non-employees such as the self-employed or non-working individuals.

[5] Government policies in Japan are usually implemented based on the cycle of a fiscal year, starting from April 1st and ending on March 31th of the following year.

several decades, the age-adjusted death rate of chronic liver diseases has gradually declined from 20 to 16.5 per 100,000 people during the period 1990-2017 (Cheemerla and Balakrishnan, 2021). On the other hand, lifestyle risk factors are becoming more responsible for chronic liver diseases (Yuan, et al., 2022). For example, alcohol consumption is the major cause of alcoholic hepatitis, it also deteriorates type B and type C hepatitis, accelerating the fibrosis progression to cirrhosis. Overall, alcohol use is found to account for 50% of mortality related to cirrhosis (Asrani, et al., 2019). Recently, nonalcoholic liver fat disease (NALFD), a liver condition when excessive fat is built-up in livers of people who drink little or no alcohol, is becoming the most prevalent chronic liver diseases. NAFLD is highly related to obesity, as well as other metabolic risk factors such as diabetes and hyperlipemia (Yuan, et al., 2022). Worldwide prevalence of NAFLD is estimated to be around 24% and is expected to continue grow in the next decade. NAFLD could further develop into nonalcoholic steatohepatitis (NASH), or even cirrhosis and liver cancer. (Asrani, et al., 2019; Cheemerla and Balakrishnan, 2021)

Both GHCs and SHCs in Japan include several routine liver function tests, one of which is the aspartate aminotransferase (AST) test, a common blood test to find out the concentration of AST in blood. When AST level is higher than the normal range, there is a concern of liver damages and health checkup providers usually send a warning sign of the risk (e.g., a colored mark or a low-grade out of the range A~E), together with the test results to the examinee. In a serious case, a secondary test is recommended. Follow-up medical care includes an abdominal ultrasound scan, one of the most common tests to examine the conditions of liver and blood vessels in real time. More detailed lab tests (i.e., HBs and HCV tests) are also recommended when viral infection is suspected. In general, according to the suggestions of the Japan Society of Hepatology, prevention of liver

diseases includes: (a) avoiding alcohol drinking which often imposes extra burden on the liver; (b) eating a healthy diet; (c) exercising regularly if possible; (d) avoiding smoking if possible. (JSH, 2020)

3.3 Notification of health checkup results

Implementation of health checkups, both GHC and SHC, is usually entrusted to medical institutes. In addition to reporting to health insurers, results are also sent to individuals after finishing the health checkup. For each biomarker measured in the health checkup, the MHLW proposes a guideline regarding the normal reference range and recommended health guidance (MHLW, 2013a). Particularly, regarding the AST test, the MHLW recommends to provide guidance to those with AST≧31U/I and recommend a secondary test to those with AST≧51U/I. However, medical institutes are not obliged to comply with these guidelines. For example, the National Federation of Industrial Health Organizations (called "Zeneiren" in Japanese) conducted an annual survey on normal reference ranges for AST used in more than 300 medical institutes randomly selected nationwide. According to the survey, medical institutes often determine their own criteria mainly based on their discretions on the guidelines of Japan Society of Comprehensive Medical Checkup, academic literature and manuals of reagents used in lab tests.[6]

More specifically, Figure 2 shows the distribution of the upper limit of normal reference range used in practice by medical institutes in 2014, 2018 and 2021. Despite the recommendation of MHLW, in 2014, only 31% of medical institutes set the upper limit of normal reference range for AST as 30U/I, with 40U/I being more prevalent

[6] The normal reference range for AST set by the Japan Society of Comprehensive Medical Checkup is the same as that of MHLW.

because it has long been used as the standard criterion (Wroblewski, 1959). However, the trend changed rapidly in the following years: the share of medical institutes using 30U/I increased to 0.43 in 2018 and 0.55 in 2021; while that of those used 40U/I decreased from 0.36 in 2014 to 0.23 in 2021, suggesting an increasingly tighter screening standard among medical institutes. In fact, according to the survey, approximately 8~17% of medical institutes reported that they had changed the standard from the previous year every year during 2014-2019.[7]

Individuals' AST test results are usually reported together with an indicator showing whether they are within normal range or abnormally too high or too low, the risk information of interest. In more serious cases, doctors' opinions may be provided in the end of the report, suggesting lifestyle changes, a secondary test or immediate treatments. There may be some variation in doctors' judgements, but it is probably relatively small for AST levels slightly higher than the cutoff which are our focus in this study.

4. Data description

4.1 Data

We use data from the Claims Database of the Japan Medical Data Center (JMDC), Japan's largest epidemiological receipt database available for academic and industrial use (Nagai et al., 2021). Rich data on medical expense receipts (inpatient, outpatient, and prescription) and health checkups are collected from multiple employer health insurance societies since 2005.[8] All these data can be linked by a patient ID at the individual level.

[7] The Zeneiren survey is not a part of the data used in our analysis, so the normal reference ranges reported in the survey may not be exactly the same as the ones used in the health checkups in our data.

[8] Note that the self-employed and retirees are not included in the data because they usually join a

The JMDC data are particularly important for our study in three manners. Frist, the data provide detailed information on all participants' annual health check, including the specific timing and a survey on lifestyle and existing health conditions (see the Appendix). Second, the rich health insurance claims data enables us to examine the prevalence and occurrence of diseases in the general population, including healthy people, and to track patients at multiple healthcare providers over time. Third, such large and longitudinal data are sufficient to implement a RDD and allow us to examine long-term trends in health status.

4.2 Sample

Considering the political changes related to health checkup in Japan discussed in Subsection 3.1, we focus on the JMDC date from the fiscal year of 2008 to the fiscal year of 2017. We define two study periods for the analysis: fiscal years of 2008-2012 and fiscal years of 2013-2017 because of two reasons: (a) we generally find quite different trends in the results for these two periods (to be discussed in Section 6), so that we think it is important to differentiate them; (b) the timing is aligned with the two periods of the implementation of SHC and SHG discussed in Subsection 3.1.

We first drop the data before April of 2008, integrate the data for those who have multiple health checkup records in a single month, and then exclude all the dependents and observations under 30 or over 64 years old, which leads to a sample of 9,174,246 observations, or 2,167,804 individuals. In our analysis, we mainly focus on those who were not diagnosed of major liver diseases in the past 12 months, approximately 7.5% of the total sample. Furthermore, samples used for the analysis of the short-run and the long-

community-based health insurance society in Japan.

run impacts are differently defined (see more details in Figure 1).

Table 1 describes the size of three samples before excluding observations with past diagnosis of major liver diseases and their attrition rates from 2008 to 2018. JMDC database was able to increase the sample size over time by including more health insurance societies. We are not aware of any systematic reasons why the sample increases besides the rising recognition of JMDC. But there is a possibility that more large health insurance societies, with higher ratio of female and older employees, joined over years, as the share of females increased, from 9% in 2008 to 23.3% in 2018, and mean age increased from 44.7 to 47.1 during this period. We therefore explicitly control these characteristics in our analysis. Moreover, it is implausible for the increase in the sample size to be correlated with the treatment cutoff in AST, causing a bias in the RDD estimates.

The attrition rate over time is relatively small, as Japanese employees are usually required to join the health insurance societies determined by their employers or occupations. Note that the attrition rates are large for the sample used for the long-run impact analysis in 2016 and 2017, because there are not enough years to observe the outcomes in the following 24-48 months after an initial checkup. We actually extract the data until August of 2019, the latest data available, to get as many observations as possible, so that the attrition rate in 2017 is less than 1.

4.3 Descriptive statistics

Table 2 presents the summary statistics for the variables used in the analysis. In total, we have more than 6.87 million observations in the sample during the fiscal years of 2008-2017, including those with the diagnosis of major liver diseases in the past 12 months. Approximately 81% of the entire sample are men, with a mean age of 47.

Approximately 7.3% were diagnosed of major liver diseases within the 12 months before the health checkup. Health behaviors and health outcomes in the baseline years are shown in the middle and bottom panels. On average, 40% of the full sample reported taking dinner within 2 hours before bed for more than 3 times a week. We also find that nearly 30% of individuals drink every day in the sample, and 57% drink more than 180ml Japanese sake on a drinking day.[9] A small but significant proportion, 5.7%, reported drinking more than 540ml on a drinking day. Six health outcomes are displayed, including the AST which is the running variable in the RDD analysis. Comparing the 2008-2012 sample against the 2013-2017 sample, except for the sample size, no apparent differences in individual characteristics are observed between these two periods.

Furthermore, we examine the distribution of AST value between the two study periods in Figure 3, grey for the first period and white for the second. We use the bin size of one for the histograms as AST is available only in integer. Figure 3 shows that the distributions almost overlap, indicating that there is no systematic difference in the running variable in the two periods. Moreover, a smooth distribution of AST value, at either the cutoff of 31U/L or 41U/L, suggests that the non-random heaping problem is probably not a concern in our analysis.

5 Empirical framework

We rely on a standard sharp regression discontinuity design (RDD) to estimate the effect of providing risk information on AST to GHC/SHC participants. We mainly

[9] The questionnaire uses one cup of Japanese sake (180ml) as a standard drink to measure the amount of alcohol use. According to the provided instruction, the pure alcohol contained is equivalent to one bottle of regular beer (500ml), two cups of table wine (240ml) or a cup of whisky (60ml).

examine three sets of outcomes: (a) health behaviors including eating behavior (i.e., eating speed and whether eating dinner late frequently) and drinking behavior (i.e., frequency and amount of drinking); (b) healthcare usage measured by the expenses on following biological tests, medical imaging tests and medication; (c) health outcomes measured by major biomarkers, namely AST, BMI, systolic blood pressure, fasting blood sugar, LDL-cholesterol and triglyceride.[10] In this setting, AST in year t is considered as the running variable. Based on discussion in the Subsection 3.3, as most medical institutes use either 30U/I or 40U/I as the upper limit of normal reference range for AST (Figure 2), we set the cutoff as 31U/I and 41U/I for our RDD analyses. Let c denote the cutoff, we estimate the following local polynomial regression:

$$Y_{it+1} = \alpha_0 + \alpha_1 D_{it} + \alpha_2 (AST_{it} - \mathrm{c}) + \alpha_3 (AST_{it} - \mathrm{c})^2 + \boldsymbol{\beta X}_{it} + \varepsilon_{it}, \qquad (1)$$

where Y_{it+1} represents the outcome variables for individual i in year $t+1$, D_{it} a dummy variable that equals one if AST_{it} greater than or equal to c, and ε_{it} the idiosyncratic error term. A vector of covariates $\boldsymbol{X}$ is included to control for basic individual characters, such as age, age squared, sex, and year-specific effects. We clustered standard errors by individuals for the analyses since the same individuals were observed repeatedly in different years. Following the literature, a triangular kernel is used to weight the sample. We have experimented with various bandwidth and 5U/I is preferred. An even larger bandwidth may lead to biased results which could pick up the effect of the neighboring

[10] Technically, a fuzzy RDD will be more appropriate, because the normal reference range varies by health checkup providers. However, unfortunately, since we do not have data on the exact cutoff used for each individual's AST test, we choose a sharp RDD as the second-best choice, focusing on two most commonly used cutoffs: 31U/I and 41U/I. We believe the analysis is still meaningful to provide important evidence, yet the sharp RDD is likely to under-estimate, providing only the estimates of the lower limits of the true effects.

cutoff, 10U/I away, while a smaller bandwidth results in the loss of estimation precision.

Empirically, we examine the outcomes in the short run defined as those observed during the following 24 months after the initial health checkup, and in the long run within the following 24-48 months. Note that we examine health care usage only for the following 12 months to focus on the follow-up medical care.

6 Results

6.1 Preliminary checks

Before estimating the impact of risk information in a RDD, it is useful to first check the relationship between AST and the key outcomes of interest by simple charts. Figures 4-6 plot the averages of health behaviors, healthcare usage and health outcomes in the short run, while Figure 7 shows the averages of some key long-term outcomes. Both of the two most commonly used cutoff, 31U/I and 41U/I, are marked by red straight lines. As explained in Sections 3 and 4, we separate the study period into two: fiscal years of 2008-2012 (or 2008.4-2013.3) and fiscal years of 2013-2017 (or 2013.4-2018.3). Straight lines are the best-fitting lines, without controlling for covariates, left and right to the cutoffs.

As shown in Figure 4, regarding eating behavior, a relatively obvious increase is observed at the cutoff of 41U/I for the average self-rated eating speed score (implying a decrease in speed) and a drop in the share of people eating dinner late (a drop), only in the first period. When it comes to drinking frequency, a drop is visible at both 31U/I and 41U/I for the share of people who reported drinking every day, and there appears an increase in the share of those who do not drink at 41U/I. Again, these changes are only observed in the first period. The last two health behavior outcomes measure whether

individuals drink more than 180ml or 540ml Japanese sake on a drinking day. The charts show that there is a similar drop in both outcomes at the cutoff of 41U/I in the first period.

Figure 5 presents the distributions of the averages of the medical expenses on four types of healthcare usage: biological tests, abdominal ultrasound, computerized tomography (CT) and/or magnetic resonance imaging (MRI) scan, and drug prescription for all kinds of diseases.[11] A small increase in the expenses on biological tests and abdominal ultrasound scans is observed at the cutoff of 41U/I in the first period. The use of expensive CT/MRI scans is relatively rare, reflected by the large fluctuation by AST level, suggesting that the decrease at 31U/I and the increase at 41U/I may be picking up a random variation. Lastly, no obvious changes are observed at the cutoffs for drug prescription expenses in both time periods.

Furthermore, we look at the health outcomes measured and the probability of being diagnosed of major liver diseases during a doctor visit in the short-run. As shown in Figure 6, except for LDL-cholesterol, most biomarkers generally increase at AST around the cutoffs of 31U/I and 41U/I stably and smoothly, with no obvious discontinuity, in both periods. Plots for the LDL-cholesterol are noisier, with larger standard deviations, and there seems a small drop at 41U/I in both study periods. Lastly, there appears an increase at the cutoffs of 31U/I and 41U/I, during the first study period, in the share of people who were diagnosed of major liver diseases within 12 months after the AST test.

