

NBER WORKING PAPER SERIES

IS PREVENTIVE CARE WORTH THE COST? EVIDENCE FROM MANDATORY  
CHECKUPS IN JAPAN

Toshiaki Iizuka  
Katsuhiko Nishiyama  
Brian Chen  
Karen Eggleston

Working Paper 23413  
<http://www.nber.org/papers/w23413>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2017

We are grateful to Hyuncheol Kim and seminar participants at the 2016 EuHEA conference in Hamburg for helpful comments and suggestions. Financial support from the Kikawada Foundation is gratefully acknowledged. All remaining errors are our own. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The study of which this research was a part was reviewed and approved by the Stanford University IRB. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Toshiaki Iizuka, Katsuhiko Nishiyama, Brian Chen, and Karen Eggleston. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Is Preventive Care Worth the Cost? Evidence from Mandatory Checkups in Japan  
Toshiaki Iizuka, Katsuhiko Nishiyama, Brian Chen, and Karen Eggleston  
NBER Working Paper No. 23413  
May 2017  
JEL No. I10

**ABSTRACT**

Using unique individual-level panel data, we investigate whether preventive medical care triggered by health checkups is worth the cost. We exploit the fact that biomarkers just below and above a threshold may be viewed as random. We find that people respond to health signals and increase physician visits. However, we find no evidence that additional care is cost effective. For the “borderline type” (“pre-diabetes”) threshold for diabetes, medical care utilization increases but neither physical measures nor predicted risks of mortality or serious complications improve. For efficient use of medical resources, cost effectiveness of preventive care must be carefully examined.

Toshiaki Iizuka  
Graduate School of Economics  
The University of Tokyo  
Japan  
Bunkyo-ku, Tokyo 113-0033  
Japan  
toshi.iizuka@gmail.com

Katsuhiko Nishiyama  
Mizuho Information & Research Institute  
Japan  
katsu.nishiyama@gmail.com

Brian Chen  
Department of Health Services Policy  
and Management  
Arnold School of Public Health  
University of South Carolina  
915 Greene Street #354  
Columbia, SC 29205  
bchen@mailbox.sc.edu

Karen Eggleston  
Shorenstein Asia-Pacific Research Center  
Stanford University  
616 Serra Street  
Stanford, CA 94305  
and NBER  
karene@stanford.edu

## 1. Introduction

Prevention of chronic disease has become a key health policy initiative in recent years. For example, the World Health Organization (WHO) provides a road map and menu of policy options that aim to reduce premature deaths due to chronic non-communicable diseases such as cardiovascular disease, cancer, and diabetes (WHO 2013). An important part of prevention is monitoring an individual's health condition and intervening early enough to make a difference in the course of a disease. Traditional approaches include routine health checkups, cancer screening, and disease management programs. More recently, wearable and portable devices are gaining popularity, allowing people to monitor their own health in real time. Advocates suggest that such real-time health signals will lead to appropriate preventive care and improve health outcomes at a lower cost compared to conventional approaches, although others recognize that such signals are no panacea.<sup>1</sup>

While the importance of prevention is hard to deny, relatively little attention has been paid to whether preventive care along different margins is worth its cost. The aim of this paper is to investigate this issue in the context of mandatory health checkups in Japan, focusing on risk for diabetes mellitus (DM). We first graphically look at whether health signals about risk of developing DM embodied in health checkup reports affect individuals' medical care utilization, health behaviors, and health outcomes. We then econometrically examine whether the additional care triggered by a health signal is worth the cost.

To identify the cost effectiveness or net value of preventive care, we apply a regression discontinuity (RD) design. We exploit the fact that health checkup results just below and above a threshold, e.g., the level of fasting blood sugar (FBS), may be viewed as random. People with measured values just above the threshold may receive more preventive care – such as further diagnostic tests and diabetes-related physician visits – compared to those with values just below the threshold. This additional care may lead to better health outcomes for the individuals just above the threshold, compared to those just below the threshold. By comparing the cost of care and health outcomes of these people, we can assess the cost effectiveness of providing preventive care around the threshold. Our RD analysis is a “fuzzy” one,

---

<sup>1</sup> See for example discussion in Patel, Asch, and Volpp (2015).

where we use the status of surpassing a threshold as an instrumental variable for medical care utilization. This alleviates the endogeneity concern that an omitted variable such as the person's health status, rather than the preventive care, affects health outcomes.

Using Japanese data provides several key advantages. First, we can construct unique individual-level panel data, which consist of medical claims data, health survey data, and health checkup data. These data sets can be linked by a patient ID. This rich longitudinal data set allows us to examine how health signals embodied in a checkup affect the individual's medical care utilization and health outcomes after the checkup. Second, an annual health checkup is mandatory in Japan; this mitigates concern about sample selection bias. Typically, health-conscious people are more likely to obtain health information by, for example, participating in health checkups or by using wearable devices, and this sample selection is likely to bias estimation results. Health checkups are mandatory in Japan, which alleviates this concern. Third, we have health outcome variables suitable for examining prevention. We apply a Japan-specific risk prediction model, the JJ risk engine (Tanaka et al. 2013), to our data to predict the 5-year risk of mortality and significant DM complications for each individual. These measures allow us to examine directly whether additional preventive care promotes health as measured by medium- and longer-run health outcomes. This is an advantage compared to only examining intermediate health measures such as FBS, HbA1c, and BMI that are more easily available but are also more difficult to interpret (Lipska and Krumholz 2017).

DM is an important case to study because it is a costly and incurable chronic disease of growing incidence and prevalence, and accordingly one of the primary targets for prevention (WHO 2013). DM is often called a "silent killer": individuals at first are asymptomatic and often not aware of the condition, but in the long-run suffer from various serious complications, including problems of the eye, heart, kidney, nerves, and feet. Recent research underscores the economic and human cost of DM: in 2014, approximately 422 million adults have diabetes worldwide, incurring costs estimated to total \$825 billion per year (NCD-RisC 2016). DM can generally be prevented by early intervention to reduce lifestyle risk factors (such as smoking, unhealthy diet, sedentary lifestyle, and obesity). DM and pre-diabetes can be detected by elevated blood sugar levels (i.e., as measured

by fasting blood sugar (FBS) or hemoglobin A1c (HbA1c)), a diagnostic test commonly included in regular health checkups. Indeed, in Japan, policymakers consider this so important that in 1972 they mandated that all employees receive annual screening for elevated blood sugar, as we describe below.

We have three main findings. First, at a relatively low diagnosis threshold (i.e., FBS=110 mg/dl) that corresponds to “borderline type” DM in Japan (sometimes called “pre-diabetes”), we find strong evidence that surpassing the threshold significantly increases medical care utilization as measured by DM-related physician visits and DM-related outpatient expenditures, including on medications. This finding indicates that people do respond to health signals by undertaking follow-up visits with physicians, and thus health signals can potentially promote preventive care. However, the absolute impact of the signal is small: exceeding the threshold increases the probability of visiting a physician for DM treatment by only 5 percentage points (albeit representing a 50% increase, i.e. from 10% to 15%). This small magnitude indicates that health signals do not effectively translate into preventive care for the majority of individuals in the present circumstances. Indeed, we also find no evidence that individuals improve their health-related behaviors (whether on their own or in response to physician advice during preventive care). One of the reasons for this low response rate may be the lack of intervention: currently, after receiving a warning, whether or not to visit a physician is entirely up to the individual; no one monitors response or reminds individuals about the importance of a follow-up visit.

Second, despite the significant increase in medical care utilization at the “borderline threshold”, we find no evidence that the additional care improves health outcomes. This is true both for intermediate health measures (such as FBS, BMI, and SBP) and for predicted risks of mortality and serious complications using the JJ Risk Engine. Thus, we conclude that there is no evidence that DM-related medical care is cost effective around this threshold. The results hold both in the short-run (one year after a checkup) as well as in the medium- to longer-run (three years after a checkup). These results suggest that the threshold may need to be reexamined from the perspective of cost-effectiveness.

Third, at a higher diagnostic threshold (i.e., FBS=126 mg/dl) above which the person is a “diabetic type,” we do not find robust evidence that

crossing the threshold increases medical care utilization or improves health outcomes. At first glance, these results are surprising, because the results indicate that people are less responsive to a signal of higher risk. However, inspections of actual checkup reports revealed that employers rarely flag this threshold in their health reports, and thus most individuals do not receive a health signal when crossing that threshold. Since almost all employers focus on the lower threshold to signal a warning of pre-diabetes, and neglect the threshold signifying the higher risk category of diabetes, we interpret our empirical results as suggesting that policymakers should re-consider the importance of sending a separate signal at each threshold when multiple diagnosis thresholds are of independent clinical significance.

Assessing the cost-effectiveness of health interventions has a venerable, if challenging, history (Garber 2000). Whereas randomized controlled trials (RCTs) are the recognized gold standard for judging cost-effectiveness of various medical interventions, including prevention, RCTs are not always possible and their results may have limited external validity for several policy-relevant questions.<sup>2</sup> Economists have employed multiple techniques to assess cost-effectiveness outside the context of a RCT, including exploiting national experiments (e.g. Finkelstein et al. 2012) and instrumental variables.<sup>3</sup> A notable pioneer in using clinical thresholds to further this line of research is Almond et al. (2010), estimating marginal returns to medical care for at-risk newborns. Focusing on the “very low birth weight” threshold for newborns, Almond et al. (2010) find that those whose birth weights are just below the threshold receive more medical care and experience lower one-year mortality rates, compared to newborns with birth weight just above the threshold. These discontinuities allow them to conclude that medical care for at-risk newborns is cost effective around the threshold.

---

<sup>2</sup> There is a huge literature on the cost effectiveness of pharmaceuticals and devices based on RCTs. However, such studies are often limited to a single medication or device and very few are able to capture the heterogeneity of treatment effects in the general population. The recent widespread application of RCTs in development economics has also highlighted the strengths and limitations of this approach (see for example Duflo, Glennerster, and Kremer 2007 and Deaton 2010).

<sup>3</sup> See Doyle, Graves, and Gruber (2015) for an interesting recent example using random ambulance assignment, and Cawley (2015) for an excellent review highlighting common and creative instruments such as relative distance to a medical care provider offering the treatment, the provider’s historic tendency to administer the treatment, day of week of admission, or randomization of treatment for reasons other than research. Soumerai and Koppel (2016) provide a cautionary view.

While inpatient mortality is a salient outcome for at-risk newborns, it is not a feasible metric for cost-effectiveness of many other health interventions such as routine outpatient care, including preventive care. The difficulty of obtaining an appropriate measure of health outcome is a major obstacle in calculating cost effectiveness. We circumvent this problem by calculating predicted risks of mortality and significant DM complications, using the JJ Risk Engine.

Recently, the value of annual physicals has received renewed attention and our study also relates to this strand of research (e.g., Mehrotra and Prochzka, 2015; Goroll, 2015). For example, systematic reviews have found little evidence that annual physical checkups reduce morbidity or mortality, “though they may be associated with reduced patient worry and increased use of preventive care” (Mehrotra and Prochzka 2015, p. 1485).<sup>4</sup>

Our study is also closely related to Zhao, Konishi and Glewwe (2013), Oster (2015), and Kim, Lee and Lim (2017). Using data from the China Health and Nutrition Survey (CHNS), Zhao, Konishi and Glewwe (2013) apply regression discontinuity analyses to estimate the causal effect of diagnosis with hypertension (in the 3-4 years since the previous wave of CHNS) on food consumption and use of anti-hypertensives. They find a significant increased use of anti-hypertensives, as well as reduced fat intake, particularly among the higher-income individuals told they had high blood pressure. Their results are indicative that health signals from check-ups can lead to behavioral change and preventive care, at least along some margins for some specific population groups.

Evidence in higher-income contexts have generally been less encouraging about modifying long-term behavior with individual health signals. For example, studying consumers in the US, Oster (2015) finds that households with a newly diagnosed diabetic – inferred from household scanner data recording purchase of blood sugar testing strips – exhibit little change in their food consumption behavior over the following months. She suggests that relatively modest “sin taxes” (e.g. on sugary sodas) or

---

<sup>4</sup> Our study is also related to health technology assessment (Institute of Medicine 1985) and systematic review of preventive services (e.g. US Preventive Services Task Force 2009), as well as studies of the impact of health information on individuals’ health-related behavior such as smoking and healthy diets (Chern, Loehman and Yen 1995; Kim and Chern 1999), or response to diagnosis of hypertension (Neutel and Campbell 2008; Zhao, Konishi and Glewwe 2013).

subsidies of healthy fruits and vegetables might ultimately be more effective than individual health signals.

In a recent working paper, Kim, Lee and Lim (2017) study the impact of screening for diabetes, obesity, and hyperlipidemia under the National Health Screening Program in Korea. They find that health checkup information combined with appropriate intervention (such as a follow-up consultation) can prompt positive behavioral change and improve physical measures such as BMI and waist circumference. Although closely related, our study differs from theirs in two important ways. First, we study whether preventive care is cost effective by taking into account the increase in medical expenditures triggered by health checkups, whereas the focus of Kim, Lee and Lim (2017) are the effects of health signals on health behaviors and health outcomes. Second, for the effects on health outcomes, we study not only intermediate physical measures (such as FBS and BMI) but also predicted risks of mortality or serious DM complications by utilizing risk prediction models. Studying the latter is important because it provides more direct evidence on whether health signals promote final health outcomes in addition to physical measures.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss mandatory health checkups in Japan and the key threshold values for DM diagnosis. Section 3 introduces our empirical model and Section 4 describes our data. In Sections 5 and 6, we report graphical and econometric results, respectively. Section 7 reports results from additional analysis, including long-run effects of preventive care. We conclude our study in Section 8.

## 2. Background

### 2.1. Mandatory health checkups in Japan

There are two types of mandatory health checkups in Japan. The first is the traditional mandatory health checkup based on the Industrial Safety and Health Act of 1972. Under this Act, employers must conduct a health checkup annually, and employees are also mandated to take one annually. A checkup must include a series of checkup items as listed in Appendix I, including fasting blood sugar (FBS) and hemoglobin A1c (HbA1c), both of which are commonly used to diagnose DM. The second mandatory health

checkup is the Specific Health Checkup introduced in 2008, which focuses on preventing metabolic syndrome (or lifestyle diseases). Insurers are mandated to conduct this checkup annually for the insured between age 40 and 74. The checkup also monitors FBS and HbA1c levels. Usually, people receive only one of these two checkups each year. The former checkup fulfills the requirement of the latter if all required items are covered.

In both checkups, people receive a report within one or two months after a checkup. Figure 1 shows an example of a report that an individual receives. If any checkup measure exceeds a threshold, the report typically gives a warning (such as “H” for high) for the item and recommends a visit to a physician for further consultation. Although conducting a health checkup with specific required screening items is mandatory, the government does not specify threshold values for each physical measure and employers and insurers that conduct a checkup determine their own thresholds. Also, after receiving a health warning, such as “H”, whether to visit a physician is up to the individual. The person has no obligation to make a visit, and the employer or the insurer is not obligated to monitor or enforce such a follow-up visit.

If a person makes a physician visit after a checkup, fees for the visit are covered by health insurance, and we observe all the treatments made in our claims data. A physician has to record the name of the health condition for which the visit is made, and this information in the claims data allows us to identify DM-related physician visits. When the physician is not yet definitive about the diagnosis, the physician puts a “suspicion” flag on the diagnosis, which we also observe in our data. Because many physician visits triggered by health checkups may not have a confirmed diagnosis at the time of the initial physician visit, we include DM visits with or without a “suspicion” flag in our empirical analysis.