Lastly, in comparison to the short-run outcomes, the plots for the key long-run outcomes are presented in Figure 7. The pattern of the share of everyday drinking is very

[11] Because there is a wide range of liver disease drugs and some liver diseases may be treated together with other complications, it is practically difficulty to single out drugs specific to liver disease. Expenses on CT/MRI of the examinations of head, limbs, blood vessels, breast and colon are excluded.

similar in both the short-run and the long-run: a drop is generally observed at both cutoffs in the first period but not the second period. Interestingly, for the share of no drinking, as well as those of drinking more than 180ml Japanese sake and those drinking more than 540ml on a drinking day, besides a similar pattern for the first period, the gaps at the cutoff of 41U/I become more visible in the long-run. When it comes to the long-run health outcomes, a small drop is observed at the cutoff of 41U/I for BMI, systolic blood pressure, fasting blood sugar and triglyceride in the first period. The only biomarker that also exhibits a drop at 41U/I in the second period is triglyceride.

The patterns in Figures 4-7 are suggestive of some possible impacts of risk information of liver diseases, yet they may be simply picking up a random variation. More careful analysis relies on the local polynomial estimation discussed in the following subsections.

6.2 Effects on health behaviors

This subsection discusses the results of the local polynomial regressions for the health behaviors of interest. Table 3 shows the estimates for the short-run (left-panel) and the long-run (right-panel) impacts at two cutoffs, 41U/I and 31U/I, in the top and bottom panels, respectively. Since the estimates are clearly different, we choose to conduct the analysis separately for the two study periods. All regressions include standardized AST and its squared term, sex, age, age-squared, and examination year dummies. The results do not change much without covariates. We have experimented with local linear regression and do not find much difference, either. Robust standard errors, clustered at individual level, are reported in parentheses. Results reported in Tables 3-4 are based on the estimation using our preferred bandwidth of 5U/I, and sensitivity analysis regarding

bandwidth choice will be discussed in Subsection 6.5.

In general, the coefficients reported in Table 3 are aligned with the patterns observed in Figures 3 and 7. For example, both the short- and the long-run self-rated eating speed scores increase at 41U/I in the first period in both short and long run. But they are not statistically significant after taking into account the robust standard errors. On the other hand, the coefficients are generally close to zero and, thus, statistically insignificant in the second period, despite a much larger sample and a lower standard error. Surprisingly, a negative long-run impact (i.e., a faster speed), statistically significant at the 5% level, is observed at the cutoff of 31U/I in the first period. Clinical studies have generally shown that eating fast may increase the risks of obesity and diseases such as NALFD, diabetes and cardiovascular diseases (Kolay, et al., 2021).[12] There are two possible explanations: (a) a low cutoff may reduce the risk perception and lead to an unexpected lifestyle change; (b) the estimate may be picking up a random variation. The bandwidth sensitivity analysis in Figure 9 actually suggests that the result is sensitive to a small bandwidth, such as 3U/I and 4U/I.

Eating late dinner is usually recommended to be avoided for maintaining a healthy weight. The results show that the probability of reporting frequent late dinners is about 2.6% lower at the cutoff of 41U/I within 24 months after the health checkup, statistically significant at the 10% level. The coefficient becomes very small, close to zero, in the second period though. Neither do we find any significant long-run impacts, which is consistent with the preliminary checks in Figure 7.

---

[12] Clinical studies have shown that eating slowly may suppress appetite, reduce energy intake and improve digestion, because of hormone-independent factors as well as the influences on gastrointestinal hormones that significantly affect hunger and satiety. See Kolay et al. (2021) for a complete review.

Lastly, when it comes to drinking behaviors, which is most likely to be recommended to prevent liver damages, several interesting findings are suggestive. Firstly, consistent to Figures 4 and 7, we find that, at the cutoff of 41U/I, the probability of drinking 180ml Japanese sake in the short run is reduced by 2.6%, significant at the 10% level, in the first period. The coefficient drops to -0.017 for the long-run impact, and turns insignificant. In general, the effect disappears in the second period. Secondly, although the estimates are generally small and insignificant in the short run, the long-run impact on the probability of drinking more than 540ml Japanese sake decreases by 1.6% in the first period, and 0.9% even in the second period, both significant at the 10% level. This is consistent to the pattern observed in the plots. Thirdly, even though the share of those drinking everyday appears to drop at both cutoffs in Figures 4 and 7, only the estimates at 31U/I are statistically significant, suggesting that different cutoffs may have different effects on drinking behaviors.

The results above suggest that risk information created by different cutoffs may impact drinking behavior differently, and takes the effect differently in the short and long runs. For example, a lower cutoff provides an early warning leading to a first-step lifestyle change, -i.e., a reduction in the frequency of drinking. At a higher cutoff, individuals are more likely to take a step further to reduce the amount of drinking little by little, and in the long run, the probability of heavy drinking may also be reduced. Note that the true effects are likely to be under-estimated, as the sharp RDD in our analysis assumes that the same cutoff is applied to everyone, as discussed in Section 5.

6.3 Effects on healthcare usage

Risk information may also increase the usage of healthcare, either for confirmation tests or afterward treatments. Early detection and treatment are supposed to be one of the major purposes of annual health screening, and are expected to contribute to the reduction of long-term total health expenditure.

Table 4 summaries the estimates for four outcomes measuring the usage of healthcare services within 12 months after the initial AST test. In general, we observe a significant increase of 146 yen, at the cutoff of 41U/I, only in the expenses on abdominal ultrasound test, one of the most common follow-up tests for liver diseases, in the first time period. The average expenses in our sample were 651yen (approximately $4.93), but it was 7,530yen (approximately $57) among people who have taken the test.

There are two possible explanations why we find little effect on the other healthcare usage outcomes. Firstly, an abnormal AST result slightly higher than the normal reference range may not be perceived seriously enough to trigger the use of medication or expensive high-tech diagnostic tests such as CT/MRI. Secondly, since we are not able to differentiate the usage of tests and medication for liver diseases from other purposes, we may have missed the impact on some drugs or tests specific to liver diseases. In that sense, among the four outcomes examined, abdominal ultrasound is most specific and, therefore, more likely to capture the impact of health information.

6.4 Effects on health outcomes

Improvement in health outcomes is the ultimate goal of information provision and lifestyle changes. The estimates for six biomarkers that measure health status are reported in Table 5. In general, all the estimates in the first period and most in the second period

are negative at the cutoff of 41U/I, consistent to a change towards a healthier health condition, yet many are statistically insignificant. The estimates for only three biomarkers are statistically significant in a certain period: (a) the long-run AST in the first period, (b) both the short- and long-run BMI in the first period and (c) the long-run triglyceride in the second period.

Particularly, the last finding on the long-run impact on triglyceride stands out because we rarely find any impact in the second period. Moreover, both the magnitude (6.52) and statistical significance (at the 1% level) of the estimate is quite large. Bandwidth sensitivity analysis also indicates that the result is robust to the choice of bandwidth. Given the sample mean of 119.6mg/dL, the estimate indicates a 5.4% reduction of triglyceride 2-4 years after being informed of the risk of liver diseases at a higher cutoff, 41U/I.

Lastly, we find that the probability of formal diagnosis of major liver diseases increases slightly in the following 12 months, approximately by 0.5% at the cutoff of 41U/I. In contrast, we generally do not find much significant impact at the cutoff of 31U/I, in both periods, except for a surprising increase in the short run BMI. More discussion on the robustness of this result will be given in the following subsection.

6.5 Robustness checks

The accuracy of the RDD estimates relies critically on two assumptions: (a) the assignment variable is not manipulated at the cutoff so that some individuals are more likely to be assigned to the treatment group; (b) the samples, within the range set by chosen bandwidth, on both sides of the cutoff are comparable in absence of the treatment.

Standard checks on these assumptions as well as the robustness of local polynomial estimation results are reported in this subsection in detail.

Figure 8 displays the histogram of AST, together with the McCrary density estimation for the cutoffs of 31U/I (in the top) and 41U/I (in the bottom), respectively. There appears no visible discontinuity in the distribution and estimated density at the cutoffs, suggesting that the running variable is unlikely to be manipulated. In fact, as a blood test is usually operated in labs, it is probably very difficult to manipulate AST in reality. We have also conducted the t-tests recommended by McCrary (2008) to double check the continuity in AST at placebo cutoffs for various bandwidth choices (Table 6). Although the discontinuity at some cutoffs is statistically significant, following Chetty et al. (2009) and Kim et al. (2019), we have created the cumulative distribution function of the absolute value of t-statistics from the McCrary tests and confirmed that the t-statistics at the treatment cutoffs are not disproportionally greater than those at placebo cutoffs. For example, as shown in Figure 9, the distribution does not exhibit certain noticeable patterns around the straight lines, which indicate the t-statistics at the cutoffs that are no larger than the 50^{th} percentile.[13]

Next, we examine the validity of the second assumption by estimating Eq. (1) for all the covariates and baseline outcomes.[14] If there is no discontinuity, the estimate for α_1 should be statistically insignificant. The check has been done for both cutoffs in both time periods. Results are reported in Table 7. In general, most of the estimates are statically insignificant, implying that the samples are comparable on both sides of the cutoffs. Two out of thirteen baseline

[13] Figure 9 only reports the distribution of t-statistics for the cases using the bandwidth of 5U/I. However, although not reported, the patterns observed when different bandwidths are used are similar. Results are available upon request.

[14] Of course, all the covariates are excluded from the explanatory variables in Eq.(1).

characteristics, self-rated eating speed score and BMI are statistically significant in certain time periods (see p-values in bold). This may be due to randomness, but one also needs to be cautious when interpreting the estimation results for these outcomes. For example, as discussed in Subsection 6.4, we find that BMI in the first period tends to decline at the cutoff of 41U/I. Considering that the sample on the right had a higher BMI in the baseline year, the true effect may be even larger. On the other hand, a positive impact on BMI at the cutoff of 31U/I in the second period may be just reflecting the difference already existed in the baseline year.

The estimates reported in Tables 3-5 are based on the estimation using our preferable bandwidth of 5U/I. Since AST is a discrete variable, we avoid the standard procedure of choosing optimal bandwidth for continuous variables. Instead, we have experimented a set of bandwidths. As expected, in Figure 10, a small bandwidth, such as 3U/I, generally leads to a larger standard error and a wider confidence interval, while a larger bandwidth increases the bias driven by data far away. Note that the 95% confidence intervals are shown in the figure, so the estimate that are significant at the 10% level may still lie inside of the confidence interval. Focusing on the estimates that are statistically significant at bandwidth of 5U/I reported in Tables 3-5, we find that the following estimates are relatively robust to the bandwidth choice: whether drink every day in the first period, long-run triglyceride in the second period and medical expenses on abdominal ultrasound in the first period. As for the marginally significant estimates for the eating behaviors, drinking amount, BMI and the probability of diagnosis of liver diseases, the results are sensitive to a small bandwidth.

Lastly, we also examine the sensitivity to the cutoffs for the key outcomes and the estimates at placebo cutoffs are presented in Figure 11. In general, the estimates are not statistically significant at placebo cutoffs. However, we do find a significant estimate at a few placebo cutoffs, such as 46U/I for the short-run eating speed score in the second period, 51U/I for whether drink

more than 180ml Japanese sake in the second period, and 36U/I for the probability of the diagnosis of liver diseases. This may explain partly why we do not find much impact of health information in the second period. We will further explore this issue in Subsection 6.7.

6.6 Heterogeneous effects

It is possible that health information may have heterogenous effects, so we examine three specific samples separately in this subsection: males, females and people with high risks. Tables 8 and 9 report the results for male sample (left panel), female sample (middle panel) and people with high risks (right panel), respectively. The high risk are defined as those having the metabolic syndromes based on the criteria determined by the Japanese Society of Internal Medicine (MHLW, 2013a).[15]

Since approximately 80% of our sample are males, the results tend to be dominated by those for male sample. First, taking a closer look at results in Table 8 for the cutoff of 41U/I, the estimates for drinking behaviors are generally statistically insignificant for women. The increasing use of abdominal ultrasound test remains significant for both male and female samples in the first period, yet an interesting difference is found in the second period: males tend to reduce the use of CT/MRI while females increase the use of such expensive tests and reduce that of abdominal ultrasound. Females may have a higher tendency to use expensive tests, but it is unclear why males reduce the use upon receiving risk information. Regarding health outcomes, we also observe different patterns for

[15] The criteria for the metabolic syndromes include: (a) waist circumstance $\geqq$ 85(90) for male (female); and (b) two of the following risks are identified: fasting blood sugar $\geqq$ 110 mg/dL; systolic blood pressure $\geqq$ 130 mmHg or diastolic blood pressure $\geqq$ 85 mmHg; and HDL-cholesterol<40 mg/dL or triglyceride $\geqq$ 150 mg/dL (MHLW, 2013a).

females: the effect on BMI disappears, while a strong effect is observed for LDL-cholesterol in the first period.

Lastly, as shown in Table 8, when it comes to the sample with higher metabolic syndrome risks, we find a much strong effect on drinking amount, a reduction of 6.3% (4.9%) in the probability of drinking more than 180ml (540ml) Japanese sake on a drinking day in the first period. Moreover, a larger reduction in triglyceride is also observed among this sample. The results suggest that the high-risk population are more responsive to the risk information regarding some key outcomes such as the amount of drinking.

Finally, we also examine the heterogenous effects at the cutoff of 31U/I (Table 9). Again, the effect on the probability of drinking every day is mainly driven by that of the male sample, and a different pattern is observed for the expenses on biological tests: a reduction for males but an increase for females. In general, unlike the effects at the cutoff of 41U/I, there seems little impact on health outcomes at 31U/I, except for the high-risk group for whom we observe a decrease of 1mmHg in systolic blood pressure.

6.7 Why are the estimated effects different in the two study periods?

Probably the most puzzling finding in our analysis is the drastic differences in the estimates for the two study periods. Despite a larger sample size, the estimates for the second period are generally less significant, mainly due to a smaller magnitude of the coefficient. We have also conducted the estimation for the full sample, only to find that the estimates are dominated by those of the second period. We experiment with different study periods and find that the years after 2013 tend to share similar trends. Several

hypotheses are proposed and tested in this section to explore the reasons for this issue.[16]

The first hypothesis is that individuals in the second period may have already been informed of risk information in previous health checkups, so that they are less responsive in health behaviors. To verify this possibility, we exclude from the sample the individuals with abnormal AST levels (i.e., AST over 40U/I or over 30U/I) in the previous round(s) of health checkup. Unfortunately, the results do not change much and the differences remain.

The second hypothesis is that the samples are different in the two study periods, and the differences are due to a sample bias. We first compare the sample characteristics in Table 2 and find they are very similar in all the key variables. But there may be some differences in unobservable variables, so we focus on a subsample in the second time period who were also the subjects in the first time period. Examining the same individuals in both two time periods should technically eliminate the possibility of sample bias. Again, we find the results for the same individuals remain different in two time periods.