## 2.2 Threshold values for DM diagnosis

There are several thresholds that could trigger preventive care for DM. The Japan Diabetes Society (JDS), which issues treatment guidelines for DM, specifies FBS and HbA1c thresholds for DM diagnosis. Although DM can be diagnosed by alternative measures, such measures are rarely available in

regular checkups and thus we do not consider them in this study.<sup>5</sup> FBS is the traditional measure used for DM diagnosis in many countries. The JDS-specified FBS threshold values for diabetes diagnosis and elevated risk of diabetes have remained unchanged since before our study period. Specifically, an individual with FBS greater than or equal to 126 mg/dl is considered a “diabetic type,” while an individual with FBS greater than or equal to 110 mg/dl but below 126 mg/dl is regarded as “borderline type.” People with borderline type have a high rate of developing DM (Seino et al. 2010). A FBS value below 110 mg/dl is “normal type.”<sup>6</sup> An alternative, newer measure used for DM diagnosis is HbA1c, which JDC adopted in July 2010. HbA1c greater than or equal to 6.5% signifies that an individual is “diabetic type.”<sup>7</sup> Additionally, HbA1c between 4.6% and 6.2% is considered the “standard value.”<sup>8</sup> To summarize, the clinical thresholds for DM diagnosis are FBS=126 mg/dl or HbA1c=6.5%, and for pre-diabetes or “borderline type,” it is FBS=110 mg/dl. Additionally, HbA1c=6.3% is also considered a cutoff value.

### 2.3 Empirical distribution of thresholds used in checkup reports

As discussed in the previous section, there are four thresholds (i.e., FBS=126, 110 or HbA1c=6.5, 6.3) that are most relevant for diagnosing DM. However, employers do not have to adopt these values, because they are not legally bound to any specific signal to employees and can determine their own thresholds for reporting results of health checkups. Unfortunately, our data does not have information on the clinical threshold(s) that each employer adopts. As an alternative, we searched the Internet and investigated what thresholds are typically used in actual checkup reports. We found more than 50 checkup reports posted on the Internet that contain FBS and/or HbA1c thresholds. One of our first main findings is that in all reports, only one threshold (or standard range) is specified for each physical

---

<sup>5</sup> For example, DM can be diagnosed by examining blood sugar two hours after ingesting 75 grams of glucose. Please see Seino et al. (2010) for more details.

<sup>6</sup> Starting in April 2013, JDS defines FBS value greater than or equal to 100 and less than 110 as “high normal.”

<sup>7</sup> In this paper, we express HbA1c values based on the National Glycohemoglobin Standardization Program (NGSP) values.

<sup>8</sup> Please see the following treatment guideline for DM:

[http://www.fa.kyorin.co.jp/jds/uploads/Treatment\\_Guide\\_for\\_Diabetes\\_2014-2015.pdf](http://www.fa.kyorin.co.jp/jds/uploads/Treatment_Guide_for_Diabetes_2014-2015.pdf)

exam measure. That is, no report defines two thresholds for one measure, such as both FBS=110 mg/dl and FBS=126 mg/dl.

Figure 2 shows the “empirical” distribution of the thresholds obtained from our Internet search. It shows that for FBS, clinical thresholds at 110 mg/dl and 100 mg/dl are both common. FBS=110 mg/dl corresponds to “borderline type,” as we discussed in Section 2.2. FBS=100 mg/dl is not a threshold for DM diagnosis, but corresponds to the threshold for metabolic syndrome screening. Understanding the effects of metabolic syndrome screening is also important, but it deserves a thorough investigation beyond the scope of this paper. Thus, we do not study the FBS=100 mg/dl threshold in this paper. FBS=126 mg/dl is the threshold that corresponds to “diabetic type.” However, as Figure 2 shows, almost no health checkup reports adopt this threshold. Thus, it appears that in Japan, individuals are not receiving independent signals from their required check-ups about both pre-diabetes and diagnosable diabetes FBS values.

HbA1c thresholds exhibit a different pattern. In contrast to the FBS values, many more thresholds are used for HbA1c and, more importantly, these values are in close proximity. This makes it difficult to implement an RD approach because there is not a large enough “window” to identify the impact of each threshold. For example, to empirically examine a discontinuity at HbA1c=6.3% that is part of the “standard range,” we need a large number of observations just below and above the threshold. However, because many employers also use the HbA1c=6.0% threshold, only three data points, i.e., HbA1=6.0, 6.1, and 6.2, can be used to represent the data just below the 6.3% threshold. For these reasons, we use FBS values for our RD analyses in this paper.

### 3. Empirical Model

#### 3.1 Regression discontinuity design

We attempt to identify the effect of preventive care on health outcomes using an RD design. We exploit the fact that health checkup results just below and above a threshold may be viewed as random. We believe this is a reasonable assumption because blood sugar levels can go up and down and it is difficult for individuals to precisely control those levels. Moreover, since we focus on non-DM patients in our analysis, these people are typically neither

aware of their blood sugar levels nor have any incentive to manipulate them. People with measured values just above the threshold may receive more preventive care than those with values just below the threshold, such as further diagnostic tests and diabetes-related physician visits. This additional care may lead to better health outcomes (e.g., lower likelihood of developing diabetes and lower medium-to long-term cardiovascular mortality risk) for the patients just above the threshold, compared to those just below the threshold.

A concern for identifying the effects of medical care on health outcomes is endogeneity. An omitted variable, such as the person's health condition, which may be correlated with the amount of medical care, may also affect health outcomes. If so, this will bias the results. We address this issue by implementing a "fuzzy" version of RD, where we use the status of surpassing a diagnosis threshold as an instrumental variable for medical care utilization. Specifically, in the case of the FBS=110 mg/dl threshold, we estimate the following RD model. The first stage regression is as follows:

$$Care_{it} = \alpha_0 + \alpha_1 FBS110_{it} + \alpha_2 Z_{it} + f(FBS_{it} - 110) + A_t + \mu_{it}, \quad (1)$$

where  $Care_{it}$  is medical care utilization of person  $i$  in year  $t$ .  $Care_{it}$  represents one of the following three variables: i)  $DM\_diagnosis$ : a dummy variable that equals one if person  $i$  makes at least one DM-related visit within a year after a checkup in year  $t$  and zero otherwise; ii)  $DM\_visits$ : the total number of DM-related visits for person  $i$  within a year after a checkup in year  $t$ ; and iii)  $DM\_spending$ : DM-related outpatient medical spending (including spending on DM medications) for person  $i$  within a year after a checkup in year  $t$ . We construct these variables by aggregating 12 months of claims data after a checkup. Appendix II defines all the variables used in this study.

$FBS110_{it}$  is a dummy variable that equals one if person  $i$ 's FBS is greater than or equal to 110 mg/dl in year  $t$ , and zero otherwise. We define the other threshold value,  $FBS126_{it}$ , in the same way.  $Z_{it}$  captures person  $i$ 's demographics, including age, age squared, and gender.  $f(FBS_{it} - 110)$  is a function that controls for the FBS level in year  $t$ . We experiment with both linear and quadratic functional forms with respect to  $(FBS_{it} - 110)$  and allow their effects to differ before and after the threshold.  $A_t$  are year fixed effects and  $\alpha$ s are parameters to be estimated.  $\mu_{it}$  is the error term which

we allow to be correlated over time.<sup>9</sup>

The second stage regression is given by the following:

$$y_{it+1} = \beta_0 + \beta_1 Care_{it} + \beta_2 Z_{it} + f(FBS_{it} - 110) + A_t + \epsilon_{it}, \quad (2)$$

where  $y_{it+1}$  captures a health outcome of person  $i$  in year  $t+1$ . As we discuss below,  $y_{it+1}$  represents three types of outcome variables. The remaining variables are the same as in Equation (1).  $\beta_s$  are parameters to be estimated and  $\beta_1$  is the RD coefficient, our main interest in this paper.

The fuzzy RD approach is valid if our excluded instrument,  $FBS_{110_{it}}$ , satisfies the following conditions. First, it is correlated with  $Care_{it}$  and, second, it affects health outcomes,  $y_{it+1}$ , only through  $Care_{it}$ . We will examine these conditions in more detail later, but the second condition requires special attention from the outset. In particular, one may be concerned that the health signal (e.g., “H” for FBS) may not only increase medical care utilization but also independently alter the person’s health-related behaviors such as smoking and exercise habits (without a warning from the physician to stop smoking or engage in more exercise), that could also affect health outcomes. If this is true, the second condition above will be violated. Moreover, we note that whether health signals affect health behaviors is in itself an important policy issue. For these reasons, we test this hypothesis in our data by examining changes in reported health-related behaviors the year after the checkup. As we discuss in Section 7.4, health signals appear to have little effect on health behaviors, which supports the assumption required for identification.

### 3. 2 Dependent variables

We use two types of health outcomes as our second-stage dependent variable. The first captures intermediate health measures, including blood sugar level (FBS and HbA1c), body mass index (BMI), and systolic blood pressure (SBP). These physical measures are taken from the annual checkup in year  $t+1$ . Although these measures are objective and relatively easily measured, these metrics alone cannot tell us whether preventive care significantly reduces the health risks of DM. Therefore, we apply a Japan-specific risk prediction model, the JJ Risk Engine (Tanaka et al. 2013)

---

<sup>9</sup> This is because we sometimes observe the same individual in multiple years.

to predict medium- and longer-term health outcomes. We use risk factors in year  $t+1$  to calculate the individual's predicted risk of mortality and significant DM complications in the following five years, including (i) risk of having a stroke, (ii) risk of developing coronary heart disease (CHD), and (iii) risk of non-cardiovascular (CV) mortality. The JJ Risk Engine uses patient demographics (e.g. gender and age), physical measures (e.g., BMI, HbA1c, blood pressure, cholesterol level), and health behavior (e.g., exercise and smoking) as inputs for calculating risks.<sup>10</sup> Appendix III describes how we implemented the JJ Risk Engine using our data.

It should be noted that the JJ Risk Engine is designed to predict risks for individuals with diagnosed DM without complications, not for predicting risks for individuals without DM. Most individuals in our sample are pre-DM patients whose risk factors fall outside the range used to construct the prediction model. While the engine allows us to calculate the risks even for pre-DM patients, we should be careful about interpreting the results because predicted risks of mortality and significant complications for pre-DM patients are based on extrapolation. For this reason, we checked the robustness of our results using a risk prediction model for cardiovascular outcomes of individuals without diagnosed diabetes (developed by the WHO and International Society of Hypertension). The results do not change with this alternative risk measure.

#### 4. Data

Our data consist of medical claims data, health checkup data, and health survey data. All of these data can be linked by a patient ID. The data come from several employer health insurance groups and are provided by the Japan Medical Data Center (JMDC). As of April 2014, the JMDC claims data base covers 1.6 million members.<sup>11</sup> Our data cover the period between January 2005 and December 2014. The claims data are monthly and we can track the person's medical record as long as the person works for the same employer. Individuals usually have a health checkup once annually; our data includes the year and month of the checkup. A health survey – asking

---

<sup>10</sup> We are extremely grateful for Shiro Tanaka and co-authors for providing the computer code for this project.

<sup>11</sup> JMDC claims data have been used by a number of studies, including Iizuka (2012) and Fukushima et al. (2016).

respondents about their self-assessed health and health-related behaviors – is conducted as part of a checkup and thus is usually the same month as the checkup.

In this study, we are interested in the effects of preventive care on health outcomes and thus we focus on those who are not being treated for DM at the time of a health checkup. We include a checkup in our analysis if it meets the following conditions: i) the patient was not diagnosed with DM during the 6 months before the checkup, ii) we have data for the patient at least 6 months before the checkup, iii) we have data for the patient at least 12 months after the checkup, iv) the patient was 30~64 years old at the checkup, and v) the patient had a checkup only once in a given year.

Table 1 provides summary statistics for the variables used in the analysis. We have more than 1.7 million observations in our data set. Figure 3 looks at the distribution of FBS values. It shows a smooth distribution of measured FBS values, with no apparent discontinuity at either the FBS=110 mg/dl or FBS=126 mg/dl thresholds. More than 287,000 observations are available around the “borderline type” signal, i.e., measured values which fall between  $FBS \geq 100$  and  $FBS \leq 119$ . We have fewer observations around the “diabetic type” signal, but we still observe more than 40,000 observations for the same bandwidth.

An underlying assumption of an RD approach is that covariates do not exhibit a discontinuity at the threshold. To check whether covariates are balanced just before and after the thresholds, we plot the average values of our covariates, i.e., female and age, for each FBS value. As shown in Figure 4, there is no apparent discontinuity at the two thresholds for these variables, indicating that our covariates are reasonably balanced.

In our data, we observe individuals only if they are working at the same company and while the health insurance group provides data to JMDC. To address a potential selection issue, in Figure 5 we plot whether attrition is related with the threshold values, where *Attrition* equals one if the person disappears from our data in the month after a checkup and zero otherwise. As shown in Figure 5, there is no apparent discontinuity at the thresholds, indicating that attrition is unrelated to cutoff values.

## 5. Graphical analysis

We first graphically examine how people respond to health signals. In particular, we look at the two diagnosis thresholds for DM diagnosis, i.e., FBS=110 mg/dl or 126 mg/dl, and how people's (i) health behavior, (ii) medical care utilization, and (iii) health outcomes change at the thresholds.

### 5.1 Effects of “borderline type” signal

Figures 6-9 present graphical analysis for the effects of the FBS $\geq$ 110 signal, above which the person is considered a “borderline type.” Figure 6 clearly shows that all types of medical utilization significantly increase at the threshold. In particular, the probability of visiting a physician for DM increases about 5 percentage points (from 10% to 15%) at the threshold, and the total number of DM-related visits increases by approximately 0.2 per year. Similarly, the use of an oral glucose tolerance test (OGTT), an additional test to diagnose DM, significantly increases at the threshold. DM-related outpatient medical spending also increases by around 2,000 JPY (approximately US \$20) per year.

Whereas we observe a clear jump at the threshold, its absolute impact seems limited. For example, the probability of visiting a physician for DM increases about 5 percentage points at the threshold. Although this represents a 50% increase, it does not seem to be a large absolute magnitude, given that nearly 90% of people could potentially respond to the signal at the threshold (please see Figure 6). One reason for the low response rate may be that only half of those who exceed the threshold receive a warning signal of “borderline type,” as Figure 2 indicates. Moreover, people may discount the clinical importance of the “borderline type” signal even when they receive it.

In contrast to Figure 6, we observe little change in health behavior (Figure 7) and health outcomes (Figures 8 and 9) at the threshold. In Figure 7, the probability of “walking or exercising” appears to somewhat increase at the threshold, but other health behaviors exhibit little change. In Figures 8 and 9, we observe virtually no discontinuities in health outcomes at the threshold, whether measured by intermediate health outcomes (Figure 8) or predicted 5-year risks of mortality and significant DM complications (Figure 9).

### 5.2 Effects of “diabetic type” signal

Figures 10-13 present the results for the  $FBS \geq 126$  mg/dl threshold, above which the person is considered a “diabetic type.” Although this threshold represents a more serious signal of DM, responses are weaker than are those at the  $FBS \geq 110$  mg/dl threshold. First, as shown in Figure 10, medical care utilization somewhat increases, but the impacts are smaller and less clear compared to those found for  $FBS \geq 110$  mg/dl. For example, the probability of having a DM diagnosis increases only around 4 percentage points. This is smaller the impact found at the  $FBS = 110$  threshold. The result for DM-related outpatient medical expenditure now appears insignificant. Second, Figure 11 indicates that the signal has little effect on health behaviors, similar to the case of the  $FBS \geq 110$  mg/dl threshold. Third, as shown in Figures 12 and 13, there is no clear evidence that the  $FBS \geq 126$  mg/dl signal improves health outcomes.<sup>12</sup>

At first glance, it is surprising to find weaker responses at  $FBS \geq 126$  mg/dl than at  $FBS \geq 110$  mg/dl, because individuals presumably would be more concerned about a signal of diabetes than pre-diabetes. However, the results are not surprising in light of what was revealed by Figure 2 – namely, that individuals rarely receive the diabetes signal because very few checkup reports adopt  $FBS = 126$  mg/dl as a threshold. Because people do not appear to receive a signal that they are a “diabetic type,” it is not surprising that they do not respond at the  $FBS \geq 126$  mg/dl threshold. Of course, it is a serious concern if, as we suspect, high-risk people are not alerted that they are actually high risk. Such “false reassurance” could offset any health benefits of (possibly repeated) signals at lower thresholds. One implication of these results is that if multiple threshold values exist for a physical measure, it is important that separate signals be considered at each risk level, calibrated to the strength of the evidence and seriousness of the risk.

## 6. Results from a fuzzy RD regression

### 6. 1. First-stage results

This section reports the results from the first-stage regression of the

---

<sup>12</sup> In Figure 13, “risk for stroke” and “risk for non-CV mortality” appear to decrease somewhat at the threshold. However, if we estimate a regression as specified in Equation (1) using these health outcomes as the dependent variables, the coefficients for  $FBS_{126}$  dummies are not statistically significant in either case even at the 10 percent confidence level (not reported).

fuzzy RD analysis, Equation (1). We estimate the model by using a window width of 5 mg/dl and 10 mg/dl. The 10 mg/dl width is the widest possible one in the case of FBS=110 mg/dl, as there is another cutoff at 100 mg/dl as shown in Figure 2. For the dependent variable  $Care_{it}$ , we employ two variables, i) total number of DM visits and ii) DM-related outpatient medical expenditure. To save space, we only report the coefficients for  $FBS110_{it}$  and  $FBS126_{it}$ .

The results are reported in Table 2. For the FBS $\geq$ 110 mg/dl threshold, the results clearly indicate that DM-related medical care utilization increases at the threshold (please see Panel A). These results are consistent with the graphical results presented in Figure 6. Specifically, the number of DM-related visits increases by 0.2 per year and this estimate is stable regardless of the window width (i.e., 5 mg/dl or 10 mg/dl) and the functional form (i.e., linear or quadratic). All of these estimates are significant at the one percent level. We also find that DM-related outpatient medical spending significantly increases at the threshold. Again, the coefficient for the threshold is significant at the one percent level in all models. The point estimates vary somewhat and indicate that spending increases between 1,200 and 2,000 JPY (approximately US\$ 12-20) per year, depending on the model. These results indicate that the FBS $\geq$ 110 mg/dl threshold dummy variable is highly correlated with the amount of DM-related medical care and satisfies the condition for use as an excluded instrument.

The results for the FBS $\geq$ 126 mg/dl threshold are in stark contrast. As reported in Panel B of Table 2, only one coefficient for the threshold dummy variable is statistically significant. Consistent with the graphical analysis reported in Figure 10, there is no robust evidence that crossing the “diabetic type” threshold increases DM-related medical care, perhaps because individuals are not receiving a health signal at that threshold.

## 6.2. Second-stage results

Table 3 reports the results from the second-stage regression. As we discussed in Section 3, we use two types of health outcomes as our dependent variables, namely (i) intermediate health outcomes and (ii) predicted risks of mortality and significant DM complications. For the endogenous variable in Equation (2), we experiment with two variables, i.e., DM\_visits (Panel A) and

DM\_spending (Panel B), both of which are assumed to be sufficient statistics for DM-related medical care. We only report the results for the “borderline type” threshold because in Section 6.1 we did not find that medical care utilization increases at the “diabetic type” threshold.<sup>13</sup>

The table shows that none of the coefficients are negative and statistically significant. Thus, there is no empirical evidence that the significant increase in DM-related medical care around the threshold improves health outcomes. While we only look at short-run effects in this section, as we report later in Section 7.2., we also do not find evidence of improvements in long-run health outcomes. Therefore the medical care triggered by the  $FBS \geq 110$  mg/dl threshold does not appear to be cost effective.

Please note that we have at least 54,000 observations in our regressions even when we use a narrow window width of 5 mg/dl. Moreover, the first-stage F-statistics are usually bigger than 20 except for some cases where we use the quadratic functional form with a narrow window width of 5 mg/dl (not reported), indicating that the excluded instrument is strongly correlated with DM care, as anticipated by Figure 6. Thus the insignificant second-stage results are not because of the weak instrument problem.

## 7. Additional analysis

### 7.1. Longer-run effects on health outcomes

So far, we have looked at short-run effects of health signals and found no evidence that additional care triggered by health signals improves health outcomes. However, medical care can have cumulative effects and thus we might observe stronger effects in the long-run.

To assess this possibility, we graphically examine the effects on health outcomes three and five years after a checkup, focusing on the “borderline threshold” where we found significant short-run increases in medical care utilization. Figure 14 shows the effects on intermediate health outcomes: we find no apparent discontinuity at the threshold. The results for predicted risks of mortality and significant complications are similar, as shown in

---

<sup>13</sup> Nonetheless, we also estimated the regressions for  $FBS \geq 126$  mg/dl and found that the first-stage F-stats are less than 5 in all cases. This invalidates the instrumental variable approach, as we expected.

Figure 15. The results for five years after a checkup are similar, as reported in Figures A1 and A2 in the Appendix. Note that for these predicted risks of mortality and complications, we are in effect examining 10-year outcomes, since we use risk factors 5 years after the checkup to predicted outcomes for the next 5 years. In other words, for a 2009 checkup, we use blood pressure and other risk factors as measured in 2014 to predict probabilities of suffering a stroke, developing CHD, or non-cardiovascular mortality between 2014 and 2019. Thus, even in the long-run, there is no evidence that additional care for DM (around the margin of “borderline type”) improves health outcomes.

## 7.2 Effects on individuals who did not receive a signal in the previous year

One might expect that some people are not health conscious and they may routinely ignore health warning signals even if they receive them. If we exclude these individuals, the effects of health signals on health outcomes and medical expenditures might be substantially larger. To explore such a possibility, we redo the analysis by focusing on individuals who did not get the “borderline type” signal in the previous year because their FBS values were below the threshold.

In Figure A3, we report the results for medical care utilization. As before, the “borderline type” signal clearly increases medical care utilization. In fact, as expected, the impacts of the signal are slightly larger than what we reported in Figure 6. For example, as shown in Figure A3, the probability of visiting a doctor for DM increases by approximately 6 percentage points, as opposed to the 5-percentage-point increase found in Figure 6.

In contrast, as reported in Figure A4, the effect of the health signal on health outcomes does not appear to be different. Similar to our previous result reported in Figures 8 and 9, there is no clear evidence that the “borderline type” health signal affects health outcomes. Our regression results confirm this finding (not reported). Thus, although individuals who did not receive the warning last year respond more to the signal as expected, additional medical care utilization still does not seem to improve health outcomes. Again, we find no evidence that preventive DM care around the “borderline threshold” of FBS=110 mg/dl is cost effective.

### 7.3 Alternative health outcome measure

As an alternative to the predicted risks of mortality and significant complications using the JJRE, we also experimented with a risk measure calculated by the WHO risk prediction model (please see Appendix IV for how we implemented the risk model.) Unlike the JJRE, the WHO risk measure is based on individuals without diagnosed diabetes and thus complements the JJRE measures. As shown in Figure A5, there is no clear evidence that the “borderline type” and “diabetic type” signals affect the WHO risk measure. We also run the same regression models as before and found that the coefficients for the thresholds are not statistically different from zero (not reported). These results provide additional support for our finding that the health signals have little effect on health outcomes.

### 7.4 Effects of health signals on health behaviors

As discussed in Section 3.1, one concern for our identification approach is that after receiving a health signal, individuals may alter health behavior, which in turn may also affect health outcomes. If so, this makes it difficult to identify the impact of medical care utilization on health outcomes. Although we did not find evidence that medical care utilization affects health outcomes, in this section, we empirically examine the relationship between health signals and health behaviors by estimating Equation (1), using health behavior variables as the dependent variables. Specifically, we create dummy variables for (i) exercise or walk regularly, (ii) smoke, (iii) drink every day, and (iv) eat after dinner, and use them as the dependent variables. We perform this analysis for the “borderline type” threshold where the signal can potentially work as an instrument for medical care utilization.

Table 4 reports the results. To save space, we only report the coefficients for the threshold dummy variables. As shown in the table, only one out of 32 coefficients is statistically significant at the five percent confidence level. These results are in stark contrast to the results for medical care utilization, where we found all coefficients are significant at the one percent confidence level. Thus, we have no evidence that health signals affect health behaviors, and this result provides support for our instrumental variable approach. To the extent that preventive visits to physicians might also involve counseling to reduce lifestyle risk factors such as smoking and sedentary lifestyle, this

non-response along margins of health-related behavior could also be interpreted as further evidence of the lack of cost-effectiveness of preventive care at this margin.

The result that people do not alter health behaviors after receiving a health warning is not surprising in the current context. As previously mentioned, when a physical health measure exceeds a DM diagnosis threshold such as  $FBS \geq 110$  mg/dl, a checkup report usually recommends a visit to a physician, but typically does not recommend or provide programs for lifestyle changes. Moreover, lifestyle changes such as quitting smoking are notoriously challenging. These factors may help to explain why health behaviors change little after surpassing a diagnosis threshold for DM. At the same time, the result is worrisome from the perspective of preventing DM.

## 7.5 Heterogeneous responses to health signals

In Section 5.1, we found that, on average, the “borderline type” signal increases DM-related medical care utilization. This suggests that health signals can potentially help prevent DM by triggering physician visits. Responses to health signals may differ, however, depending on individuals’ characteristics. For example, health-conscious people may respond more to health signals, while those who have a present bias may be less responsive. Information on heterogeneous responses to health signals could potentially help inform the design of effective prevention programs. The aim of this section is to investigate heterogeneous responses to health signals.

To investigate heterogeneous effects, we add two terms to Equation (1). The first is an individual characteristic of our interest, such as whether the person has a “poor eating habit,” and the second is its interaction with the threshold dummy variable, e.g.,  $FBS_{110_{it}}$ . The coefficient for the latter indicates heterogeneity and is the main interest of this section.

We look at three individual characteristics. First, we test the hypothesis that health-conscious individuals are more responsive to health signals, using “poor eating habit” as a proxy for being not health-conscious. A dummy variable, *PoorEat*, equals one if the person reports either eating before sleep or skipping breakfast.<sup>14</sup> Second, we examine whether smokers are less

---

<sup>14</sup> To be more specific, the dummy variable equals one if one “eats dinner within two hours of before sleeping three times or more a week” or “skips breakfast three times or more a week.”

responsive to health signals. Studies in behavioral economics (Laibson 1997, Chaloupka and Warner 2000, Gruber 2001, Volpp et al. 2009, Giné, Karlan, and Zinman 2010) have found that smokers have a present bias and tend to discount the future more than their non-smoking counterparts. These people may respond less to health signals, because they discount the benefits of good health in the future, which may result from today's visit to a physician. Third, similarly, we also examine whether those who drink alcohol respond more to health signals.<sup>15</sup>

Table 5 reports the results. We focus on the “borderline threshold” where DM-related medical care clearly increased. We use diagnosis of DM as the dependent variable and only report the coefficients for the FBS threshold dummy variable, individual characteristics, and their interaction term. Also, to save space, we focus on the quadratic specification.<sup>16</sup> We find that as shown in columns (1) and (2), those who have a “poor eating habit” respond less to the health signal by 1.5-1.6 percentage points relative to all others. This effect is statistically significant at the one percent confidence level. We also find that smokers and drinkers are less responsive to the health signal relative to non-smokers and rare drinkers by 1.3 and 1.5-1.8 percentage points, respectively (please see columns (3)-(6)). In columns (7) and (8), we show the results that include all of the three variables at the same time. The coefficients for the interaction terms become somewhat smaller in this model but are still statistically significant in most cases. A simple summation of the three coefficients in column (8) reveals that the impact of the “borderline type” signal is reduced by half -- from 0.065 to 0.031 -- if the person has all three of these negative health habits.

These results provide suggestive evidence that people respond to health signals differently and those who smoke, drink, and have unhealthy lifestyles (in term of eating habits) respond less to health signals. Note that these people are the primary targets of disease prevention and our results indicate that extra efforts are necessary to induce them to visit a physician. A caveat of this analysis is that our approach only suggests an association but not causation because some confounding factors may be correlated with the characteristics of our interest. A further analysis would be necessary to

---

<sup>15</sup> Drinkers are those who drink alcohol “everyday” or “sometimes” as oppose to “rarely.”

<sup>16</sup> The results are similar for the linear specification and they are available from the authors.

more rigorously identify heterogeneous effects.

## 8. Conclusions

While the importance of preventive care is hard to refute, it is also true that not all preventive care can improve welfare. Using unique individual-level panel data, we investigated whether people respond to health signals and if so, whether medical care triggered by health signals is worth the cost. We did so in the context of mandatory health checkups in Japan, focusing on preventive medical care for DM.

We find that, first, people respond to health signals and increase their probability of visiting a physician. This result confirms that health signals can potentially help prevent chronic diseases by bringing people to physicians' offices. The result also implies that the thresholds adopted in health checkups or wearable devices should reflect the cost-effectiveness of preventive care. If, on the other hand, employers or device makers can freely determine the thresholds on their own, the result may exacerbate rather than mitigate wasteful over-use of some kinds of care while not effectively promoting use of medical resources that are under-used relative to their cost-effectiveness (Baicker, Mullainathan, and Schwartzstein 2015).

Second, our results also indicate that only a small fraction of people respond to health signals along the margins that we study, and thus that the absolute impact of health signals may be limited. We can think of at least two reasons for this low response. First, although employers notify employees about their physical measures from the health checkup, employers may not specifically communicate that the employee belongs to a higher-risk group such as "borderline type" or "diabetic type." That is, the lack of a clear alert may explain the lack of response. Second, even when an individual receives a signal that he/she is a "borderline type," currently, whether to respond by visiting a physician is entirely up to the individual: neither is there a mandate for the individual in question to visit a physician, nor are employers obligated to follow up with employees about the potential benefits of a visit. In this case, if preventive care is indeed cost-effective, then to make prevention work, a more interventionist approach may be necessary. Further analysis is needed to distinguish these explanations.<sup>17</sup>

---

<sup>17</sup> Results from the National Health Screening Program in Korea suggest that clear

Third, most importantly, we do not find evidence that additional medical care triggered by health signals is cost effective. For the “borderline type” threshold, we find substantial increases in DM-related medical care utilization. However, health outcomes did not improve, either for physical measures (risk biomarkers) or for predicted risks of mortality or serious DM complications. Thus, there is no evidence that additional medical care is worth the cost around this threshold, and the current threshold may need to be reexamined. For the higher threshold value that corresponds to “diabetic type,” few people respond to this signal and thus we were unable to assess cost effectiveness at that margin.

There are a large number of diagnosis thresholds that could trigger additional preventive care – primary, secondary, and tertiary prevention – and little is known about their cost effectiveness. While we focus on DM in our analysis, our approach can easily be applied to many other health conditions and clinically-relevant diagnostic criteria. Such analyses could be useful inputs for establishing appropriate diagnosis thresholds and conveying their significance to patients, leading to more efficient use of medical resources.

---

information combined with prompting for a follow-up consultation can lead to a higher response rate, with between 16% and 19% of individuals above the 126 threshold in the baseline check-up subsequently receiving a diagnosis of DM (Kim, Lee and Lim 2017).

## Appendix I: Required checkup items (as of April 2008)

1. Physician's health interview
2. Height, body weight, abdominal girth, visual acuity, auditory acuity, subjective/objective symptom, blood pressure
3. Chest X-ray
4. Anemic analysis (hemoglobin content, erythrocyte count)
5. Liver function (GOT, GPT, gamma-GTP)
6. Blood lipid (TG, HDL-C, LDL-C)
7. Blood Sugar (FBS or HbA1c)
8. Urinalysis (Protein, Sugar)
9. Electrocardiogram

Source: Ministry of Health, Labour and Welfare (2008)  
<http://www.mhlw.go.jp/new-info/kobetu/roudou/gyousei/anzen/dl/080123-3a.pdf> (accessed 2016.12.29).

Appendix II: Definitions of variables (Do we have more variables to discuss?)