The third possible explanation lies in the impacts of increasingly tighter screening criteria used to judge the normal range of AST value. A decrease in the upper limit of the normal reference range implies that an abnormal AST value in current time period might be within the normal range before, which could result in a change in individuals' risk perception, i.e., taking the risk information less seriously. As a result, health behaviors are less likely be changed at the lower new cutoff (e.g., the cutoff of 31U/I in our analysis). On the other hand, as fewer medical institutes use the higher cutoff (e.g., 41U/I), the estimates at the cutoff are more likely to be underestimated in a sharp RDD.

To confirm this hypothesis, we have examined other biomarkers. For example, the

[16] Although not reported, all the results discussed in this subsection are available upon request.

normal reference range for triglyceride has been quite constant over the two study periods in Japan. According to the annual survey of Zeneiren, the share of medical institutes used 149mg/dL, the most commonly used cutoff, has been around 92.0-93.6% during 2014 to 2019, which is quite constant compared to the case of AST. We therefore use a RDD to analyze the impact of risk information of triglyceride on the same outcomes as we did for AST. Interestingly, we find no big difference in the estimates for the two time periods. In some case, we even find more significant effects in the second period.

7 Conclusions

Providing health information is considered as an important instrument to improve health behaviors for a better public health in the long run. This study pushes forward the frontier of the understanding of the effects of health information by analyzing the impacts of an important liver function test, AST test, that has gained little attention so far. Using rich longitudinal data in Japan, the study applies a RDD to investigate how individuals respond to the risk information of liver diseases. Although the analysis is complicated by increasingly stricter criteria of screening and multiple cutoffs, our results provide several important findings.

Firstly, despite some previous studies that find little health behavior effects of providing risk information, our results suggest that individuals do respond, to a certain extent, to the risk information of liver diseases. But the effects are limited and found only for drinking behaviors commonly recommended as the most important preventive measure for liver diseases. If we only examine behaviors such as eating, smoking and exercising, we are likely to draw a conclusion similar to previous studies.

Secondly, as medical institutes tightening their criteria to determine the normal range

of health checkup results, the impacts of health information may interact with such dynamic changes. For the case of AST, we generally find less significant effects in the second time period, when the upper limit of normal range of AST was switched to a lower level by many health checkup providers. Although we could not provide direct evidence on the impact of changing cutoffs due to the lack of data, our analysis points to the need of future research to further address this issue to better predict the effects of public policies and health interventions.

Lastly, aligned with the findings of previous studies, our results also show that individuals with higher risks tend to be more responsive to health information, reemphasizing the importance of targeting at this group. In that sense, Japan's SHC and SHG program that focus on identifying and helping the high-risk group seems to be an effective approach to improve the health of the target population. One caveat of this study needs to be noted. Due to the data limitation, we could apply only a sharp RDD to estimate the effects of risk information, which is likely to underestimate the true effects because the same treatment cutoff may not have been applied to each individual. Hence, we propose our estimates as the lower limits of the true effects.

8 References

Alalouf, M., Miller, S., Wherry, L.R., 2019. “What difference does a diagnosis make? Evidence from marginal patients”, *National Bureau of Economic Research Working Paper* (W26363).

Almond, D., Lee, A., Schwartz, A.E., 2016. Impacts of classifying New York City students as overweight. *Proceedings of National Academy of Science U.S.A.* 113 (13): 3488-3491.

Asrani, S.K., Devarbhavi, H., Eaton, J., Kamath, P.S., 2019. “Burden of liver diseases in the world”, *Journal of Hepatology* 70(1):151-171.

Becker, U., 2005. “Epidemiology and risk factors involved in alcohol liver disease” In: Preedy V.R., Watson, R.R., editors. *Comprehensive handbook of alcohol related pathology*. London, UK: Elsevier Academic Press: 467-480.

Chen, S., Sudharsanan, N., Huang, F., Liu, Y., Geldsetzer, P., Bärnighausen, T., 2019. “Impact of community based screening for hypertension on blood pressure after two years: regression discontinuity analysis in a national cohort of older adults in China”, *BMJ* 366, l4064.

Chetty, R., Looney, A., Kroft, K., 2009. “Salience and taxation: theory and evidence”, American Economic Review 99(4): 1145-1177.

Cook, W., 2019. “The effect of personalised weight feedback on weight loss and health behaviours: Evidence from a regression discontinuity design”. *Health Economics* 28(1): 161-172.

Cheemerla, S., Balakrishnan, M., 2021. “Global epidemiology of chronic”, *Clinical Liver Disease* 17(5).

Dai, T., Jiang, S., Liu, X., Sun, A., 2022. “The effects of a hypertension diagnosis on health behaviors: A two-dimensional regression discontinuity analysis”, *Health Economics* 31(4): 574-596.

Dalton, A.R., Soljak, M., 2012. “The nationwide systematic prevention of cardiovascular disease: the UK’s Health Check programme”, *Journal of Ambulatory Care Manage* 35(3): 206-215.

Folland, S., Goodman, A.C., Stano, M., 2017. *The Economics of Health and Health Care*. Routledge.

Fukuma, S., Iizuka, T., Ikenoue, T., Tsugawa, Y., 2020. "Association of the National Health Guidance Intervention for obesity and cardiovascular risks with health outcomes among Japanese men. *JAMA Internal Medicine* 180(12):1630-1637.

Gaggero, A., 2020. "The effect of type 2 diabetes diagnosis in the elderly", *Economics & Human Biology* 37, 100830.

Gaggero, A., Gil, J., Jiménez-Rubio, D., Zucchelli, E., 2022. "Does health information affect lifestyle behaviours? The impact of a diabetes diagnosis", *Social Science & Medicine* 314, 115420.

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

Iizuka, T., Nishiyama, K., Chen, B., Eggleston, K., 2021. False alarm? Estimating the marginal value of health signals. *Journal of Public Economics* 195 (104368).

Japan Society of Hepatology (JSH), 2020. *To Better Understand Liver Diseases.* The Japan Society of Hepatology Public Relations Committee, Tokyo.

Kaiser Family Foundation (KFF), 2013. "Summary of the Affordable Care Act", *Health Reform*, April 25, 2013. https://www.kff.org/health-reform/fact-sheet/summary-of-the-affordable-care-act/. Retrieved on Dec. 26, 2022.

Kim, B., Thomsen, M. R., Nayga, R. M., Fang, D., Goudie, A., 2022. "Impact of weight status reporting on childhood body mass index", *Child Obesity* 18(7): 485-493.

Kim, H.B., Lee, S.A., Lim, W., 2019. Knowing is not half the battle: Impacts of information from the National Health Screening Program in Korea. *Journal of Health Economics* 65:1-14.

Kolay, E., Bykowska-Derda, A., Abdulsamad, S., Kaluzna, M., Samarzewska, K., Ruchala, M., Czlapka-Matyasik, M., 2021. "Self-reported eating speed is associated with indicators of obesity in adults: A systematic review and meta-analysis", *Healthcare (Basel)* 9 (11): 1559.

McCrary, J., 2008. "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics* 142(2): 698-714.

Ministry of Health, Labor and Welfare (MHLW), 2012. Special Survey on Industrial Safety and Health, Japan.

Ministry of Health, Labor and Welfare (MHLW), 2013a. Standard health check-up and health guidance programme – 2013 (in Japanese). Retrieved from

https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/seikatsu/.

Ministry of Health, Labor and Welfare (MHLW), 2013b. Survey on State of Employees' Health (in Japanese). Retrieved from https://www.mhlw.go.jp/toukei/list/h24-46-50.html.

Nagai, K., Tanaka, T., Kodaira, N., Kimura, S., Takahashi, Y., Nakayama, T., 2021. "Data resource profile: JMDC claims database sourced from health insurance societies", *Journal of General and Family Medicine* 22(3): 118-127.

Narisada, A., Shibata, E., Hasegawa, T., Wakayama, R., Suzuki, K., 2022. "The impact of the National Health Program on diabetes incidence among working-age men with prediabetes: A regression discontinuity analysis of a nation-wide database in Japan," *Diabetes Research and Clinical Practice* 189, 109946.

Oster, E., 2018. "Diabetes and diet: purchasing behavior change in response to health information", *American Economic Journal: Applied Economics* 10 (4): 308-348.

Prina, S., Royer, H., 2014. The importance of parental knowledge: evidence from weight report cards in Mexico. *Journal of Health Economics* 37: 232-247.

Rodríguez-Lesmes, P., 2021. "Estimating the gains of early detection of hypertension over the marginal patient", *PLOS ONE* 16(7), e0254260.

Rumgay, H., Arnold, M., Ferlay, J., Lesi, O., Cabasag, C.J., Vignat, J., Laversanne, M., McGlynn, K.A., Soerjomataram, I., 2022. "Global burden of primary liver cancer in 2020 and predictions to 2040", *Journal of Hepatology* 77(6):1598-1606.

Tsushita K., Hosler, A.S., Miura, K., Ito, Y., Fukuda, T., Kitamura, A., Tatara, K.. 2017. "Rationale and descriptive analysis of Specific Health Guidance: the nationwide lifestyle intervention program targeting metabolic syndrome in Japan", *Journal of Atherosclerosis Thrombosis* 25(4): 308-322.

World Health Organization, 2009. *Global health risks: mortality and burden of disease attributable to selected major risks*. Geneva: World Health Organization.

WHO, 2017. *Noncommunicable diseases progress monitor 2017.* Geneva: World Health Organization.

Wroblewski, F., 1959. "The clinical significance of transaminase activities of serum", *American Journal of Medicine* 27: 911-23. doi: 10.1016/0002-9343(59)90175-5. PMID: 13846130.

Yuan, S., Chen, J., Li, X., Fan, R., Arsenault, B., Gill, D., Giovannucci, E.L., Zheng,

J.S., Larsson, S.C., 2022. “Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study”, European Journal of Epidemiology 37(7): 723-733.

Zeneiren, 2014, 2015, 2016, 2017, 2018, 2019. Report on Medical Screening Accuracy Survey. https://www.zeneiren.or.jp/research_list/index.html (in Japanese). Retrieved on Dec. 26, 2022.

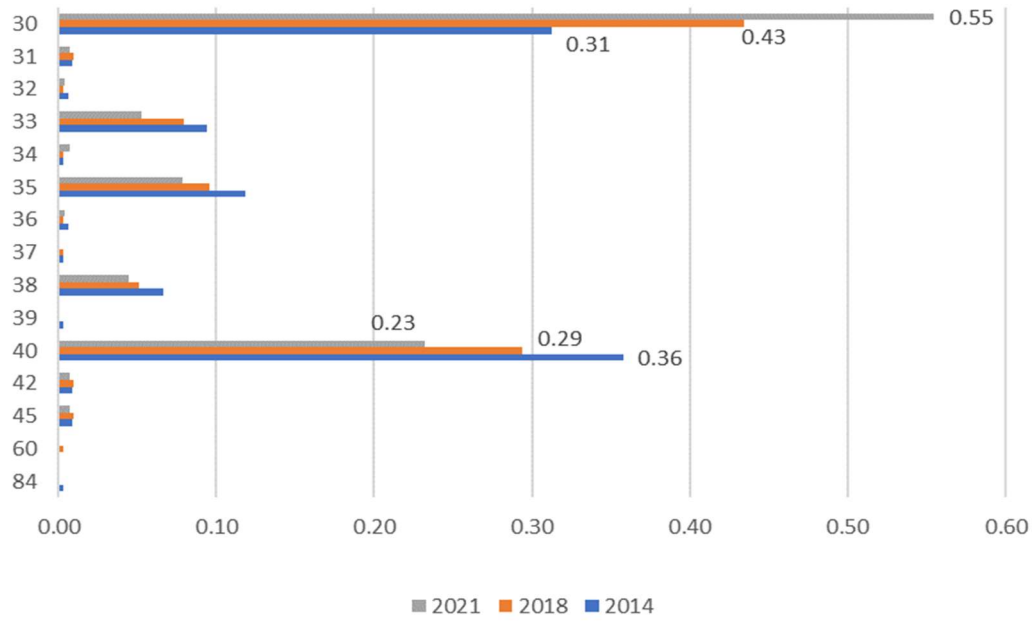
Zhao, M., Konishi, Y., Glewwe, P., 2013. “Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption”, *Journal of Health Economics* 32: 367-385.

Figure 1: Description of sample generation

	No. of observations	No. of indiviudals
Observations with health checkup information from 2005~2019 in JMDC.	13,157,681	3,233,271
↓ (44,533 dropped)		
Step 1: Some individuals had their health checkup results and their answers to a routine questionnaire implemented duing the health checkup recorded in separate observations in the same month. We integrate the data and drop one of the duplicates.	13,113,148	3,233,271
↓ (212,238 dropped)		
Step 2: Drop data before April of 2008.	12,900,910	3,224,045
↓ (1,941,101 dropped)		
Step 3: Drop data of dependants.	10,959,809	2,645,605
↓ (1,785,563 dropped)		
Step 4: Drop those aged younger than 30 or older than 64.	9,174,246	2,167,804
↓ (696,397 dropped)		
Step 5: Drop those who were diagnosed of major liver diseases in the past 12 months	8,477,849	2,101,607
↓		
Step 6-a: Sample for short-run impact analysis: Individuals with health checkup information in two continous periods within 24 months.*	5,866,345	1,674,585
Step 6-b: Sample for long-run impact analysis: Individuals with information available in the next time period 24~48 months later.*	3,791,667	1,181,395

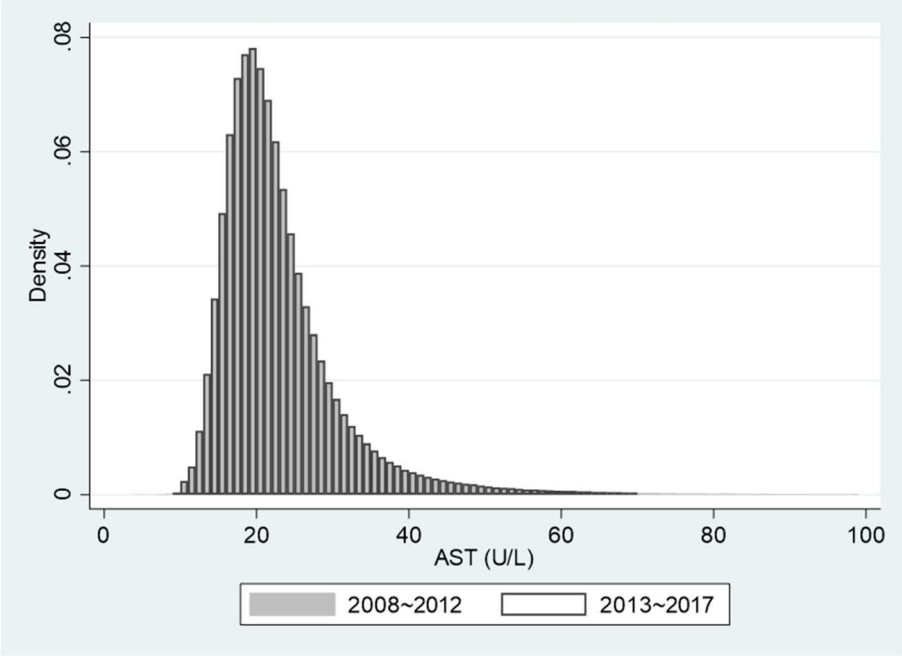
Notes:
* Sample size varies slightly by outcomes due to the missing values. Since we are using outcomes in following time periods, we define the end of the study period as March of 2018, though some information in 2019 is extracted.