Outcome variables	Definition
DM-related outpatient medical spending	<ul style="list-style-type: none"> <li>• Total expenditures within 1 year of the checkup of all outpatient claims which include DM as one of the reported conditions being diagnosed or treated;</li> <li>• “Expenditures” are calculated from points as follows: (Total points of all outpatient claims which include “DM” as one of the conditions + Total points of pharmacy claims which include an anti-diabetic medication as one of drugs)</li> <li>• “DM drugs” are drugs categorized as A10 in ATC code</li> </ul> <p>Note: 1 point = 10 YEN</p>
FBS / HbA1c / BMI / SBP	<ul style="list-style-type: none"> <li>• The values of FBS / HbA1c / BMI/SBP measured at the checkup a year after the index checkup</li> </ul>
Risk of Stroke / CHD / non-CV mortality	<ul style="list-style-type: none"> <li>• The predicted 5-year risk of developing macro- and micro-vascular complications from Type 2 Diabetes, based on the risk factors (e.g. age, blood pressure, HbA1c) measured at the checkup a year after the index checkup.</li> <li>• These predicted risks are calculated from the JJRE equations as developed by Tanaka et al. (Diabetes Care 2013)</li> </ul>
walking or exercising / drinking everyday / smoking / eating after dinner / poor eating	<ul style="list-style-type: none"> <li>• Self-report of health habits (walking or exercising / drinking frequency) as measured in the health survey associated with the checkup a year after the index check-up</li> <li>• “walking or exercising” = 1 if the individual reports exercising enough to work up a sweat for 30 minutes or more per day &amp; 2~7 days per week. (“walking or exercising” = 0 otherwise.)</li> <li>• “drinking everyday” = 1 if he/she drinks every day. (“drinking everyday” = 0 otherwise.)</li> <li>• “smoking” = 1 if he/she has a habit of smoking.</li> </ul>

	<p>(“smoking” = 0 otherwise.)</p> <ul style="list-style-type: none"> <li>• “eating after dinner” = 1 if he/she eats a midnight snack 3 days or more per week. (“eating after dinner” = 0 otherwise.)</li> <li>• “poor eating” = 1 if he/she has dinner within 2 hours of sleep or skip breakfast 3 days or more per week. (“poor eating” = 0 otherwise.)</li> </ul>
DM diagnosis (DMDX)	<ul style="list-style-type: none"> <li>• The rate of a diagnosis of diabetes (or at least rule-out diagnostic testing for diabetes) within 1 year of the index checkup</li> <li>• “diagnosis” is defined by the following 2 conditions: <ol style="list-style-type: none"> <li>1. The individual has a claim that includes DM as one of the conditions; and</li> <li>2. Consultation days &gt; 0.</li> </ol> </li> <li>• “DM” are diseases categorized as E10~E14 in ICD10 code.</li> </ul>
DM visits(DMDAY)	<ul style="list-style-type: none"> <li>• Consultation days for DM within 1 year of the checkup</li> </ul>
OGTT	<ul style="list-style-type: none"> <li>• Rate of conducting an oral glucose tolerance test (OGTT) within 1 year of the checkup</li> </ul>

### Appendix III: Implementing the JJRE

Measurement of medium and long-run health outcomes utilizes 5-year risk of developing CHD, stroke, or non-cardiovascular mortality as predicted by the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine (JJRE) (Tanaka et al. 2013). The JJRE risk prediction model is similar to many other risk prediction models, such as the well-known Framingham cardiovascular disease risk model or the UK Prospective Diabetes Study (UKPDS) risk prediction model often used for estimating medium- and longer-term risks for individuals with diabetes. Such risk models use data from research studies to model how “risk factors” (or predictor variables), such as age, sex, and blood pressure, can predict specific health outcomes in the next 5 or 10 years. Most such models have been calibrated for non-Asian populations, and thus are not appropriate for our sample. The JJRE is specifically designed for predicting risks for a Japanese population. We are grateful to the JJRE authors for sharing their SAS program code with us.

The JJRE incorporates 11 risk factors to predict macro- and microvascular complications among Japanese patients with diabetes (without diabetes complications except mild retinopathy): sex, age, HbA<sub>1c</sub>, years after diagnosis, BMI, systolic blood pressure, non-HDL cholesterol, albumin-to-creatinine ratio, atrial fibrillation, current smoker, and leisure-time physical activity. The model was developed based on data from 1,748 Japanese type 2 diabetic patients pooled from two clinical trials. The JJRE “separately calculates each risk of the first occurrence for five events: fatal and nonfatal CHD, fatal and nonfatal stroke, noncardiovascular mortality, overt nephropathy, and progression of retinopathy” (Tanaka et al. 2013, p.1194).

We have used the JJRE code to calculate the risk of CHD, stroke, and noncardiovascular mortality in our data. We do not estimate risk of overt nephropathy or progression of retinopathy for Iizuka et al (2016), because our sample includes patients not diagnosed with diabetes and therefore the default (diabetic) values from the risk engine are not necessarily appropriate

for determining their 5-year risks of these diabetes complications.<sup>18</sup>

### Defining JJRE input variables

We must use the JJRE default values for those values that we lack for our sample. For each risk factor used in one of the JJRE risk prediction equations, we code the risk factor as follows.

### LTPA (Leisure Time Physical Activity)

LTPA is defined from two questions asked at the health check-up: EXERCISE==1 if the patient answers yes to the question “Have you been exercising at least twice a week (at least 30 minutes per session of light sweating) for over one year?” and 2 otherwise; WALK==1 if answer yes to the question “Do you walk or exercise to a similar degree daily for at least one hour?” and 2 otherwise). The amount of LTPA fits at least the JJRE categories if the self-reported answer to either of these questions is “Yes.”

Then the variable LTPA is coded as follows:

```
gen LTPA = .
```

```
replace LTPA = 1 if EXERCISE == 1 | WALKING == 1
```

```
replace LTPA = 0 if EXERCISE == 2 & WALKING == 2
```

### Duration of DM Diagnosis:

Duration of diagnosis is defined as follows: In the JMDC dataset (4\_diseases\_x., where x signifies a year such as 2014), a variable named “FIRSTDX” (first diagnosis date) associated with each diagnosis code exists. We identified the first diagnosis year associated with diabetes for each patient in the cohort using data from 2005 to 2014. In the great majority of cases, this first diagnosis date remains constant throughout all diabetes-related visits in the data. However, where there are different values for FIRSTDX for a given patient over multiple visits, we took the earliest of the FIRSTDX variable. Duration of diabetes diagnosis in a given year is

---

<sup>18</sup> However, to run the JJRE SAS model requires inputting values for all risk factors and outputting all five predicted risks; therefore, we utilize the JJRE default values for the risk factors that are missing in the JMDC data.

calculated as current year – min(FIRSTDX). Patients without two separate non-suspect (i.e. suspicion flag<sup>19</sup>=0) diagnoses of diabetes in two or more years are assigned a duration of diagnosis of 0 in all years.

### ACR

JMDC does not report ACR values. We therefore assigned ACR a value of 60 (the default value for a diagnosed diabetic in the JJRE) if a patient has two or more confirmed (non-suspect) diagnoses of diabetes in two separate years. Otherwise, a patient is assigned an ACR value of 30.

### NHDL-C (Non-High Density Lipoprotein Cholesterol)

We estimated NHDL-C from the JMDC checkup data on HDL, LDL, and triglycerides, using the Friedewald formula: If  $TG < 400$ ,  $NHDL = LDL + (TG/5)$ ; otherwise if  $TG \geq 400$ , NHDL is set to missing.

### AF (Atrial Fibrillation) and DR (Diabetic Retinopathy)

We used JMDC's disease data files to determine whether an individual had a prior history of AF or DR. For each individual, we identified the earliest year he or she had a non-suspect diagnosis of AF (I48 Atrial fibrillation and flutter) and the earliest year the individual had a non-suspect diagnosis of DR (H36, E103, E113, E123, E133 or E143). Then, for each observation for which a JJRE risk calculation was conducted, we identified whether the earliest year of AF or DR is prior to the current year. If so, we set the dummy variables AF and/or DR to 1, and 0 otherwise.

### BMI (Body Mass Index)

Height and weight are available in the JDMC checkup data and are reported in centimeter and kilograms. We calculated the BMI using the standard

---

<sup>19</sup> The JMDC claims data includes a "suspicion flag" to demarcate claims in which the physician may suspect a given condition but has not definitively diagnosed it, such as a diagnostic rule-out test for a given medical condition like diabetes. We use "non-suspect" to describe claims lacking this suspicion flag (i.e. SUSPECT==0).

formula,  $\text{weight (kg)} / [\text{height (m)}]^2$  after converting height in centimeters to meters.

All other variables (age, female, systolic blood pressure, smoking status, and HbA1c) were taken directly from the patient demographic file or the checkup file.

We confirmed, for a random sample of PIDs with check-up data, perfect congruence between our JJRE predicted risks (from applying the SAS code to our JMDC data) and the JJRE predicted risks output from the web engine of JJRE ([www.biostatistics.jp/prediction/jjre](http://www.biostatistics.jp/prediction/jjre)).

#### Appendix IV. WHO Risk Prediction Model

As an alternative risk prediction measure, we used the risk charts prepared by the World Health Organization (WHO) and International Society of Hypertension (ISH) to estimate 10-year risk of a fatal or non-fatal cardiovascular event. Specifically, we use Figure 25 Western Pacific region A which includes Japan (WPR A). This risk prediction chart, summarizing the WHO/ISH prediction model for 10-year risk of a fatal or non-fatal cardiovascular event by gender, age, systolic blood pressure, total blood cholesterol, smoking status and presence or absence of diabetes mellitus, is available at [http://ish-world.com/downloads/activities/colour\\_charts\\_24\\_Aug\\_07.pdf](http://ish-world.com/downloads/activities/colour_charts_24_Aug_07.pdf).

The WHO/ISH predicted risks summarized in the charts are in categories or ranges (not a continuous variable based on the continuous risk factors of each patient, such as the JJRE model). Therefore we produced the output variable “whorisk” as categorical variables, defined as follows:

0 is a ten year risk of cardiovascular event of <10%;

1 is 10% - < 20%

2 is 20% - < 30%

3 is 30% - < 40%

4 is  $\geq$  40%

The WHO/ISH chart lists age as 70, 60, 50 and 40. We classified patients into the WHO/ISH age categories as follows:

a person falls in the 40 category if he/she is < 50;

a person falls in the 50 category if (s)he is 50 to < 60;

a person falls in the 60 category if (s)he is 60 to < 70;

a person falls in the 70 category if (s)he is 70+.

Systolic blood pressure (SBP) categories in the WHO/ISH chart are 180, 160, 140 and 120 only; we classified an individual’s SBP as follows:

if a person’s SBP is  $\geq$  180, he falls in the 180 category;

if a person’s SBP is 160 to < 180, then he falls in 160;

if a person’s SBP is 140 to < 160, then he falls in 140;

if a person’s SBP is < 140, then he falls in 120.

Total cholesterol is designated as 4, 5, 6, 7, and 8 mmol/l in the WHO/ISH model.

We calculated total cholesterol as  $TC = \text{triglycerides} * 0.2 + HDL + LDL$  in mg/dl; we multiplied this result by 0.02586 to convert mg/dl to mmol/l. Then individuals' total cholesterol is classified into the following categories:

$< 5 \rightarrow 4$

$\geq 5 \text{ to } < 6 \rightarrow 5$

$\geq 6 \text{ to } < 7 \rightarrow 6$

$\geq 7 \text{ to } < 8 \rightarrow 7$

$\geq 8 \rightarrow 8$

The unit of observation is the checkup. Each individual has a predicted risk estimated for each checkup observed in the data.

## References

- Almond, Douglas, Joseph J. Doyle Jr., Amanda E. Kowalski and Heidi Williams (2010) "Estimating marginal returns to medical care: evidence from at-risk newborns," *Quarterly Journal of Economics*, 591-634.
- Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance," *The Quarterly Journal of Economics* 1623 (2015): 1667.
- Cawley, John. "A selective review of the first 20 years of instrumental variables models in health-services research and medicine." *Journal of Medical Economics* 18.9 (2015): 721-734.
- Chaloupka, Frank J., and Kenneth E. Warner. "The economics of smoking." *Handbook of Health Economics* 1 (2000): 1539-1627.
- Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. "Healthcare Exceptionalism? Performance and Allocation in the US Healthcare Sector." *American Economic Review* 2016, 106(8): 2110
- Deaton, Angus. "Instruments, randomization, and learning about development." *Journal of Economic Literature* 48.2 (2010): 424-455.
- Doyle, Joseph, John Graves, and Jonathan Gruber. *Uncovering Waste in US Healthcare*. No. w21050. National Bureau of Economic Research, 2015.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. "Using randomization in development economics research: A toolkit." *Handbook of Development Economics* 4 (2007): 3895-3962.
- Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, and Katherine Baicker. *The Oregon health insurance experiment: evidence from the first year. The Quarterly Journal of Economics* (2012), 1057–1106.
- Fukushima, K, Mizuoka, S, Yamamoto, S, and Iizuka T. 2016. "Patient cost sharing and medical expenditures for the Elderly." *Journal of Health Economics*, 45:115–30.
- Garber, Alan M. "Advances in cost-effectiveness analysis of health interventions." *Handbook of Health Economics* 1 (2000): 181-221.
- Giné, Xavier, Dean Karlan, and Jonathan Zinman. "Put your money where your butt is: a commitment contract for smoking cessation." *American Economic Journal: Applied Economics* 2.4 (2010): 213-235.
- Goroll, Allan H. "Toward Trusting Therapeutic Relationships—In Favor of

- the Annual Physical." *New England Journal of Medicine* 373.16 (2015): 1487-1489.
- Gruber, Jonathan. "Tobacco at the crossroads: the past and future of smoking regulation in the United States." *The Journal of Economic Perspectives* 15.2 (2001): 193-212.
- Iizuka, T. 2012. "Physician Agency and Adoption of Generic Pharmaceuticals." *American Economic Review*. 102(6):2826–58.
- Institute of Medicine. *Assessing Medical Technologies*. Washington, DC: National Academy Press; 1985.
- Kim, Hyuncheol Bryant, Suejin A. Lee, and Wilfredo Lim. 2017. "Knowing Is Not Half the Battle: Impacts of the National Health Screening Program of Korea." Working paper.
- Laibson, David. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics* (1997): 443-477.
- Lipska, Kasia J., and Harlan M. Krumholz. "Is HbA1c the right outcome for studies of diabetes?" *JAMA* Published online January 26, 2017. doi:10.1001/jama.2017.0029.
- Mehrotra, Ateev, and Allan Prochazka. "Improving value in health care—Against the annual physical." *New England Journal of Medicine* 373.16 (2015): 1485-1487.
- NCD Risk Factor Collaboration (NCD-RisC) 2016. "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants" *Lancet* 387: 1513–30.
- Patel, Mitesh S., David A. Asch, and Kevin G. Volpp. "Wearable devices as facilitators, not drivers, of health behavior change." *Jama* 313.5 (2015): 459-460.
- Soumerai, Stephen B., and Ross Koppel. "The Reliability of Instrumental Variables in Health Care Effectiveness Research: Less Is More." *Health Services Research* (2016).
- Tanaka, Shiro, Sachiko Tanaka, Satoshi Iimuro, Hidetoshi Yamashita, Shigehiro Katayama, Yasuo Akanuma, Nobuhiro Yamada, Atsushi Araki, Hideki Ito, MD, PHD8, Hirohito Sone, Yasuo Ohashi, for the Japan Diabetes Complications Study Group and the Japanese Elderly Diabetes Intervention Trial Group. 2013. "Predicting Macro- and Microvascular Complications in Type 2 Diabetes," *Diabetes Care*. 36(5): 1193-1199.