Figure 2: Distribution of upper limit of normal reference range for AST



Sources: Medical Screening Accuracy Survey Report, Zeneiren (2014, 2018, 2021)

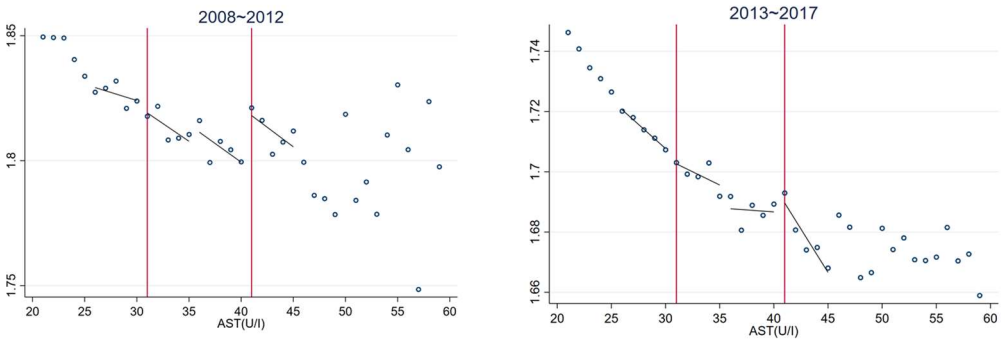
Figure 3: Distribution of AST by time periods



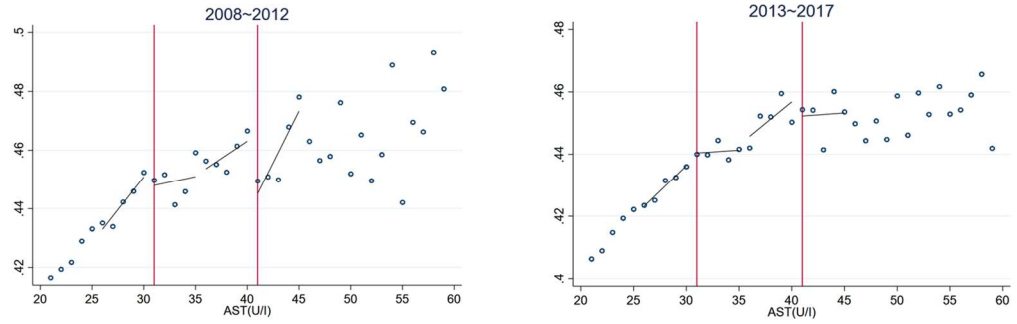
Notes:
Bin size 1 is used and very few observations with AST over 100U/I are suppressed for clearer vision.

Figure 4: Averages of health behavior measures in the short run by AST

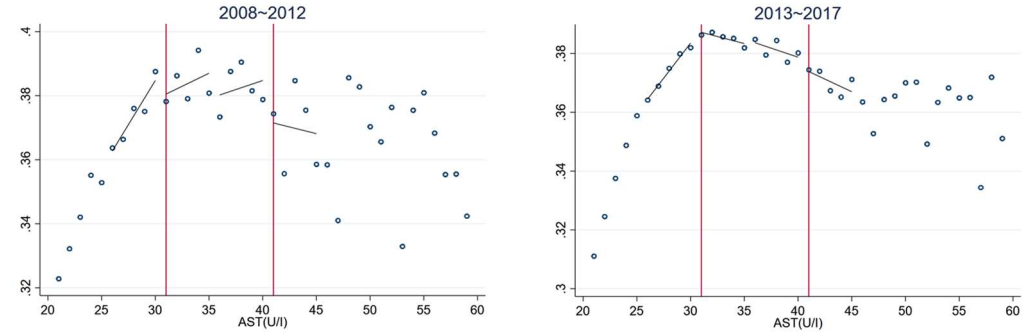
1. Self-rated eating speed



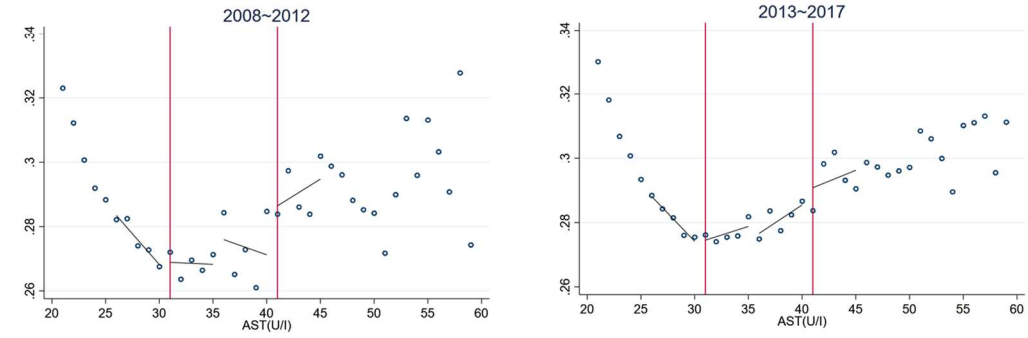
2. Whether take dinner late (1=yes)



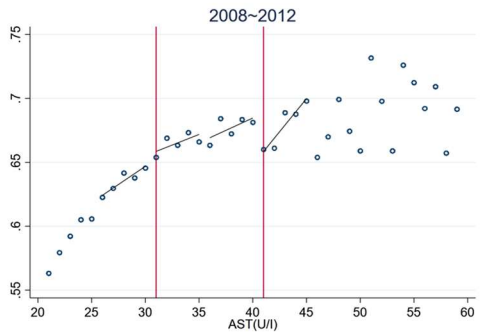
3. Whether drink everyday (1=yes)

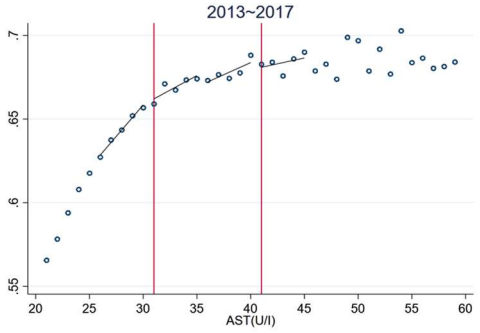


4. Whether doesn't drink (1=yes)

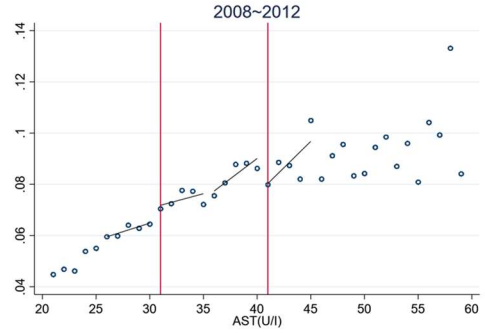


5. Whether drink more than 180ml on a drinking day (1=yes)





6. Whether drink more than 540ml on a drinking day (1=yes)



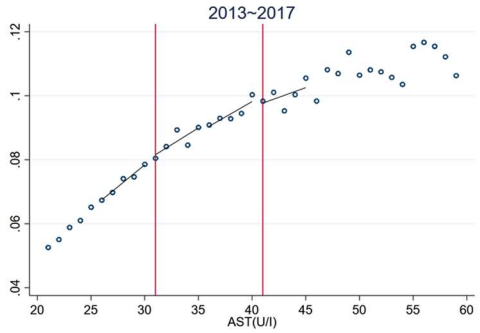
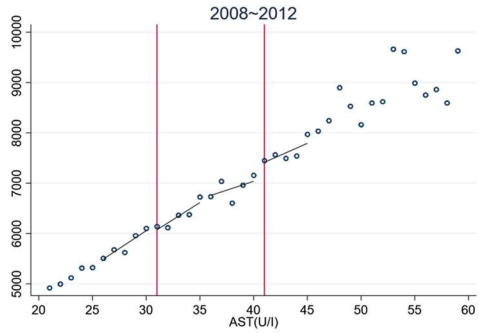
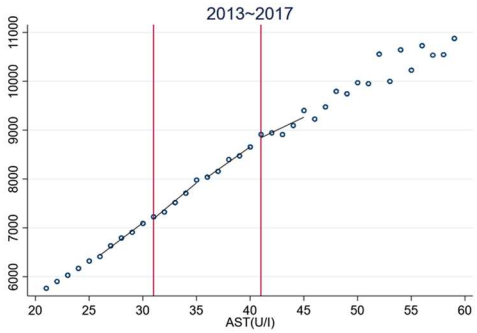


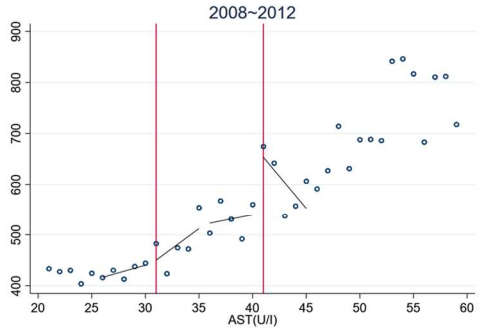
Figure 5: Averages of healthcare usage in the short run by AST

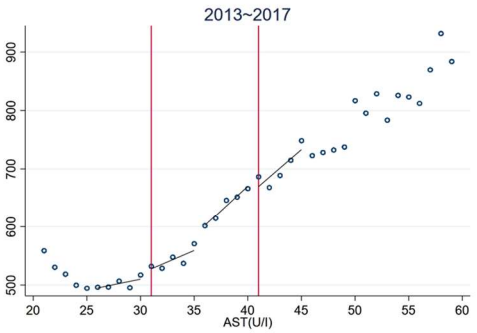
1. Medical expenses on biological tests (yen)



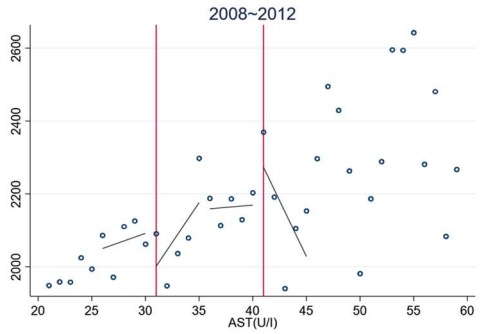


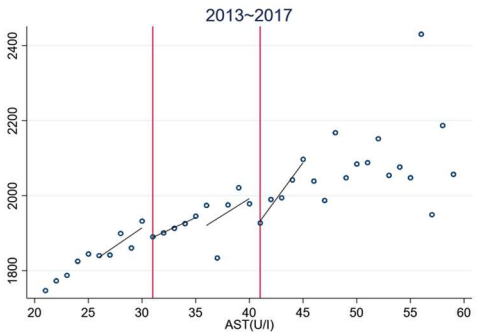
2. Medical expenses on abdominal ultrasound tests (yen)



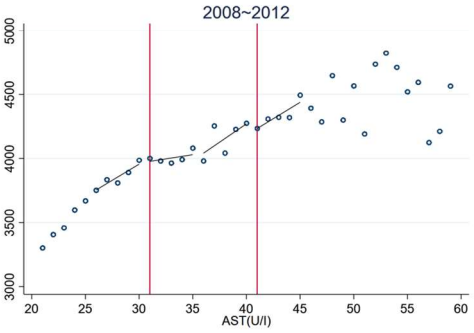


3. Medical expenses on CT/MRI tests (yen)





4. Medical expenses on drug prescription (yen)



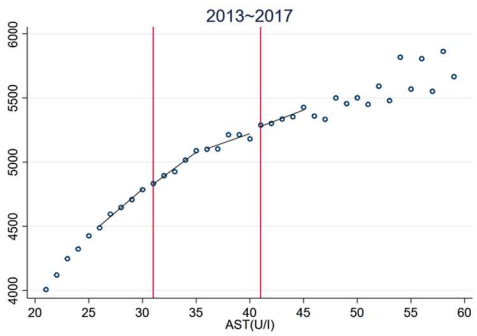
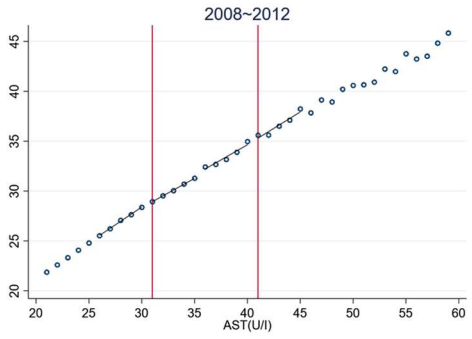
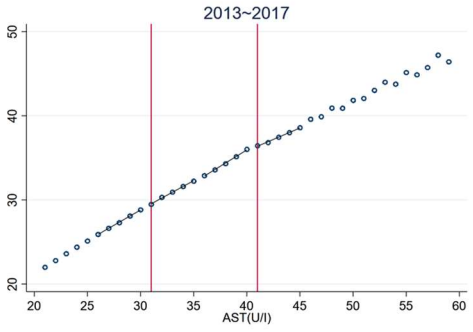


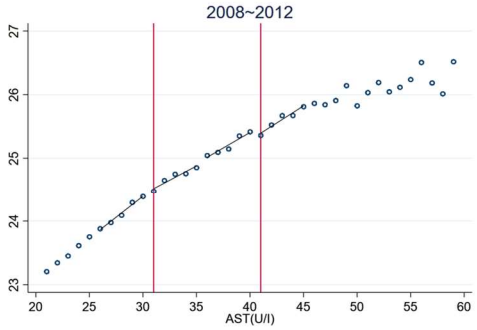
Figure 6: Averages of health outcomes in the short run by AST

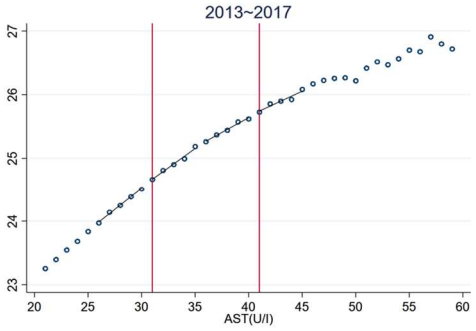
1. Aspartate Aminotransferase (AST) (U/l)



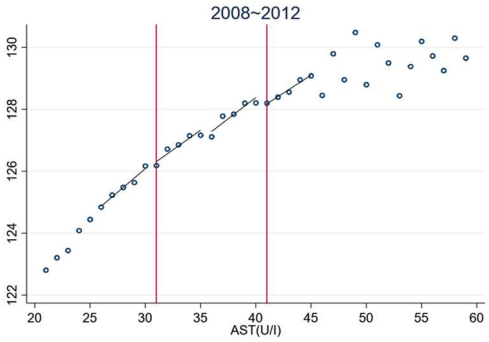


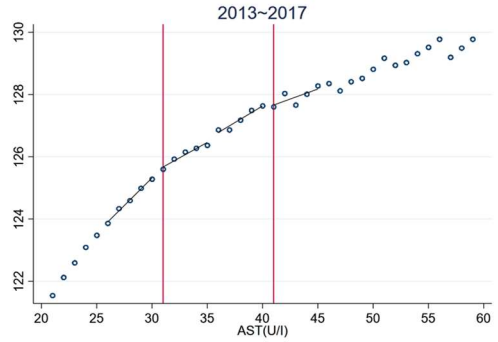
2. Body mass index (kg/m2)



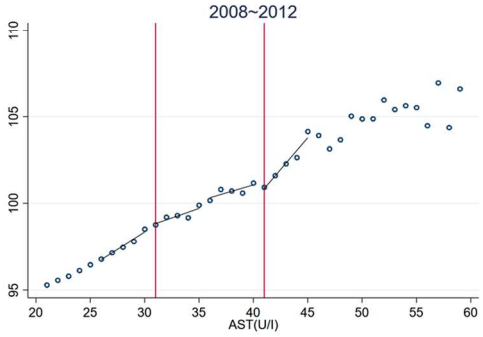


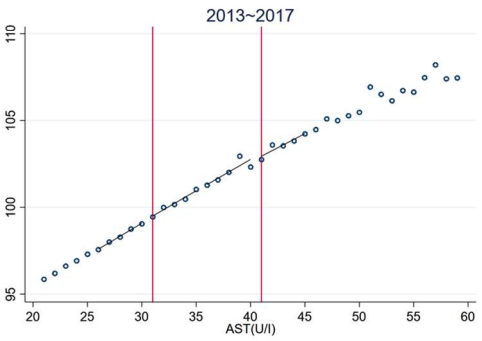
3. Systolic blood pressure (mmHg)



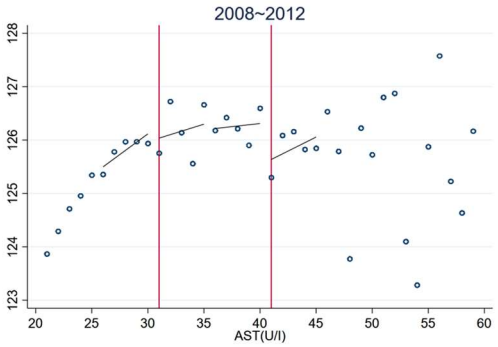


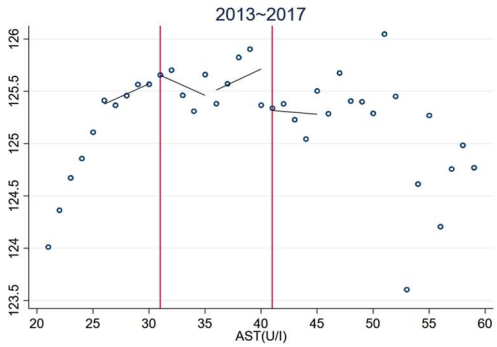
4. Fasting blood sugar (mg/dL)



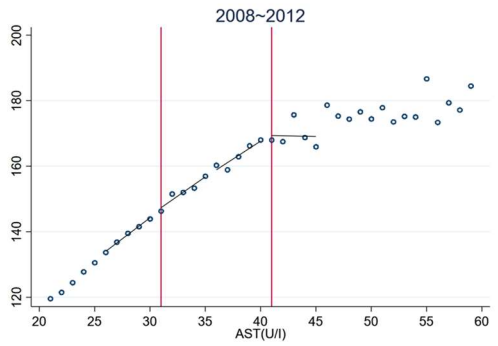


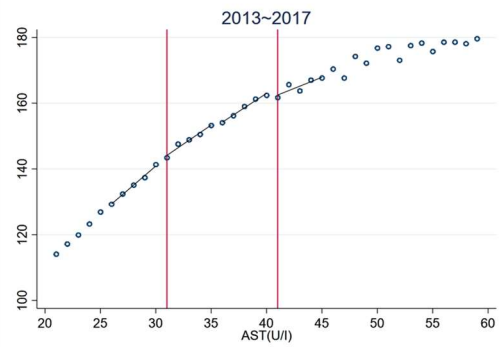
5. Cholesterol (mg/dL)



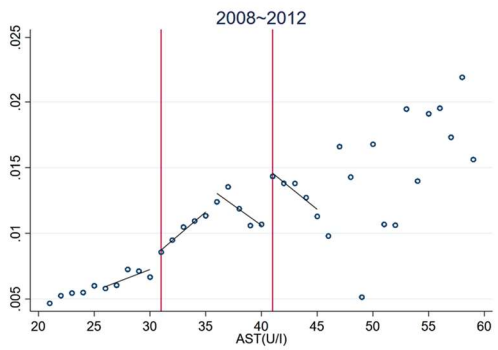


6. Triglyceride (mg/dL)





7. Prob. of liver diseases diagnosis in following 12 months



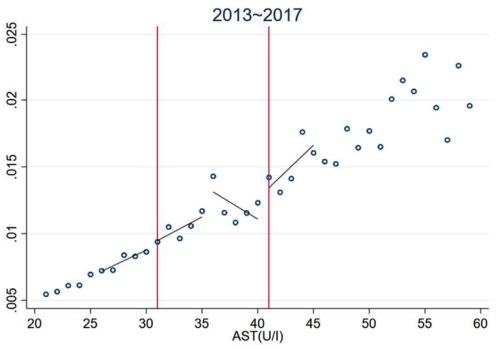
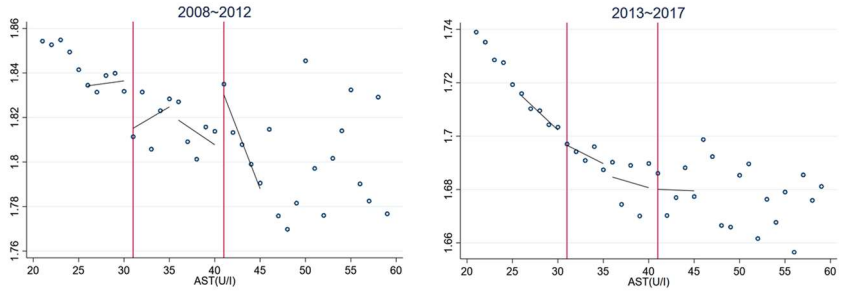
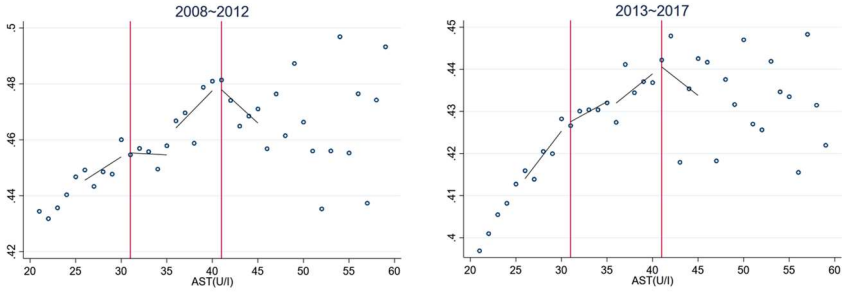


Figure 7: Averages of major outcomes in the long run by AST

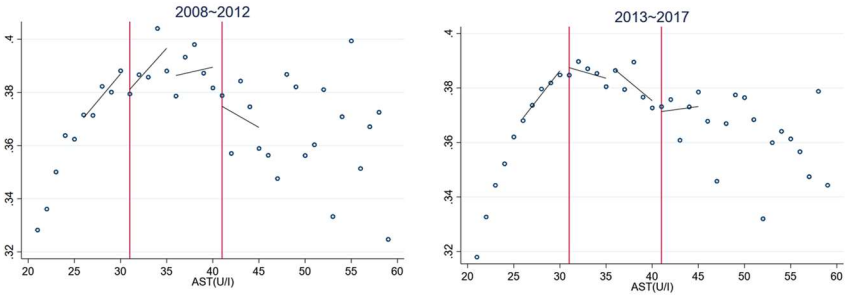
1. Self-rated eating speed score



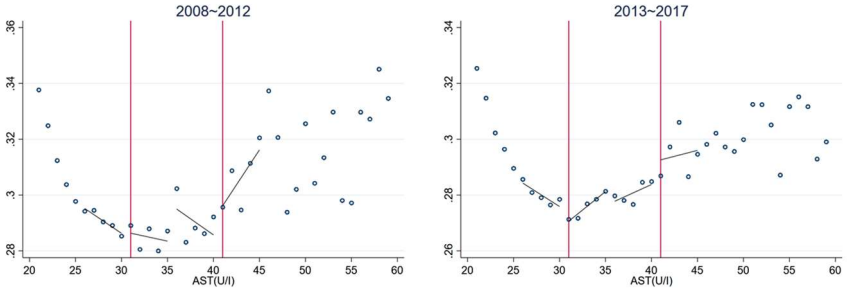
2. Whether take dinner late (1=yes)



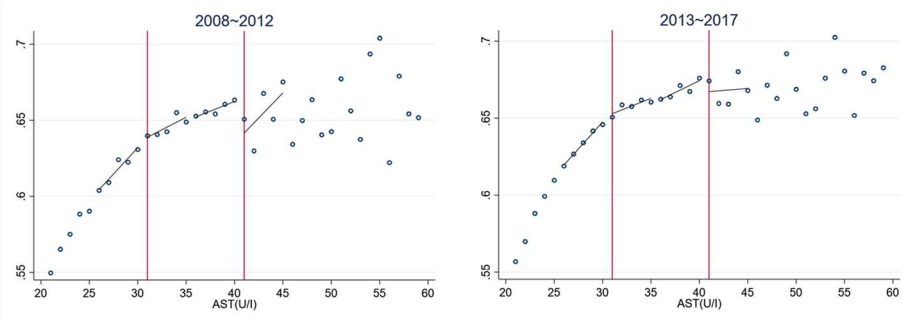
3. Whether drink everyday (1=yes)



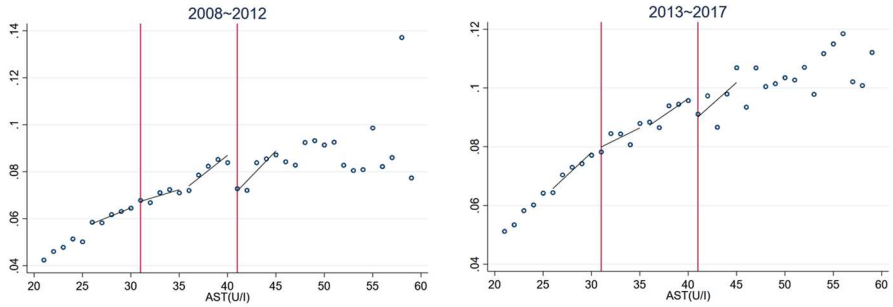
4. Whether doesn't drink (1=yes)



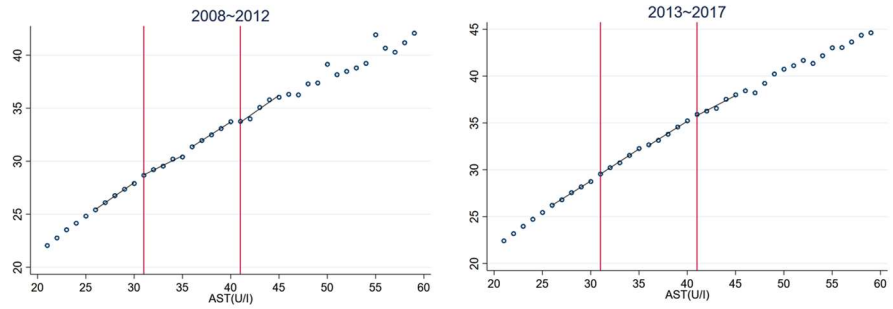
5. Whether drink more than 180ml on a drinking day (1=yes)



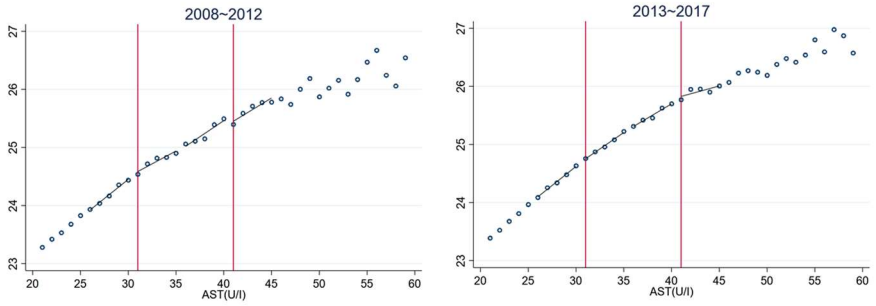
6. Whether drink more than 540ml on a drinking day (1=yes)



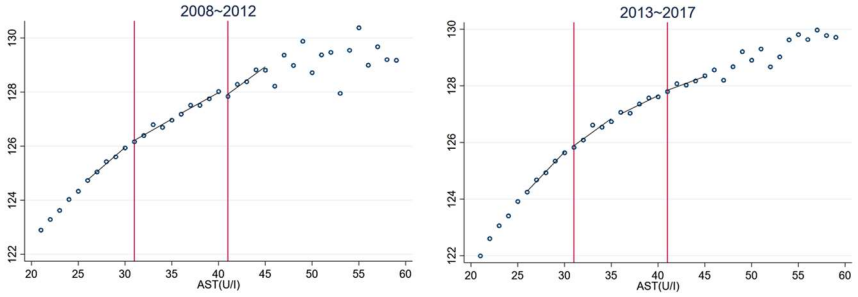
7. Aspartate Aminotransferase (AST) (U/l)