- US Preventive Services Task Force. "Screening for breast cancer: US Preventive Services Task Force recommendation statement." *Annals of Internal Medicine* 151.10 (2009): 716.
- Volpp KG, Troxel AB, Pauly MV, et al. A randomized, controlled trial of financial incentives for smoking cessation. *N Engl J Med* 2009;360:699-709
- World Health Organization 2013. *GLOBAL ACTION PLAN FOR THE PREVENTION AND CONTROL OF NONCOMMUNICABLE DISEASES*. Available from [http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf) (accessed October 11, 2016).
- Yutaka Seino, Kishio Nanjo, Naoko Tajima, Takashi Kadowaki, Atsunori Kashiwagi, Eiichi Araki, Chikako Ito, Nobuya Inagaki, Yasuhiko Iwamoto, Masato Kasuga, Toshiaki Hanafusa, Masakazu Haneda, Kohjiro Ueki "Report of the Committee on the classification and diagnostic criteria of diabetes mellitus." *Diabetol Int* (2010) 1:2–20.

Figure 1. An example of a checkup report

CERTIFICATE OF HEALTH

Name		Date of birth	
Under Medical Treatment	None		
Medical History	None		
Subjective Symptoms	None		
Objective Symptoms	No findings		

This year

Last year

		2012/4/20	2011/4/10	2010/4/15	2009/4/18	Reference(or Normal)
Age		47	46	45	44	
Physical Examination	Height(cm)	171.8	171.9	171.8	171.8	
	Weight(kg)	65.5	66.7	65.2	60.1	
	BMI	22.2	22.6	22.1	20.4	
	Waist Circumference	72.4	72.8	71.3	70.5	
Eye sight	Without glasses(R/L)	-	-	-	-	
	With glasses(R/L)	0.9/1.0	0.8/0.7	0.9/0.9	1.0/1.0	
Hearing	Right 1000Hz	normal	normal	normal	normal	
	Right 4000Hz	normal	normal	normal	normal	
	Left 1000Hz	normal	Impaired	normal	normal	
	Left 4000Hz	normal	Normal	normal	impaired	
	Method	audiometer	audiometer	audiometer	audiometer	
Chest X-ray	Findings	no findings	no findings	no findings	no findings	
	Method	Direct	direct	Direct	direct	
	Film No.	No.314	No.201	No.55	No.308	
Sputum examination		normal	normal	normal	normal	
Electrocardiogram examination		normal	normal	normal	normal	
Liver function	ASL(GOT)	29	33	30	28	≤ 35(U/L)
	ALT(GPT)	27	42	28	26	≤ 35(U/L)
	γ-GTP	44	49	42	38	≤ 55(U/L)
Serum lipid concentration	HDL cholesterol	45	41	43	44	≥ 40(mg/dL)
	LDL cholesterol	110	113	103	99	<120(mg/dL)
	Neutral Fats	107	119	110	100	<150(mg/dL)
Glucose metabolism	FBS	108	<b>H 112</b>	104	<b>H 115</b>	≤ 109(mg/dL)
	HbA1c(NGSP)	5.5	H 5.9	5.2	H 6.0	≤ 5.8(%)
Anemia test	RBC	470	465	480	472	≥ 400, ≤ 539(10 <sup>4</sup> /μL)
	Hemoglobin	15.9	16.2	14.6	16.7	≥ 13, ≤ 16.6(g/dL)
Blood pressure	SBP	102	108	102	98	≤ 130mmHg
	DBP	70	72	70	65	≤ 85mmHg
Uric acid	Glucose	(-)	(-)	(-)	(-)	
	Protein	(-)	(-)	(-)	(-)	

FBS value is high!  
(112,115)

Figure 2. Empirical distribution of clinical threshold values

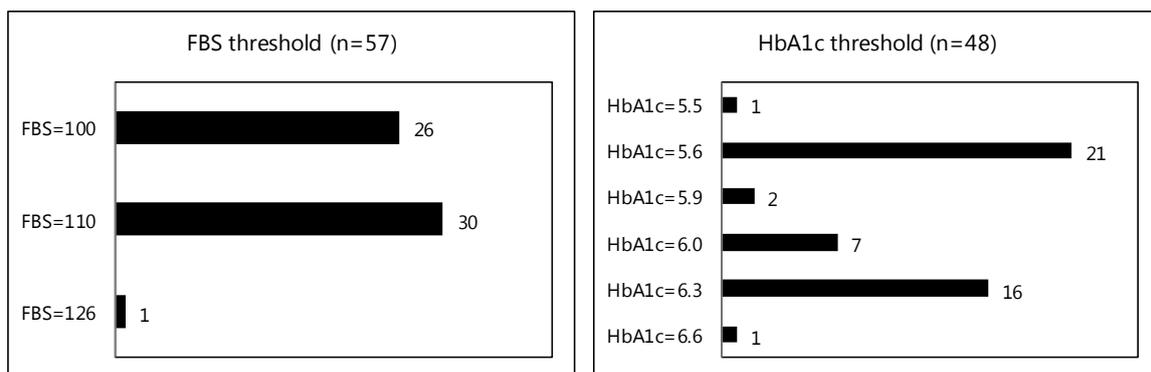


Figure 3. Distribution of FBS values

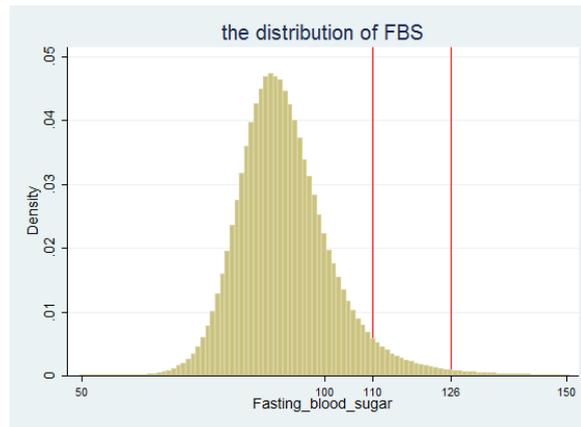


Figure 4. Distribution of Covariates

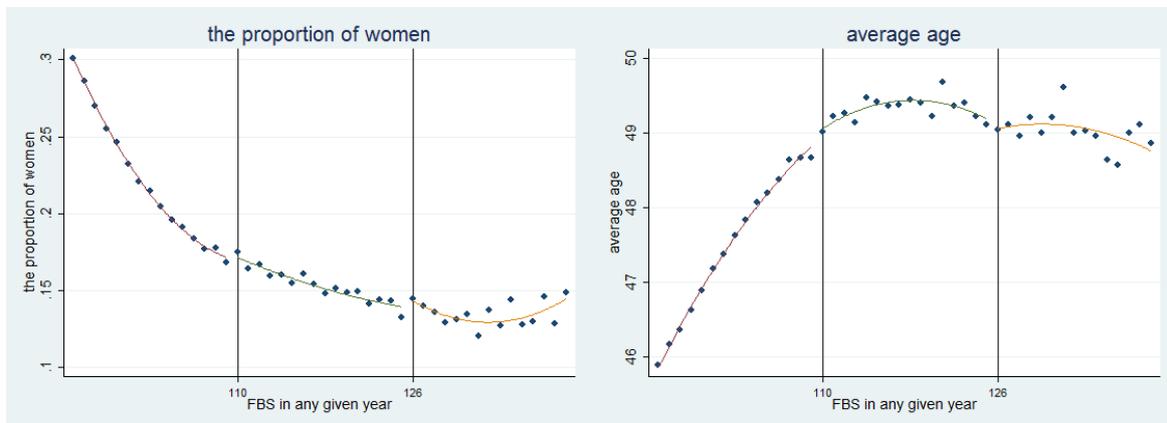


Figure 5. Attrition rate

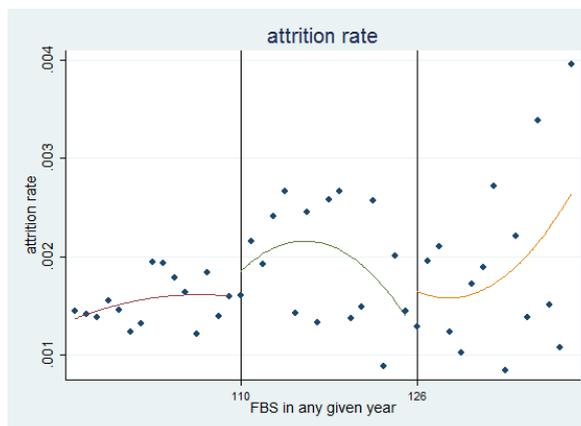


Figure 6. Effects of  $FBS \geq 110$  signal on medical care utilization

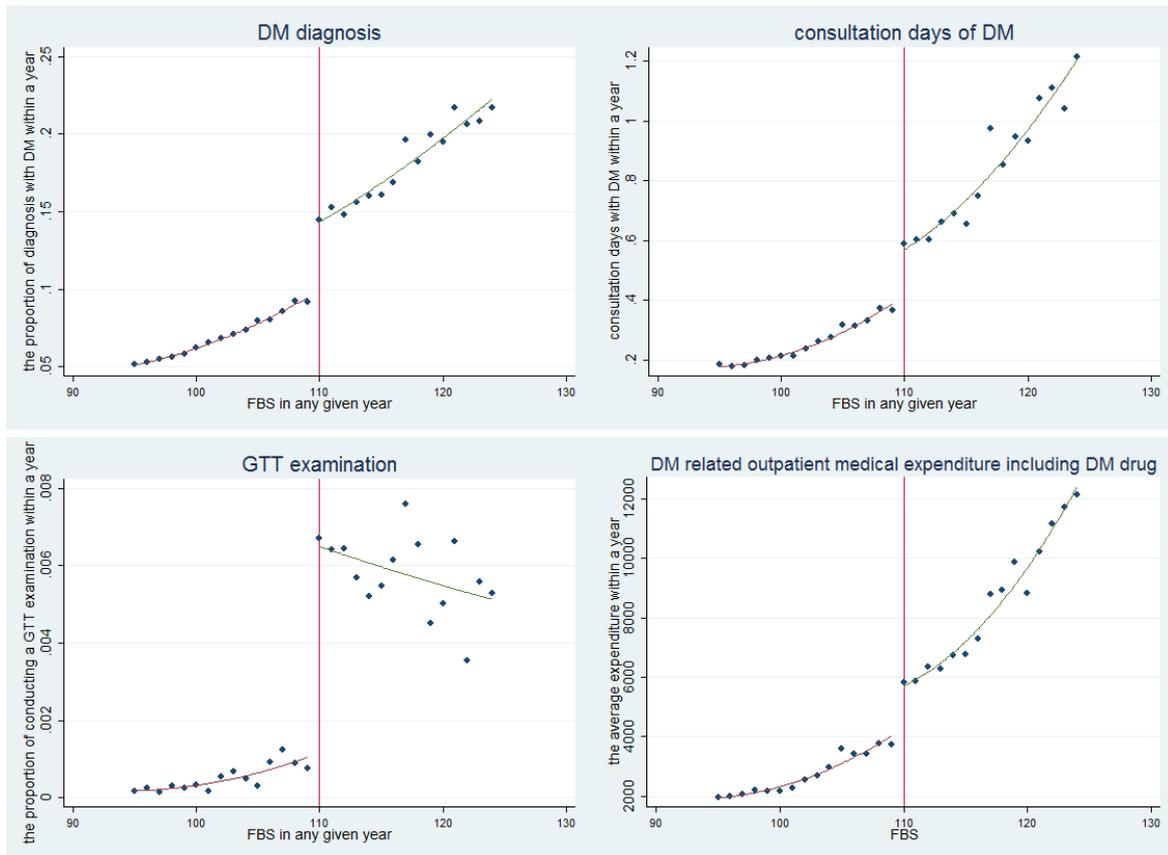


Figure 7. Effects of  $FBS \geq 110$  signal on health behavior

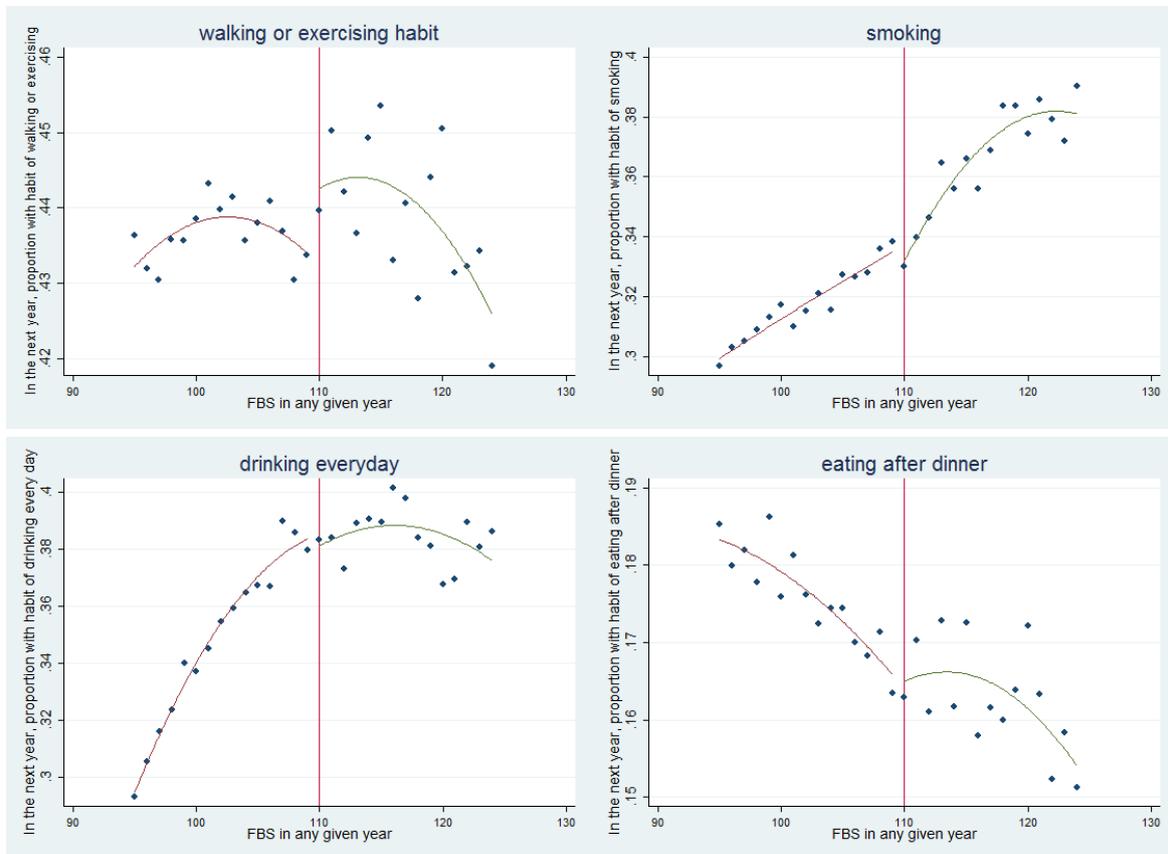


Figure 8. Effects of  $FBS \geq 110$  signal on intermediate health outcomes

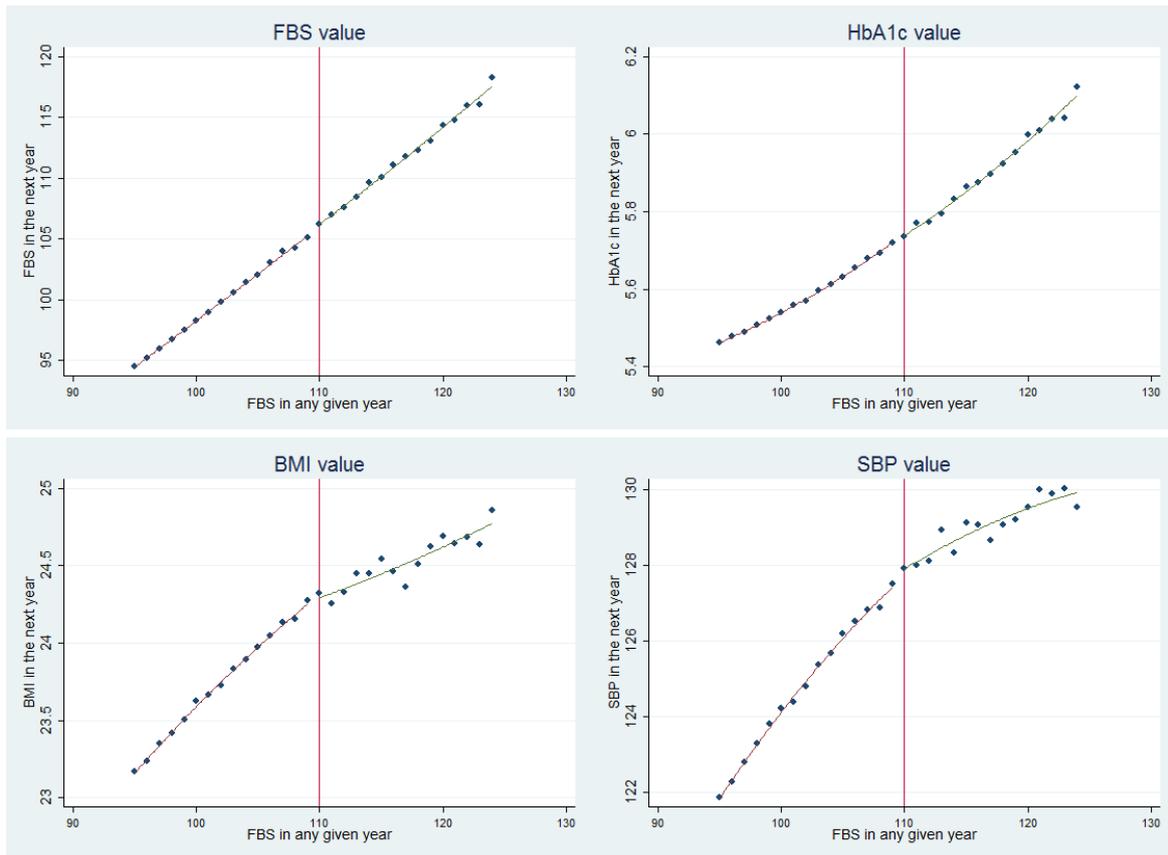


Figure 9. Effects of  $FBS \geq 110$  signal on predicted risks of DM complications