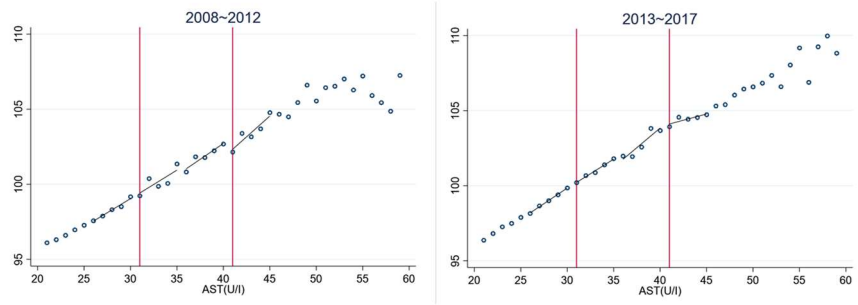
8. Body mass index (kg/m2)



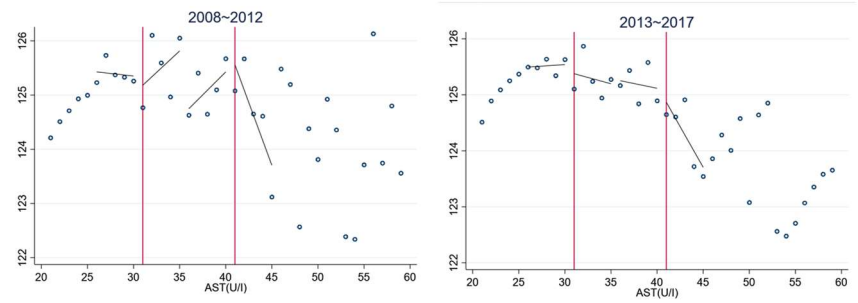
9. Systolic blood pressure (mmHg)



10. Fasting blood sugar (mg/dL)



11. Cholesterol (mg/dL)



12. Triglyceride (mg/dL)

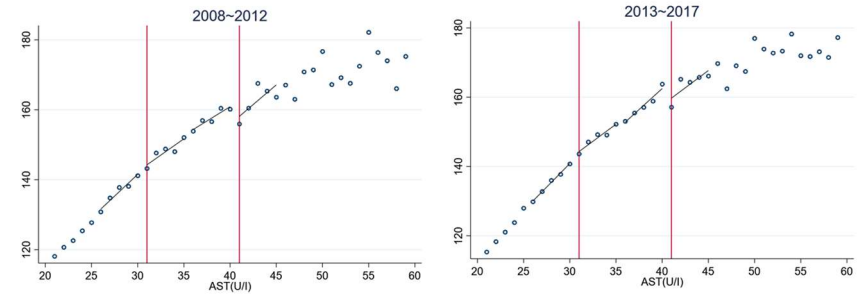


Figure 8: Histogram of AST and McCrary density tests at 31U/I and 41U/I (bandwidth = 5)

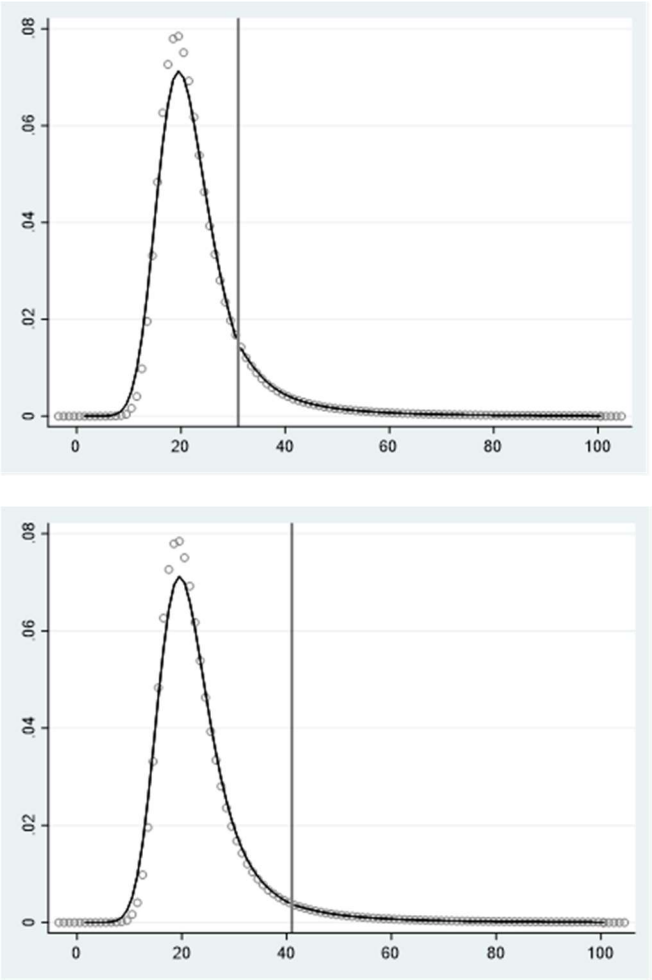


Figure 9: Cumulative distribution of t-statistics from McCrary tests at placebo cutoffs

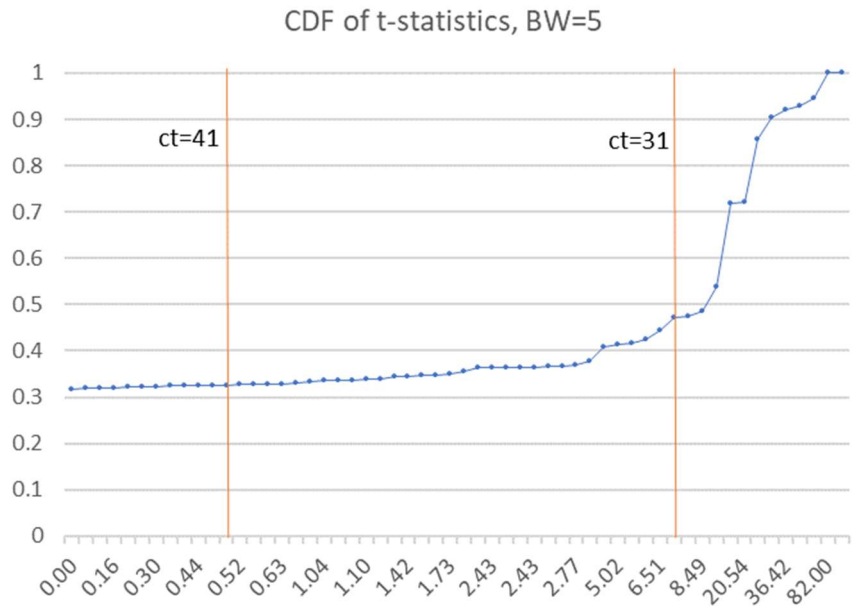
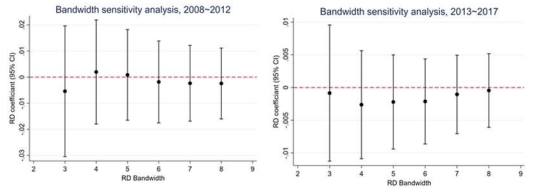
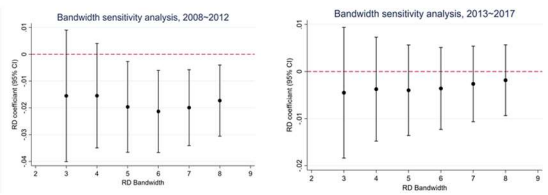


Figure 10: RD estimate sensitivity to bandwidth for key outcomes

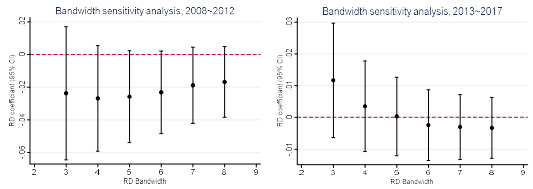
Self-rated eating speed score at cutoff = 31 (short-run)



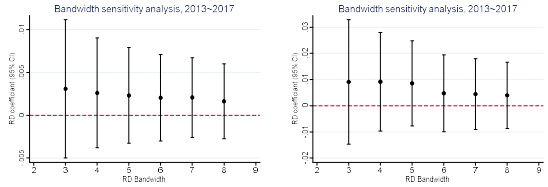
Self-rated eating speed score at cutoff = 31 (long-run)



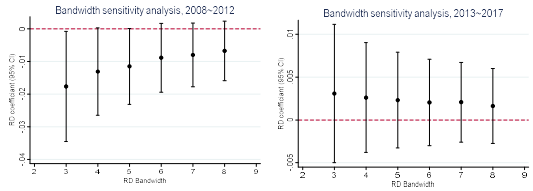
Whether take dinner late at cutoff = 41 (short-run)



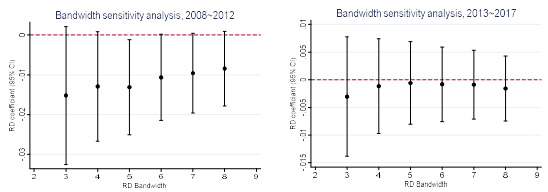
Whether take dinner late at cutoff = 41 (long-run)



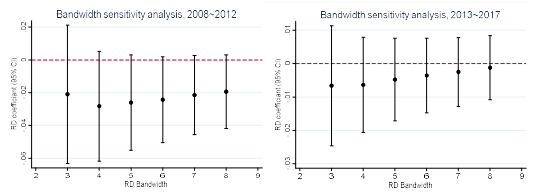
Whether drink everyday at cutoff = 31 (short-run)



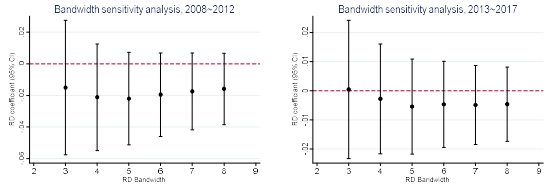
Whether drink everyday at cutoff = 31 (long-run)



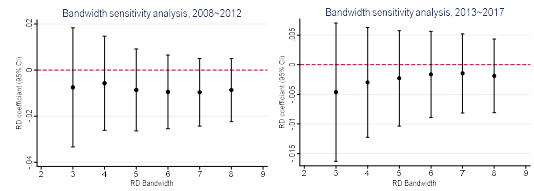
Whether drink more than 180ml at cutoff = 41 (short-run)



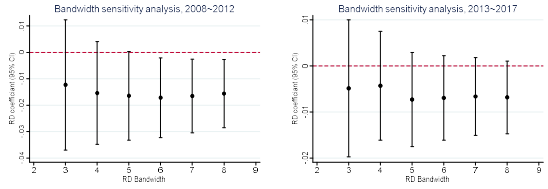
Whether drink more than 180ml at cutoff = 41 (long-run)



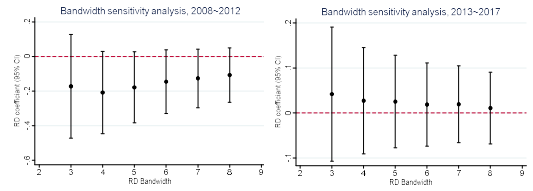
Whether drink more than 540ml at cutoff = 41 (short-run)



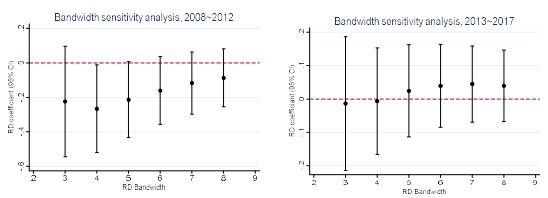
Whether drink more than 540ml at cutoff = 41 (long-run)



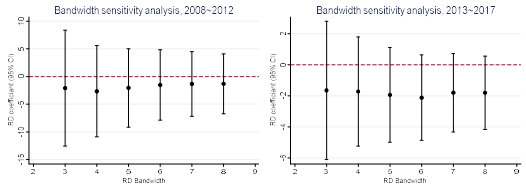
Body mass index (kg/m2) at cutoff = 41 (short-run)



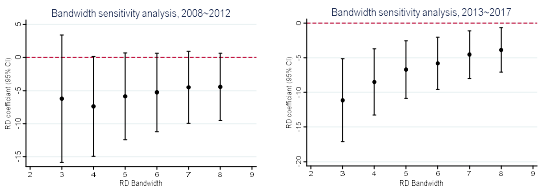
Body mass index (kg/m2) at cutoff = 41 (long-run)



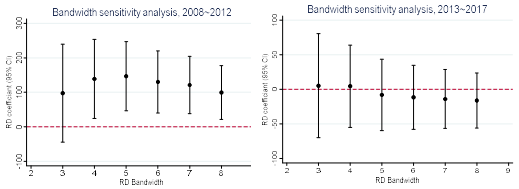
Triglyceride (mg/dL) at cutoff = 41 (short-run)



Triglyceride (mg/dL) at cutoff = 41 (long-run)



Medical expenses on abdominal ultrasound tests at cutoff = 41 (short-run)



Prob. of liver diseases diagnosis in following 12 months at cutoff = 41 (short-run)