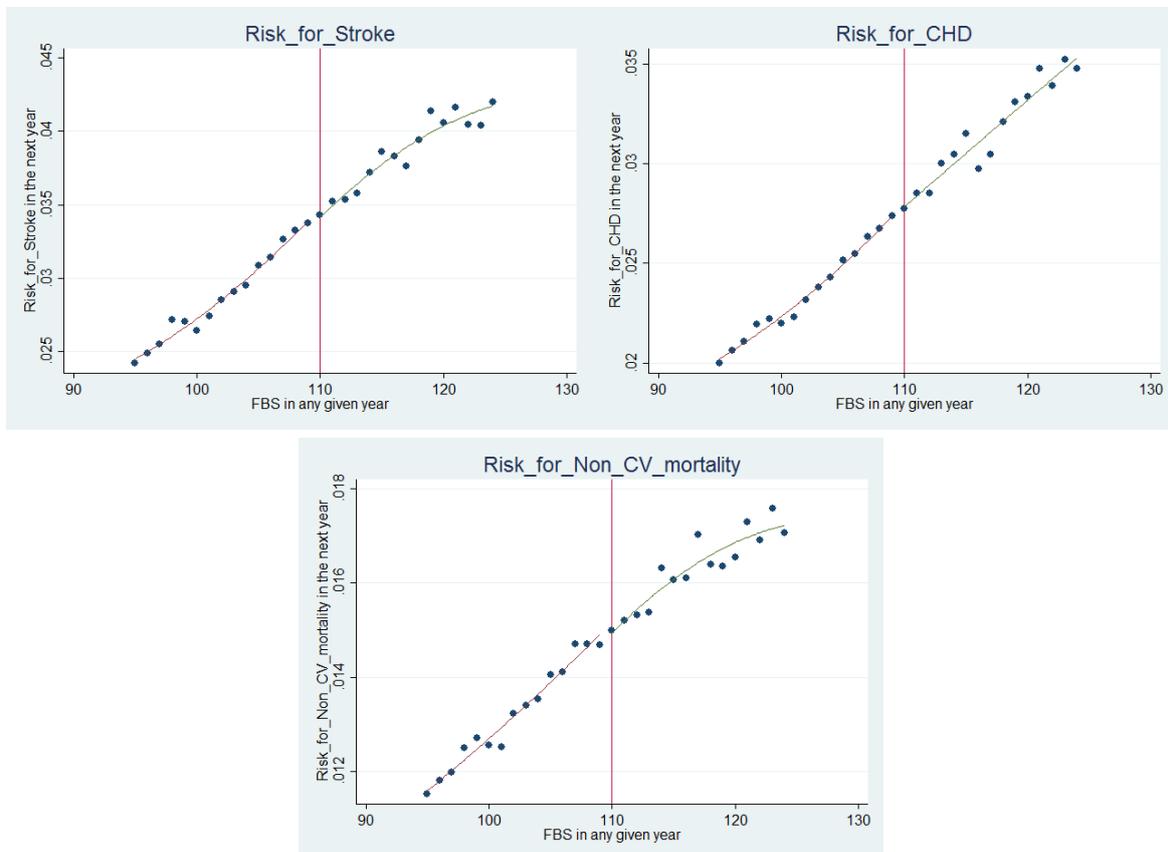


Figure 10. Effects of FBS $\geq$ 126 signal on medical care utilization

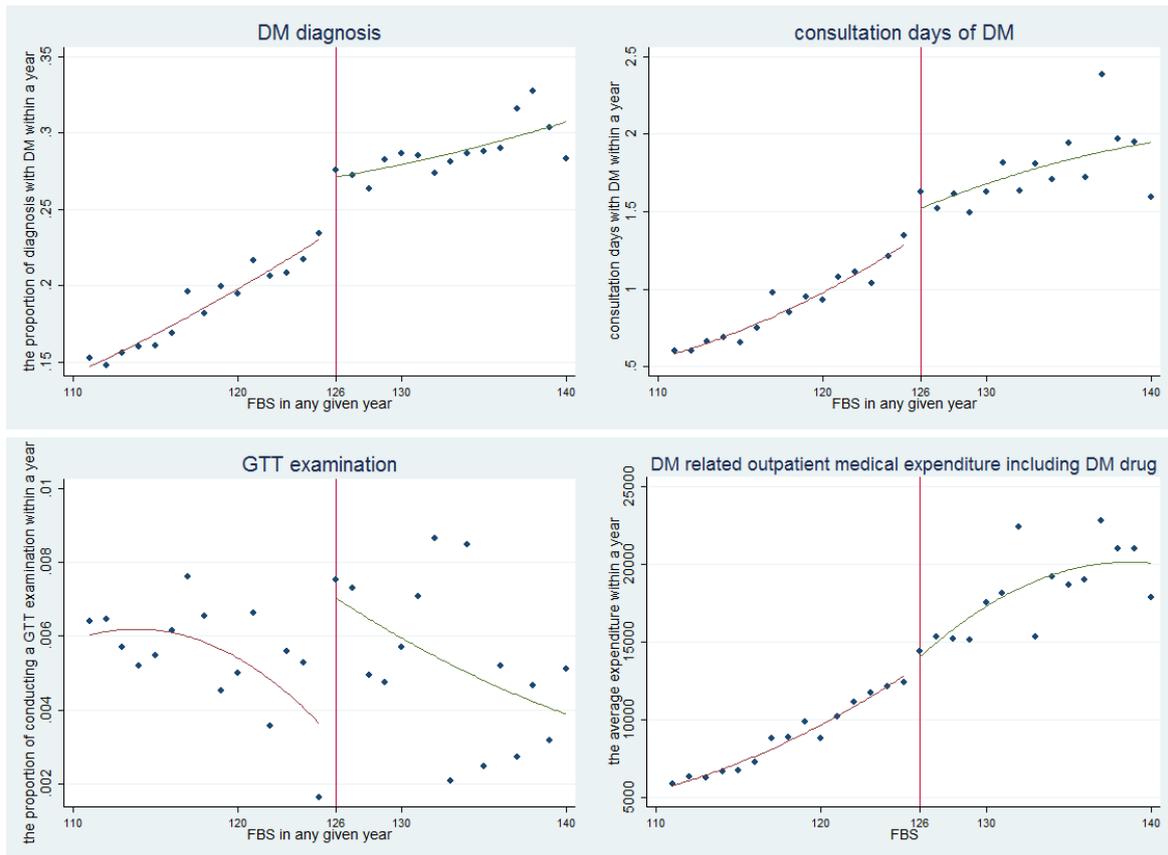


Figure 11. Effects of FBS $\geq$ 126 signal on health behavior

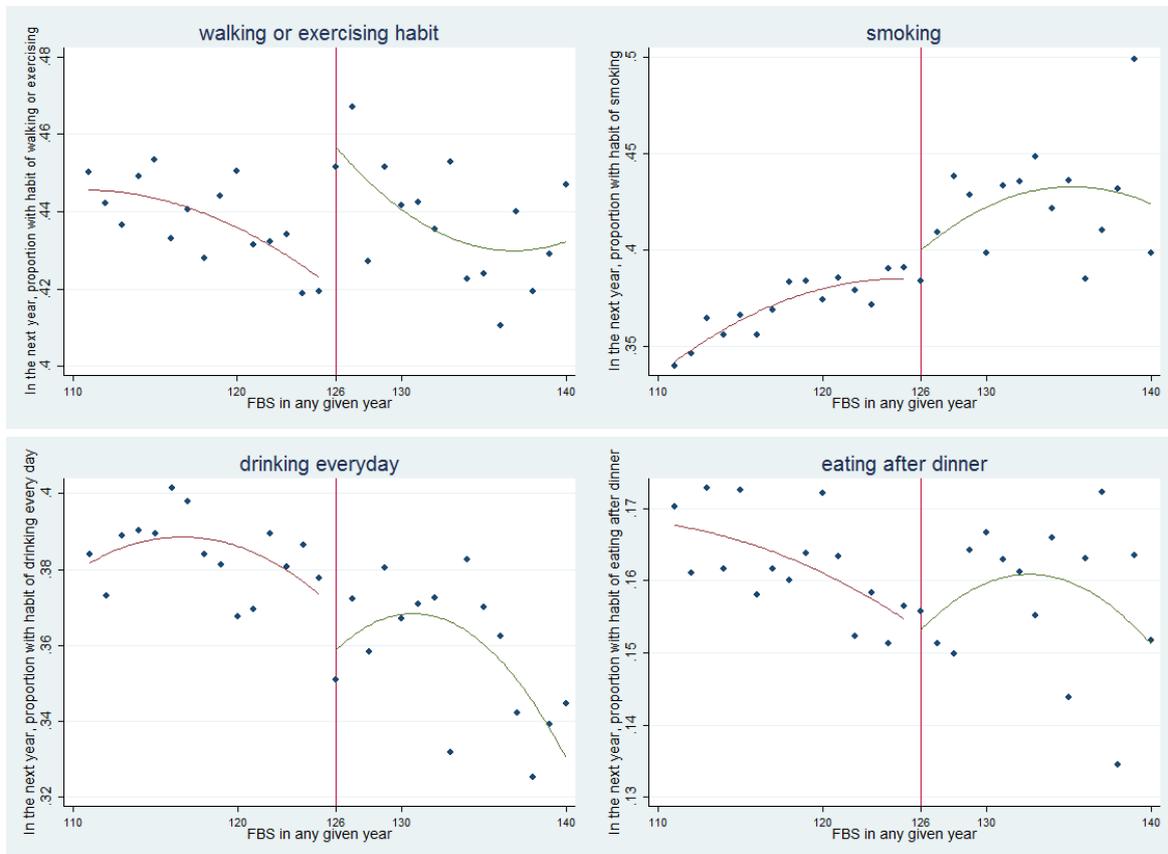


Figure 12. Effects of FBS $\geq$ 126 signal on intermediate health outcomes

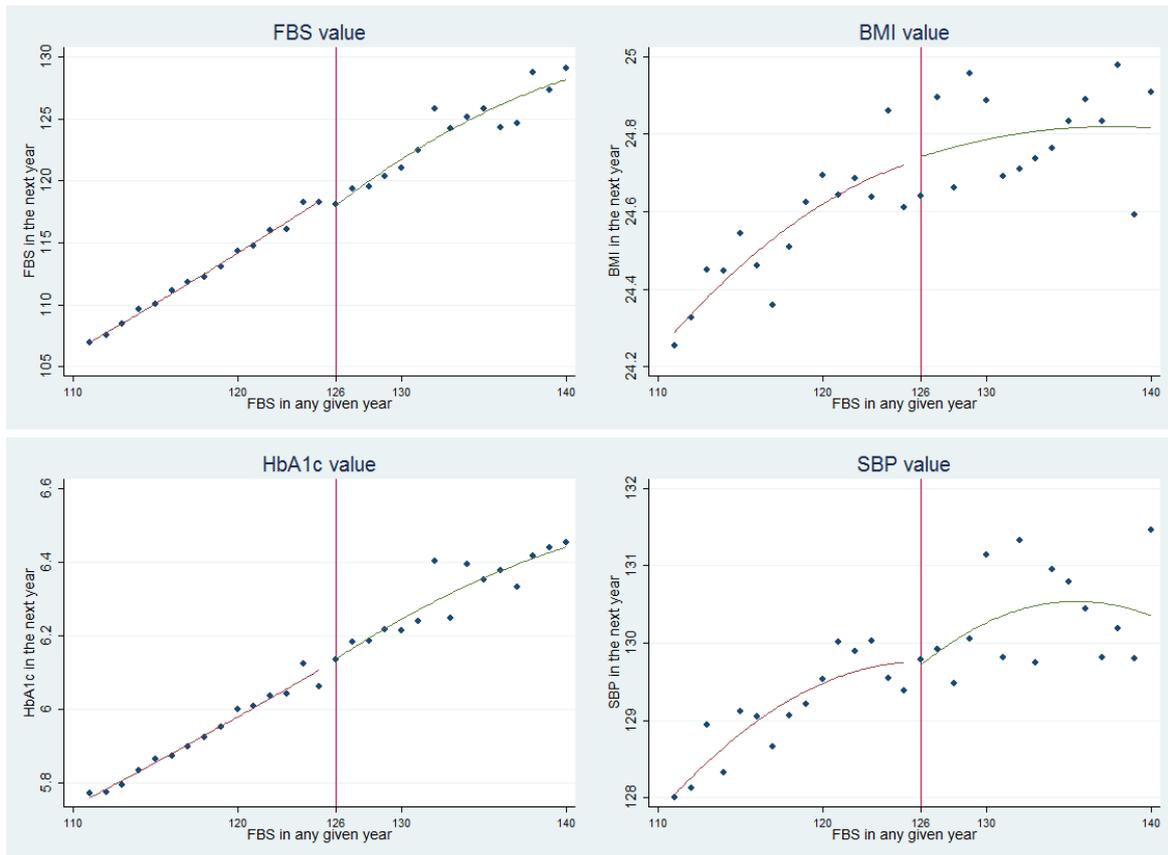


Figure 13. Effects of FBS $\geq$ 126 signal on predicted risks of DM complications