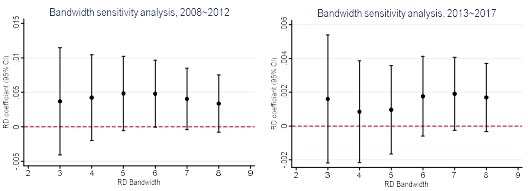
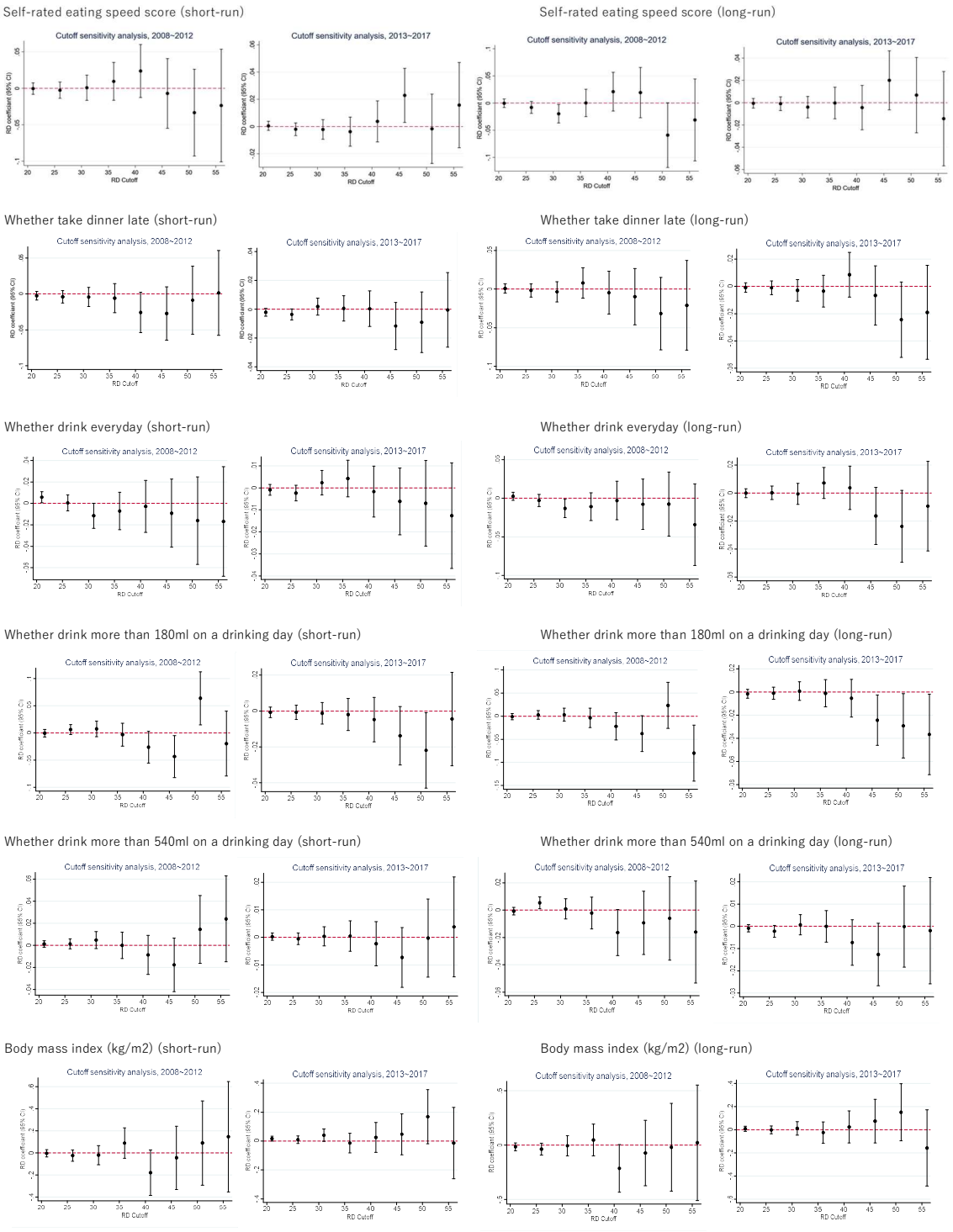
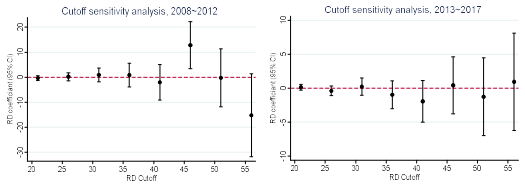


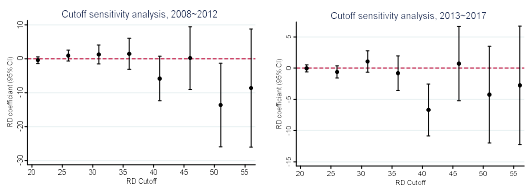
Figure 11: RD estimate sensitivity to cutoff for key outcomes (95% C.I.), bandwidth = 5



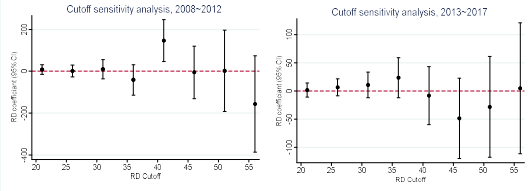
Triglyceride (mg/dL) (short-run)



Triglyceride (mg/dL) (long-run)



Medical expenses on abdominal ultrasound tests (short-run)



Prob. of liver disease diagnosis in following 12 months (short-run)

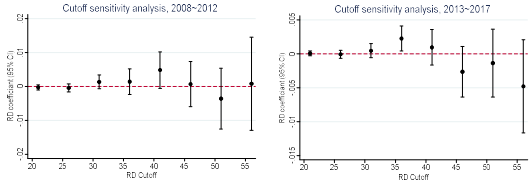


Table 1: Sample size and attrition rates by year

	No. of observations	No. of obs. with short-run info	Short-run attrition rate	No. of obs. with long-run info	Long-run attrition rate
2008	57,252	53,342	0.07	49,946	0.13
2009	133,616	126,535	0.05	117,730	0.12
2010	206,415	192,649	0.07	175,953	0.15
2011	289,897	270,938	0.07	247,598	0.15
2012	449,305	411,790	0.08	384,777	0.14
2013	720,973	675,910	0.06	620,543	0.14
2014	817,786	768,491	0.06	707,921	0.13
2015	1,101,035	1,029,632	0.06	905,077	0.18
2016	1,309,468	1,227,503	0.06	733,272	0.44
2017	1,570,050	1,389,195	0.12	129,764	0.92
2018	215,342	180,137	0.16	-	-

Notes:
1. The sample reported here is created based on steps 1-4 described in Figure 1, for April of 2008 to March of 2018.

Table 2: Descriptive statistics

	Full[1]			2008~2012			2013~2017		
	Obs.	Mean	S.D.	Obs.	Mean	S.D.	Obs.	Mean	S.D.
Age	6,871,139	46.9	8.4	1,216,731	45.8	8.5	5,654,408	47.2	8.4
Sex (1=male)	6,871,139	0.811	0.392	1,216,731	0.866	0.340	5,654,408	0.799	0.401
Whether diagnosed of liver diseases (1=yes)[2]	6,871,139	0.073	0.260	1,216,731	0.062	0.242	5,654,408	0.075	0.264
Health behaviors									
Self-reated eating speed[3]	5,270,184	1.760	0.606	776,442	1.829	0.620	4,493,742	1.748	0.602
Whether take dinner late (1=yes)[4]	5,241,732	0.405	0.491	788,829	0.408	0.491	4,452,903	0.404	0.491
Whether drink everyday (1=yes)	5,813,058	0.298	0.457	1,023,208	0.307	0.461	4,789,850	0.296	0.456
Whether doesn't drink (1=yes)	5,813,058	0.348	0.476	1,023,208	0.338	0.473	4,789,850	0.350	0.477
Whether drink more than 180ml on a drinking day (1=yes)	4,328,026	0.570	0.495	660,850	0.564	0.496	3,667,176	0.572	0.495
Whether drink more than 540ml on a drinking day (1=yes)	4,328,026	0.057	0.232	660,850	0.049	0.215	3,667,176	0.059	0.235
Health oucomes									
Aspartate Aminotransferase (AST) (U/l)	6,552,876	23.1	11.6	1,143,662	23.0	11.5	5,409,214	23.2	11.6
Body mass index (kg/m2)	6,723,766	23.4	3.6	1,207,835	23.4	3.5	5,515,931	23.5	3.7
Systolic blood pressure (mmHg)	6,728,352	121.8	15.6	1,209,294	122.9	15.2	5,519,058	121.6	15.7
Fasting blood sugar (mg/dL)	5,342,063	96.7	18.9	952,896	96.2	19.7	4,389,167	96.8	18.7
Cholesterol (mg/dL)	6,548,168	122.3	30.8	1,142,250	122.5	30.9	5,405,918	122.3	30.8
Triglyceride (mg/dL)	6,545,366	119.6	95.7	1,145,903	123.9	96.9	5,399,463	118.7	95.4

Notes:
1. The sample reported here is created based on steps 1-4 described in Figure 1.
2. Liver diseases in past 12 months, including all types of hepatitis, liver fat, liver disorders and liver cirrhosis.
3. Codes for self-rated eating speed: 1: fast; 2: average; 3: slow.
4. Defined as whether take dinner within 2 hours before bed for more than 3 times a week.

Table 3: Local polynomial regression estimates for health behaviors

	Cutoff=41							
	Short-run impact				Long-run impact			
	2008~2012		2013~2017		2008~2012		2013~2015	
Eating behaviors	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.
Self-reated eating speed	21,545	0.024	115,089	0.004	22,925	0.018	65,787	-0.004
		(0.019)		(0.008)		(0.018)		(0.010)
Whether take dinner late (1=yes)	21,999	-0.026 *	113,684	0.000	23,097	-0.005	65,477	0.009
		(0.014)		(0.006)		(0.014)		(0.008)
Drinking behaviors								
Whether drink everyday (1=yes)	27,145	-0.003	121,076	-0.002	26,293	-0.001	68,774	0.004
		(0.012)		(0.006)		(0.013)		(0.008)
Whether doesn't drink (1=yes)	27,145	0.001	121,076	-0.001	26,293	-0.003	68,774	0.001
		(0.012)		(0.006)		(0.012)		(0.007)
Whether drink more than 180ml on a drinking day	17,720	-0.026 *	98,167	-0.005	18,713	-0.017	58,463	-0.006
		(0.015)		(0.006)		(0.015)		(0.008)
Whether drink more than 540ml on a drinking day	17,720	-0.009	98,167	-0.002	18,713	-0.016 *	58,463	-0.009 *
		-(0.009)		(0.004)		(0.009)		(0.005)
	Cutoff=31							
Eating behaviors	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.
Self-reated eating speed	95,875	0.001	507,427	-0.002	102,172	-0.019 **	290,400	-0.004
		(0.009)		(0.004)		(0.009)		(0.005)
Whether take dinner late (1=yes)	98,134	-0.004	501,256	0.002	102,978	-0.005	288,899	-0.003
		(0.007)		(0.003)		(0.007)		(0.004)
Drinking behaviors								
Whether drink everyday (1=yes)	121,457	-0.012 **	534,798	0.002	117,207	-0.013 **	303,482	0.000
		(0.006)		(0.003)		(0.006)		(0.004)
Whether doesn't drink (1=yes)	121,457	0.005	534,798	0.002	117,207	0.004	303,482	-0.007 **
		(0.005)		(0.003)		(0.006)		(0.003)
Whether drink more than 180ml on a drinking day	80,279	0.008	436,100	-0.001	84,335	0.004	258,974	0.001
		(0.007)		(0.003)		(0.007)		(0.004)
Whether drink more than 540ml on a drinking day	80,279	0.005	436,100	0.000	84,335	0.002	258,974	0.001
		(0.004)		(0.002)		(0.004)		(0.002)

Notes:
1. Robust standard errors, clustered at individual level, are reported in parentheses.
2. All regressions include covariates..
3. Triangle kernel is used to weight the sample.
4. * Statistical significance at the 10% level; ** Statistical significance at the 5% level; *** Statistical significance at the 1% level.

Table 4: Local polynomial regression estimates for healthcare usage

	Cutoff=41			
	Short-run impact			
	2008~2012		2013~2017	
	Obs.	Coef.	Obs.	Coef.
Medical expenses on biological tests (yen)	30,285	303.8	138,549	130.1
		(333.6)		(170.0)
Medical expenses on abdominal ultrasound tests (yen)	30,285	146.4 ***	138,549	-8.0
		(51.1)		(26.3)
Medical expenses on CT/MRI tests (yen)	30,285	191.7	138,549	-111.4
		(179.8)		(74.4)
Medical expenses on drug prescription (yen)	30,285	-29.6	138,549	74.5
		(131.6)		(70.0)
	Cutoff=31			
Medical expenses on biological tests (yen)	135,058	-168.0	610,971	-27.7
		(144.1)		(74.8)
Medical expenses on abdominal ultrasound tests (yen)	135,058	10.0	610,971	10.9
		(23.5)		(11.7)
Medical expenses on CT/MRI tests (yen)	135,058	-69.8	610,971	-44.1
		(82.0)		(35.3)
Medical expenses on drug prescription (yen)	135,058	-36.4	610,971	-2.0
		(61.2)		(32.4)

Notes:
1. Robust standard errors, clustered at individual level, are reported in parentheses.
2. All regressions include covariates.
3. Triangle kernel is used to weight the sample.
4. * Statistical significance at the 10% level; ** Statistical significance at the 5% level; *** Statistical significance at the 1% level.
5. CT/MRI tests on head, limbs, breast, colon and blood vessel are excluded.
6. Expenses on drug prescription are for all kinds of diseases and do not include drug costs.

Table 5: Local polynomial regression estimates for health outcomes

	Cutoff=41							
	Short-run impact				Long-run impact			
	2008~2012		2013~2017		2008~2012		2013~2015	
	Obs.	Coef.	Obs.	Coef.				
Aspartate Aminotransferase (AST) (U/l)	29,866	-0.135	135,136	-0.269	27,679	-0.7575 *	76,304	-0.00931
		(0.452)		(0.209)		(0.422)		(0.275)
Body mass index (kg/m2)	30,010	-0.179 *	136,197	0.0252	27,724	-0.20212 *	76,520	0.02599
		(0.105)		0.05246		(0.111)		(0.070)
Systolic blood pressure (mmHg)	30,042	-0.1945	136,322	-0.152	27,723	-0.29929	76,524	-0.01286
		0.37498		(0.183)		(0.379)		(0.244)
Fasting blood sugar (mg/dL)	23,383	-0.349	112,508	0.035	21,957	-0.57236	66,698	-0.15289
		(0.658)		(0.295)		(0.704)		(0.402)
Cholesterol (mg/dL)	28,462	-0.987	135,027	-0.131	27,651	-0.1194	76,263	-0.15566
		(0.855)		(0.396)		(0.866)		(0.518)
Triglyceride (mg/dL)	28,459	-2.041	134,938	-1.941	27,665	-5.00877	76,287	-6.52015 ***
		(3.619)		(1.563)		(3.330)		(2.111)
Prob. of liver diseases diagnosis[5]	30,285	0.005 *	138,549	0.001				
		(0.003)		(0.001)				
	Cutoff=31							
Aspartate Aminotransferase (AST) (U/l)	132,997	-0.074	596,502	-0.034	123,116	0.24407	336,556	0.12651
		(0.134)		(0.064)		(0.165)		(0.094)
Body mass index (kg/m2)	133,952	-0.020	601,305	0.040 *	123,331	-0.00323	337,626	0.00678
		(0.044)		(0.022)		(0.047)		(0.029)
Systolic blood pressure (mmHg)	134,106	-0.112	601,803	0.096	123,321	0.03154	337,641	-0.09038
		(0.175)		(0.086)		(0.180)		(0.115)
Fasting blood sugar (mg/dL)	104,528	0.032	496,123	0.087	98,152	-0.09061	294,074	-0.00995
		(0.277)		(0.116)		(0.275)		(0.157)
Cholesterol (mg/dL)	126,731	0.018	596,152	0.066	123,019	-0.14219	336,408	-0.25888
		(0.389)		(0.180)		(0.396)		(0.239)
Triglyceride (mg/dL)	126,703	0.940	595,823	0.225	123,051	1.52022	336,502	1.03762
		(1.421)		(0.650)		(1.416)		(0.878)
Prob. of liver diseases diagnosis[5]	135,058	0.001	610,971	0.000				
		(0.001)		(0.001)				

Notes:
1. Robust standard errors, clustered at individual level, are reported in parentheses.
2. All regressions include covariates.
3. Triangle kernel is used to weight the sample.
4. * Statistical significance at the 10% level; ** Statistical significance at the 5% level; *** Statistical significance at the 1% level.
5. Probability of major liver diseases diagnosed in the following 12 months after the health checkup.