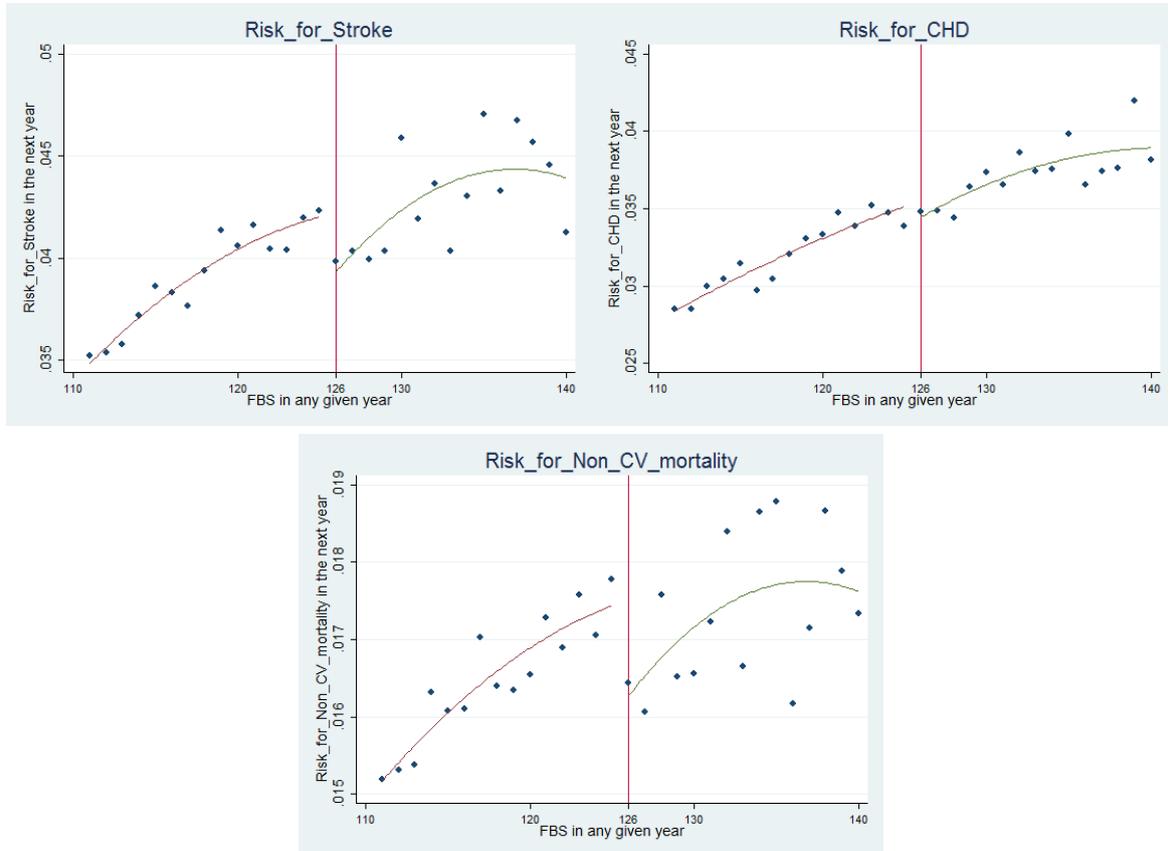


Figure 14. Longer-run effects of additional DM care on intermediate health outcomes (FBS $\geq$ 110 mg/dl, 3 years after a checkup)

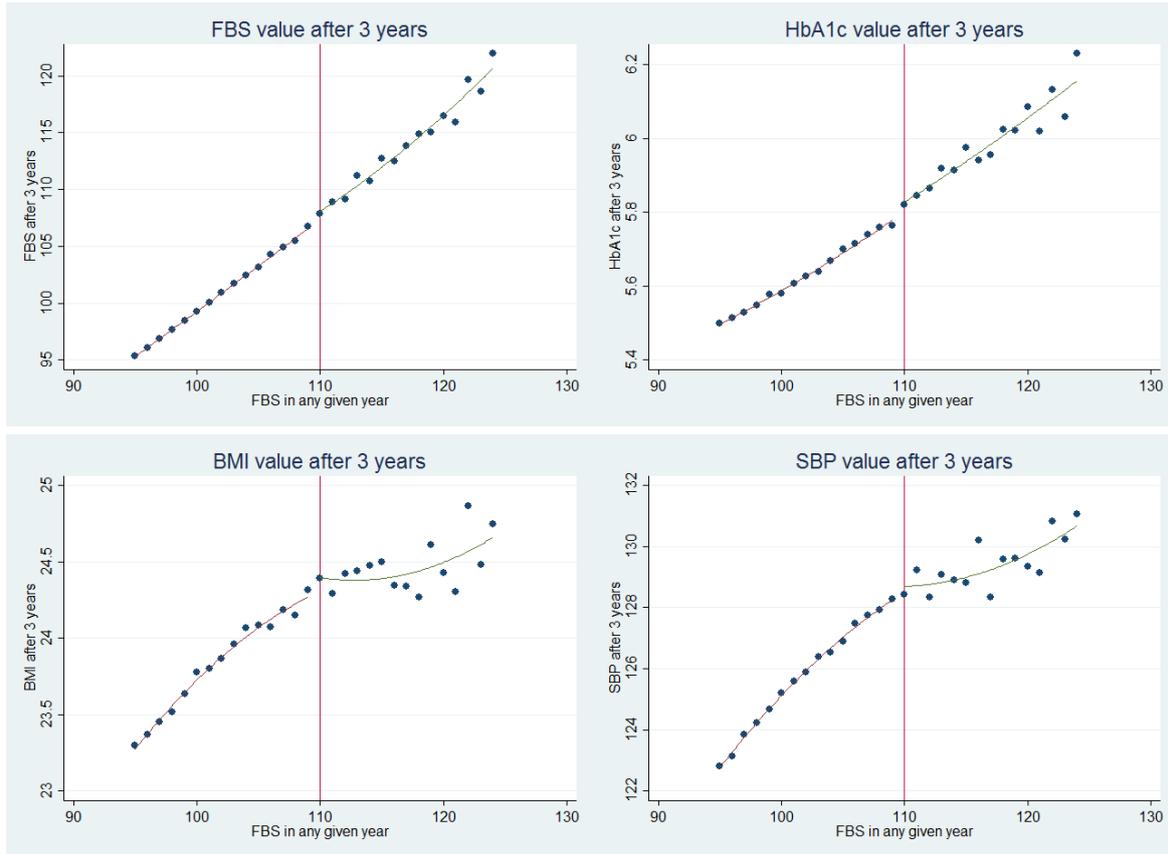


Figure 15. Longer-run effects of additional DM care on predicted risks and complications (FBS $\geq$ 110 mg/dl, 3 years after a checkup)

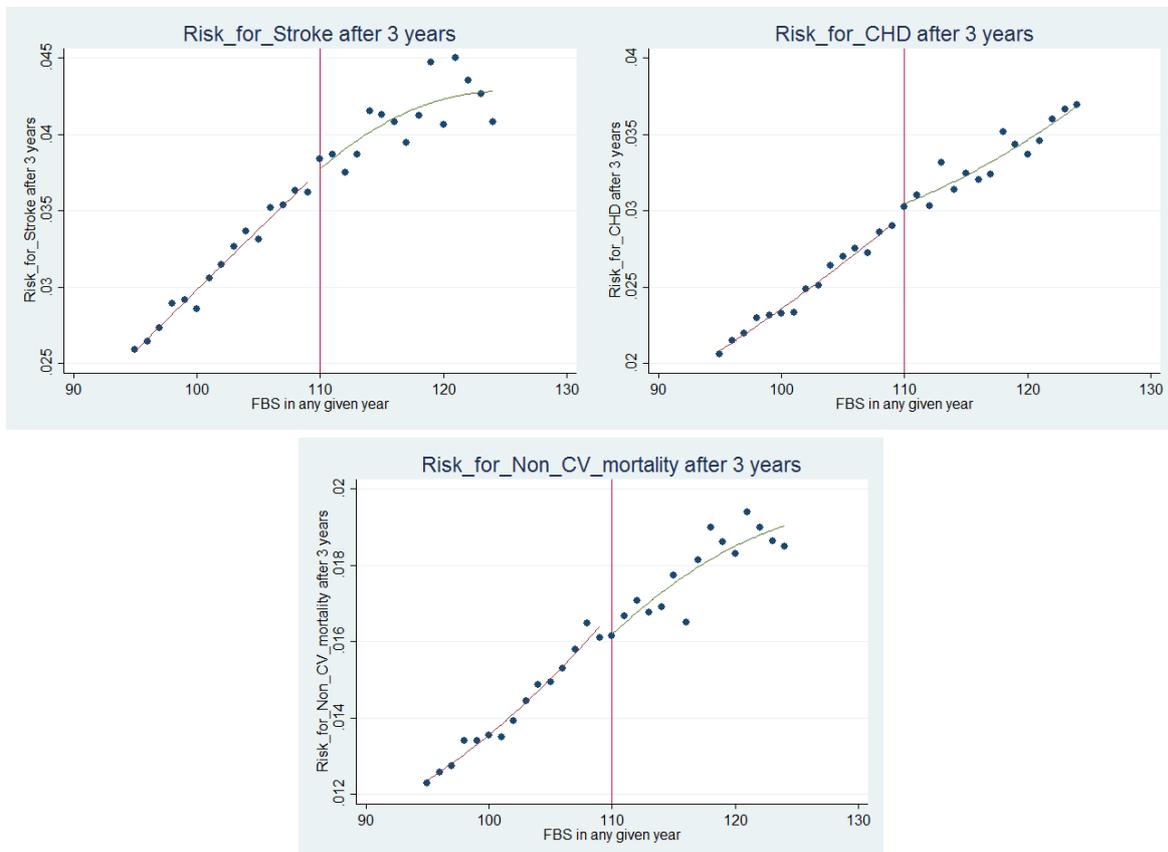


Figure A1. Longer-run effects of additional DM care on intermediate health outcomes (FBS $\geq$ 110 mg/dl, 5 years after a checkup)

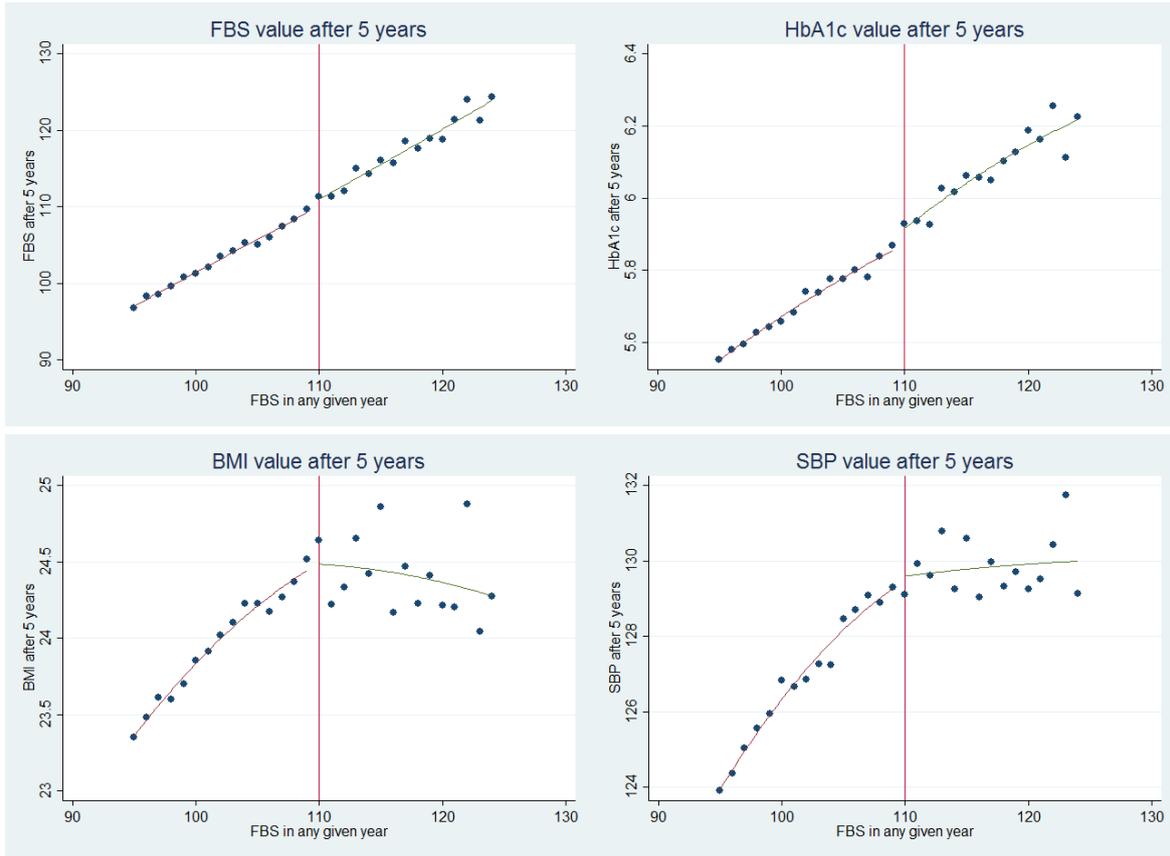


Figure A2. Longer-run effects of additional DM care on predicted risks of DM complications (FBS $\geq$ 110 mg/dl, 5 years after a checkup)

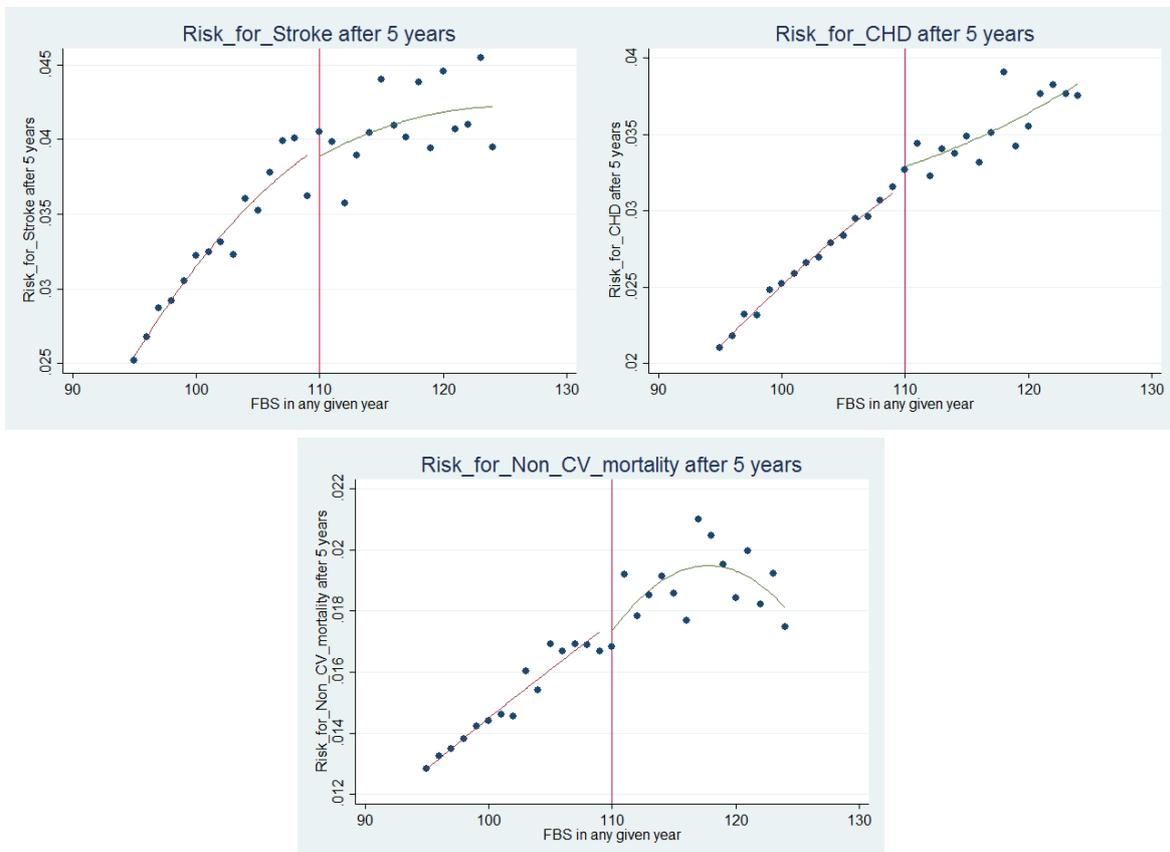


Figure A3. Effects of “borderline type” signal on medical care utilization (for those who did not exceed the threshold in the previous year)