Table 6: McCrary test results

	Bandwidth						
Cutoff	2	3	4	5	6	7	8
26	2.359	2.101	1.121	-1.572	-7.578	-18.675	-37.222
27	-1.598	-0.902	-0.439	-1.572	-1.469	-5.326	-14.203
28	0.830	0.301	0.886	2.541	4.587	6.070	5.274
29	-0.276	0.838	2.411	5.054	8.978	13.966	19.220
30	1.400	2.930	5.296	8.495	13.413	20.505	29.460
31	**0.938**	**1.898**	**4.017**	**7.994**	**13.871**	**22.123**	**33.199**
32	-1.360	-0.722	1.424	5.484	11.846	20.691	32.733
33	2.041	2.875	4.603	7.847	13.358	21.970	34.005
34	0.032	1.715	3.627	6.513	11.451	19.241	30.823
35	0.537	0.376	1.780	4.656	9.126	16.128	26.534
36	-1.115	-0.469	0.645	-0.469	7.330	13.594	22.721
37	2.139	2.769	3.773	2.769	8.941	14.285	22.115
38	-0.762	0.096	1.620	0.096	6.498	11.027	17.953
39	0.182	0.403	1.216	0.403	5.374	9.214	15.056
40	0.982	1.045	1.162	1.045	4.343	7.794	12.820
41	**-1.533**	**-1.287**	**-0.534**	**0.518**	**2.349**	**5.446**	**10.002**
42	1.543	1.698	2.172	3.135	4.641	7.057	10.798
43	-0.244	0.616	1.352	2.426	4.016	6.133	9.182
44	0.124	-0.112	0.586	2.426	2.910	4.821	7.439
45	-0.315	0.136	0.244	2.426	1.825	3.378	5.687
46	1.085	0.823	0.827	2.426	1.457	2.657	4.489
47	-1.671	-1.401	-1.226	2.426	-0.338	0.601	2.269
48	1.665	1.219	1.233	1.409	1.794	2.511	3.802
49	-1.307	-0.366	0.057	0.435	1.118	2.066	3.302
50	1.651	1.090	1.211	1.733	2.401	3.336	4.595

Notes:
1 T-statistics for the cutoffs used for the analysis are shown in bold.
2. Estimates are obtained by the DCdensity command in Stata provided by McCrary at https://eml.berkeley.edu/~jmccrary/DCdensity/.

Table 7: Checks on the continuity of covariates

	Cutoff = 31				Cutoff = 41			
	2008~2012		2013~2017		2008~2012		2013~2017	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
Age	0.060	0.524	-0.031	0.479	-0.217	0.260	-0.085	0.347
Sex (1=male)	-0.001	0.590	-0.001	0.453	-0.005	0.364	-0.003	0.403
Health behaviors								
Self-reated eating speed[3]	-0.001	0.935	0.001	0.879	0.036	**0.062**	0.007	0.379
Whether take dinner late (1=yes)[4]	0.002	0.794	0.003	0.406	-0.014	0.371	0.004	0.512
Whether drink everyday (1=yes)	-0.008	0.174	-0.002	0.505	0.006	0.632	-0.001	0.850
Whether doesn't drink (1=yes)	0.004	0.425	0.001	0.710	0.011	0.347	0.000	0.944
Whether drink more than 180ml on a drinking day (1=yes)	-0.007	0.328	0.001	0.836	-0.007	0.660	-0.006	0.396
Whether drink more than 540ml on a drinking day (1=yes)	0.003	0.441	-0.003	0.151	0.002	0.864	-0.003	0.465
Health oucomes								
Body mass index (kg/m2)	-0.006	0.890	0.044	**0.043**	-0.185	**0.082**	0.044	0.400
Systolic blood pressure (mmHg)	-0.092	0.603	0.028	0.748	-0.171	0.651	-0.186	0.313
Fasting blood sugar (mg/dL)	0.262	0.786	0.049	0.180	-0.296	0.731	-0.266	0.580
Cholesterol (mg/dL)	-0.291	0.461	0.063	0.726	-0.514	0.566	-0.450	0.269
Triglyceride (mg/dL)	1.270	0.391	-0.653	0.329	-3.322	0.381	-2.280	0.181

Notes:
1. P-values that indicate statically significance at the 10% and 5% levels are shown in bold.
2. Triangle kernel is used to weight the sample.
3. All covariates in **X** in Eq.(1) are excluded from the explanatory variables.

Table 8: Local linear regression estimates for heterogenous short-run impact (cutoff=41)

	Cutoff=41											
	Male sample				Female sample				High risk group defined by the Metabolic Syndrom Guidelines			
	2008~2012		2013~2017		2008~2012		2013~2017		2008~2012		2013~2017	
	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.
Eating behaviors												
Self-reated eating speed	20,476	0.027	105,952	0.001	1,069	-0.021	9,137	0.032	4,918	0.009	27,493	0.002
Whether take dinner late (1=yes)	20,918	-0.027 *	104,584	0.003	1,081	-0.023	9,100	-0.027	5,031	-0.026	27,126	0.004
Drinking behaviors												
Whether drink everyday (1=yes)	25,992	-0.001	111,738	-0.002	1,153	-0.029	9,338	0.003	6,030	-0.022	28,646	-0.008
Whether doesn't drink (1=yes)	25,992	0.001	111,738	-0.002	1,153	0.003	9,338	0.001	6,030	-0.011	28,646	0.013
Whether drink more than 180ml on a drinking day	17,049	-0.025 *	91,610	-0.004	671	-0.040	6,557	-0.021	4,087	-0.063 **	22,986	-0.018
Whether drink more than 540ml on a drinking day	17,049	-0.009	91,610	-0.002	671	0.001	6,557	-0.007	4,087	-0.049 ***	22,986	-0.003
Usage of healthcare services												
Medical expenses on biological tests (yen)	29,000	295.4	127,482	70.7	1,285	564.3	11,067	810.7	6,743	388.9	32,970	575.5
Medical expenses on abdominal ultrasound tests (yen)	29,000	114.1 **	127,482	20.0	1,285	973.7 **	11,067	-346.4 *	6,743	99.0	32,970	56.7
Medical expenses on CT/MRI tests (yen)	29,000	186.5	127,482	-177.9 **	1,285	241.8	11,067	671.8 **	6,743	300.1	32,970	-25.5
Medical expenses on drug prescription (yen)	29,000	-64.5	127,482	73.5	1,285	782.9	11,067	77.6	6,743	-181.4	32,970	109.1
Health outcomes in the next time period												
Aspartate Aminotransferase (AST) (U/l)	28,590	-0.163	124,267	-0.392 *	1,276	0.677	10,869	1.151	6,648	0.516	32,308	-0.607
Body mass index (kg/m2)	28,741	-0.195 *	125,306	0.027	1,269	0.174	10,891	0.004	6,642	-0.260	32,415	0.058
Systolic blood pressure (mmHg)	28,774	-0.132	125,422	-0.149	1,268	-2.035	10,900	-0.137	6,653	-0.788	32,465	-0.531
Fasting blood sugar (mg/dL)	22,433	-0.212	103,804	0.030	950	-3.158	8,704	0.117	5,377	-0.709	27,492	0.520
Cholesterol (mg/dL)	27,211	-0.713	124,168	-0.221	1,251	-8.254 **	10,859	1.011	6,399	0.429	32,312	0.101
Triglyceride (mg/dL)	27,208	-2.215	124,091	-2.067	1,251	0.046	10,847	-0.480	6,377	-14.662 *	32,277	-4.295
Prob. of liver diseases diagnosis in following 12 months	29,000	0.005 **	127,482	0.001	1,285	-0.010	11,067	0.001	6,743	0.007	32,970	0.000

Notes:
1. Robust standard errors, clustered at individual level, are reported in parentheses.
2. All regressions include covariates.
3. Triangle kernel is used to weight the sample.
4. * Statistical significance at the 10% level; ** Statistical significance at the 5% level; *** Statistical significance at the 1% level.

Table 9: Local linear regression estimates for heterogenous short-run impact (cutoff=31)

	Cutoff=31											
	Male sample				Female sample				High risk group defined by the Metabolic Syndrom Guidelines			
	2008~2012		2013~2017		2008~2012		2013~2017		2008~2012		2013~2017	
	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.	Obs.	Coef.
Eating behaviors												
Self-reated eating speed	89,769	0.000	457,750	-0.002	6,106	0.015	49,677	-0.006	14,812	0.025	78,536	0.016 *
Whether take dinner late (1=yes)	91,921	-0.004	451,744	0.003	6,213	-0.014	49,512	-0.008	15,091	0.034 *	77,368	-0.002
Drinking behaviors												
Whether drink everyday (1=yes)	114,810	-0.012 **	484,070	0.003	6,647	0.006	50,728	-0.003	18,039	-0.015	81,902	0.004
Whether doesn't drink (1=yes)	114,810	0.006	484,070	0.002	6,647	-0.011	50,728	0.006	18,039	-0.001	81,902	0.001
Whether drink more than 180ml on a drinking day	76,300	0.007	400,183	0.000	3,979	0.021	35,917	-0.017	12,253	0.010	66,447	-0.002
Whether drink more than 540ml on a drinking day	76,300	0.005	400,183	0.000	3,979	-0.008	35,917	0.003	12,253	-0.003	66,447	-0.001
Usage of healthcare services												
Medical expenses on biological tests (yen)	95,266	-348.3 **	412,155	16.3	7,469	1970.3 **	59,638	-253	19,969	-198.4	94,059	-164.4
Medical expenses on abdominal ultrasound tests (yen)	95,266	16.0	412,155	18.37	7,469	148.1	59,638	-59.1	19,969	34.7	94,059	-0.7
Medical expenses on CT/MRI tests (yen)	95,266	6.1	412,155	-59.6	7,469	473.1	59,638	8.0	19,969	-11.5	94,059	-142.1
Medical expenses on drug prescription (yen)	95,266	-84.1	412,155	0.9	7,469	245.4	59,638	-60.9	19,969	37.4	94,059	-63.1
Health outcomes in the next time period												
Aspartate Aminotransferase (AST) (U/l)	125,614	-0.041	537,864	-0.002	7,383	-0.658	58,638	-0.361 *	19,682	0.506	92,242	-0.106
Body mass index (kg/m2)	126,570	-0.034	542,524	0.040 *	7,382	0.262	58,781	0.039	19,694	-0.080	92,564	-0.057
Systolic blood pressure (mmHg)	126,721	-0.059	542,998	0.113	7,385	-1.056	58,805	-0.074	19,740	-1.043 **	92,670	-0.035
Fasting blood sugar (mg/dL)	98,811	-0.061	448,961	0.106	5,717	1.742	47,162	-0.145	15,772	0.224	77,990	0.320
Cholesterol (mg/dL)	119,428	0.018	537,522	0.057	7,303	0.080	58,630	0.360	18,938	0.761	92,247	0.143
Triglyceride (mg/dL)	119,435	1.013	537,246	0.316	7,268	-0.206	58,577	-0.731	18,887	-4.876	92,190	0.288
Prob. of liver diseases diagnosis in following 12 months	127,589	0.001	551,333	0.001	7,469	0.003	59,638	0.007	19,969	0.000	94,059	-0.002

Notes:
1. Robust standard errors, clustered at individual level, are reported in parentheses.
2. All regressions include covariates.
3. Triangle kernel is used to weight the sample.
4. * Statistical significance at the 10% level; ** Statistical significance at the 5% level; *** Statistical significance at the 1% level.

Appendix:

Table A1: Standard questionnaire of specific health checkup

	English Translation of question (original in Japanese)	Answer
1	Medication to lower blood pressure	Yes / No
2	Insulin shots or medication to lower blood sugar level	Yes / No
3	Medication to lower cholesterol level	Yes / No
4	Have you been diagnosed or treated with stroke by a doctor ?	Yes / No
5	Have you been diagnosed or treated with heart disease by a doctor?	Yes / No
6	Have you been diagnosed or treated with chronic kidney failure by a doctor or gotten dialysis?	Yes / No
7	Have you been diagnosed with anemia?	Yes / No
8	Do you habitually smoke currently?	Yes / No
9	Have you gained weight 10kg or more since 20 years of age?	Yes / No
10	Do you engage in lightly sweating physical exercise with the duration of at least 30 minutes two or more times per week for at least a year?	Yes / No
11	Do you engage in walking or similar physical activity at least an hour per day in your daily life?	Yes / No
12	Do you walk faster than other people of the same age and sex as you?	Yes / No
13	Have you gained or lost weight 3 kg or more in a year?	Yes / No
14	Do you eat faster than other people?	Fast / Normal / Slow
15	Do you eat dinner within two hours before going to bed three times or more per week?	Yes / No
16	Do you snack after dinner three times or more per week?	Yes / No
17	Do you skip breakfast three times or more per week?	Yes / No
18	How often do you drink alcohol?	Every day / Occasionally / Almost no drinking (including not being able to drink)
19	How much do you drink per day on the day of drinking?	Less than 180ml / 180- 360ml / 360-540ml / 540ml or more
20	Do you take rest enough by sleep?	Yes / No
21	Do you intend to engage in improving your lifestyle such as physical exercise and dietary habits?	1. Not intending to do so 2. Intending to do so in six months 3. Intending to do so in a month or already engaged in doing so little by little 4. Already engaged in doing so (less than six months) 5. Already engaged in doing so (six months or more)
22	Will you utilize health guidance to improve your lifestyle if you have opportunity to do so ?	Yes / No

Source: MHLW, 2013a.