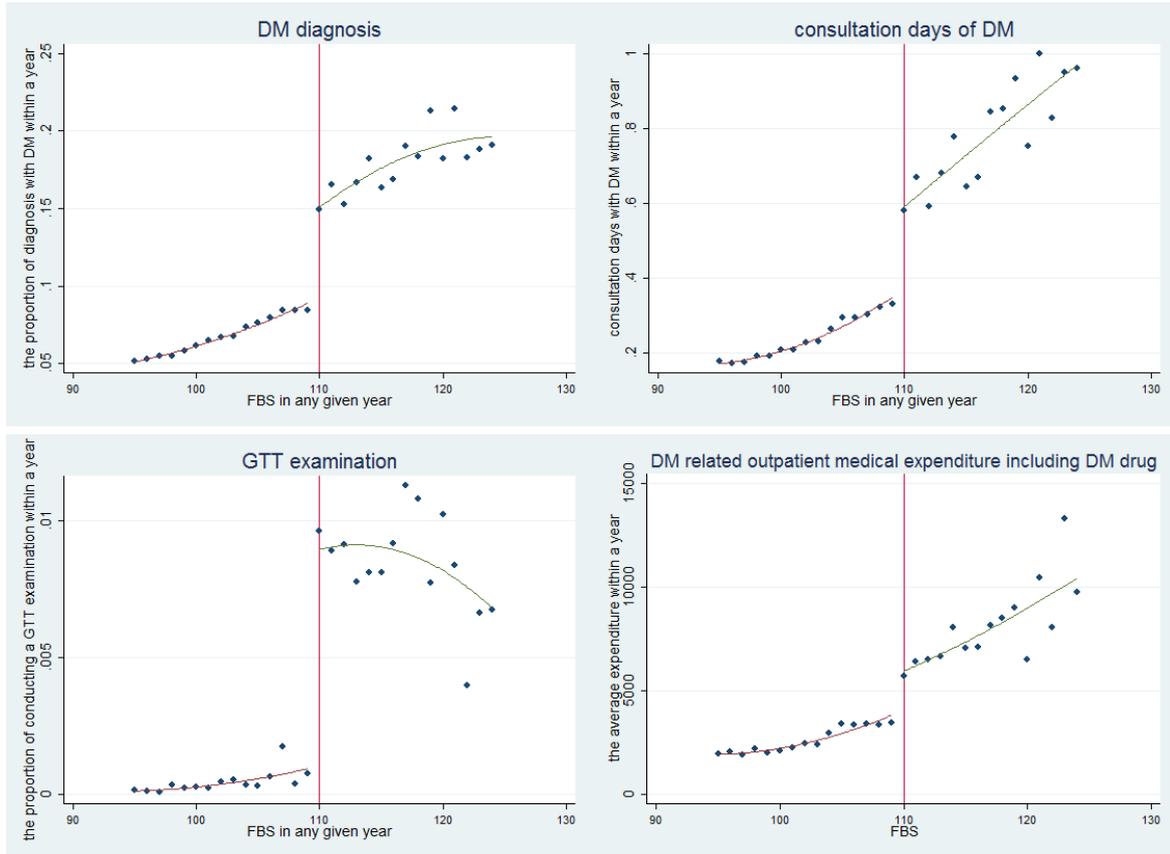


Figure A4. Effects of “borderline type” signal on selected health outcomes (for those who did not exceed the threshold in the previous year)

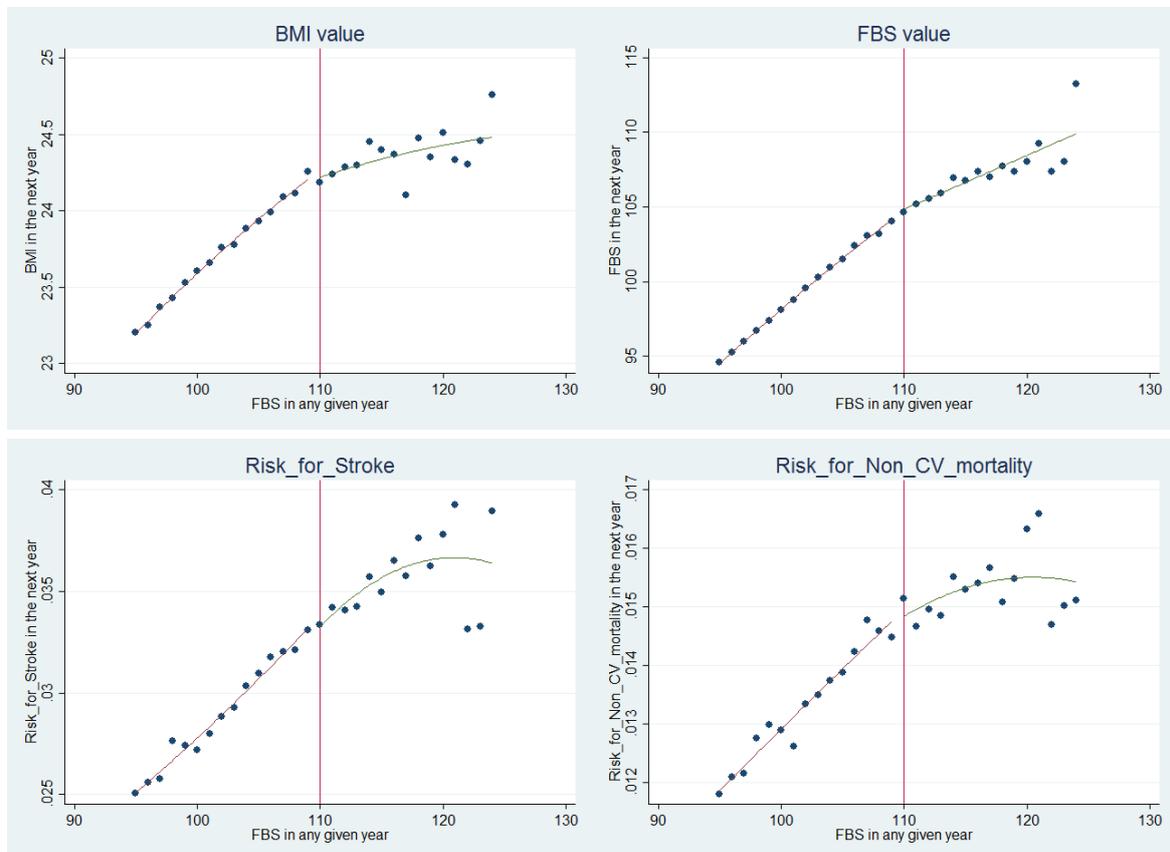


Figure A5. Effects of “borderline type” and “diabetic type” signals on WHO risk

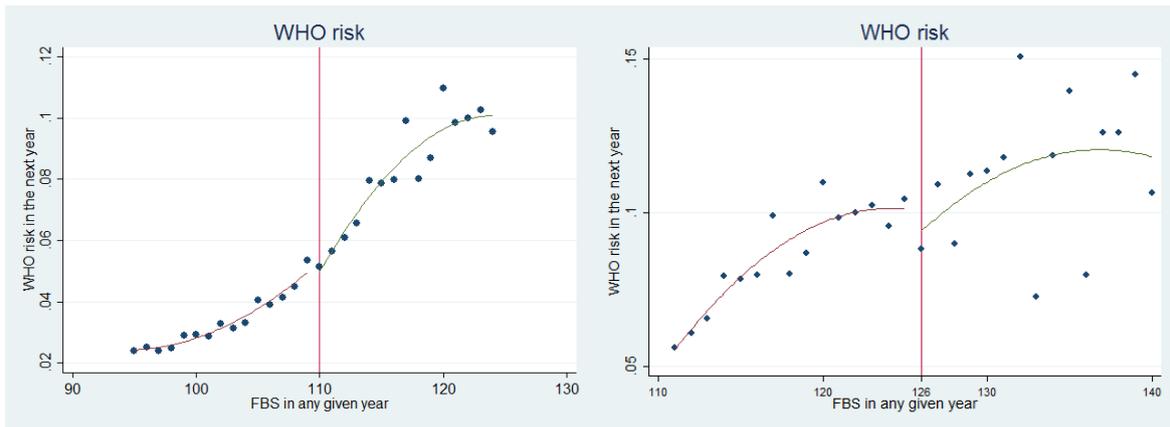


Table 1. Summary statistics

	N	mean	sd
<b>【Explanatory variables】</b>			
age	1695412	45	8.26
FEMALE(0/1)	1695412	0.36	0.48
FBS	1695412	93	14.16
<b>【health behavior】</b>			
SMOKING 1 year later(0/1)	1116472	0.30	0.46
EATAFTDN 1 year later(0/1)	923000	0.18	0.39
ALCHLFRQ 1 year later(0/1)	1045576	0.27	0.44
EXER_WALK 1 year later(0/1)	950440	0.44	0.50
SMOKING 3 years later(0/1)	473226	0.29	0.45
EATAFTDN 3 years later(0/1)	397252	0.19	0.39
ALCHLFRQ 3 years later(0/1)	444334	0.27	0.44
EXER_WALK 3 years later(0/1)	400492	0.44	0.50
SMOKING 5 years later(0/1)	134999	0.27	0.45
EATAFTDN 5 years later(0/1)	112409	0.20	0.40
ALCHLFRQ 5 years later(0/1)	127530	0.27	0.44
EXER_WALK 5 years later(0/1)	110795	0.51	0.50
<b>【medical utilization】</b>			
DMDX(0/1)	1695412	0.06	0.24
DMDAY	1695412	0.24	1.81
GTTEXAM	1695412	0.00	0.02
DMOUTEXP	1695412	2621	25654.89
<b>【health outcomes】</b>			
FBS 1 year later	1116568	93	14.03
HbA1c 1 year later	986728	5.46	0.49
SBP 1 year later	1178393	120	15.47
BMI 1 year later	1178514	22.8	3.43
CHD_R 1 year later	486752	0.02	0.02
STROKE_R 1 year later	486752	0.02	0.03
NonCV_R 1 year later	486752	0.01	0.01
whorisk 1 year later	704229	0.03	0.24
FBS 3 years later	444914	94	14.40
HbA1c 3 years later	393387	5.51	0.52
SBP 3 years later	481284	121	15.39
BMI 3 years later	481385	22.9	3.44
CHD_R 3 years later	217487	0.02	0.02
STROKE_R 3 years later	217487	0.03	0.03
NonCV_R 3 years later	217487	0.01	0.01
whorisk 3 years later	325040	0.03	0.23
FBS 5 years later	120206	96	15.34
HbA1c 5 years later	102694	5.58	0.55
SBP 5 years later	135803	122	15.15
BMI 5 years later	135807	23.0	3.46
CHD_R 5 years later	59579	0.02	0.02
STROKE_R 5 years later	59579	0.03	0.03
NonCV_R 5 years later	59579	0.01	0.01
whorisk 5 years later	100794	0.03	0.23

Table 2. Results from first-stage regressions

	Window=5 mg/dl			Obs.	Window=10 mg/dl			Obs.
	Linear	Quad	Dep.var. mean		Linear	Quad	Dep.var. mean	
Panel A: FBS $\geq$ 110 mg/dl								
Dependent variable								
DM visits	0.191*** (0.027)	0.182*** (0.042)	0.336	116,725	0.154*** (0.020)	0.185*** (0.028)	0.271	287,594
DM outpatient spending	1,968*** (340)	1,601*** (534)	3573	116,725	1,249*** (263)	1,902*** (346)	2857	287,594
Panel B: FBS $\geq$ 126 mg/dl								
Dependent variable								
DM visits	0.230 0.046	(0.150) (0.256)	1.146	18,463	0.232** (0.103)	0.222 (0.157)	0.981	42,292
DM outpatient spending	1,022 (1,251)	2,295 (2,246)	11436	18,463	1,492 (926)	906 (1,415)	9684	42,292

Note: This table shows the results from Equation (1). Only the coefficients for the FBS $\geq$ 110 mg/dl and 126 mg/dl thresholds are reported. Standard errors, corrected for clustering at the person level, are in parentheses. DM outpatient spending includes expenditures on DM drugs. “Dependent variable mean” reports the mean of dependent variable below the threshold.

\*\*\*: 1 % confidence level, \*\*: 5 % confidence level, \*: 10 % confidence level.

Table 3 Results from second-stage regressions

FBS $\geq$ 110 mg/dl	Window=5 mg/dl			Window=10 mg/dl		
	Linear	Quad	Obs.	Linear	Quad	Obs.
Panel A: Endogeneous variable = DM_visits						
Dependent variables						
FBS (next year)	0.798 (0.870)	4.016** (1.878)	77,755	1.271 (0.774)	0.963 (0.930)	191,992
HbA1c (next year)	-0.011 (0.034)	0.000 (0.062)	71,185	0.015 (0.027)	-0.031 (0.038)	170,223
BMI (next year)	-0.173 (0.240)	0.004 (0.414)	82,446	-0.101 (0.200)	-0.201 (0.256)	203,189
SBP (next year)	1.064 (1.001)	-0.351 (1.744)	82,428	0.663 (0.835)	0.326 (1.063)	203,161
Risk for stroke (next year)	-0.001 (0.002)	-0.003 (0.006)	54,157	-0.001 (0.002)	-0.001 (0.003)	130,207
Risk for CHD (next year)	-0.001 (0.001)	-0.003 (0.003)	54,157	0.000 (0.001)	-0.001 (0.001)	130,207
Non-CV mortality (next year)	0.000 (0.001)	-0.000 (0.002)	54,157	-0.000 (0.001)	-0.001 (0.001)	130,207
Panel B: Endogeneous variable = DM_spending						
Dependent variables						
FBS (next year)	0.065 (0.071)	0.390** (0.195)	77,755	0.123 (0.076)	0.077 (0.075)	191,992
HbA1c (next year)	-0.001 (0.003)	0.000 (0.006)	71,185	0.001 (0.003)	-0.003 (0.003)	170,223
BMI (next year)	-0.015 (0.021)	0.000 (0.041)	82,446	-0.010 (0.020)	-0.018 (0.022)	203,189
SBP (next year)	0.095 (0.090)	-0.035 (0.175)	82,428	0.067 (0.085)	0.029 (0.093)	203,161
Risk for stroke (next year)	-0.000 (0.000)	-0.000 (0.000)	54,157	-0.000 (0.000)	-0.000 (0.000)	130,207
Risk for CHD (next year)	-0.000 (0.000)	-0.000 (0.000)	54,157	0.000 (0.000)	-0.000 (0.000)	130,207
Non-CV mortality (next year)	0.000 (0.000)	-0.000 (0.000)	54,157	-0.000 (0.000)	-0.000 (0.000)	130,207

Note: This table shows the results from Equation (2). Only the coefficients for the endogenous explanatory variables are reported. Standard errors, corrected for clustering at the person level, are in parentheses.

\*\*\*: 1 % confidence level, \*\*: 5 % confidence level, \*: 10 % confidence level.

Table 4. Effects of health signals on health behaviors

FBS $\geq$ 110 mg/dl	Window=5 mg/dl			Window=10 mg/dl		
	Linear	Quad	Obs.	Linear	Quad	Obs.
Exercise or walk	0.005 (0.006)	0.007 (0.010)	64,751	0.008** (0.004)	0.004 (0.006)	160,850
Smoking	-0.005 (0.007)	-0.011 (0.012)	77,857	0.000 (0.005)	-0.009 (0.007)	192,176
Drink everyday	-0.010 (0.007)	0.008 (0.013)	72,511	-0.010* (0.005)	-0.008 (0.008)	179,275
Eat after dinner	0.002 (0.006)	0.003 (0.011)	63,476	0.000 (0.004)	0.002 (0.006)	157,121

Note: This table shows the results from Equation (1). Only the coefficients for the FBS $\geq$ 110 mg/dl and 126 mg/dl thresholds are reported. Standard errors, corrected for clustering at the person level, are in parentheses.

\*\*\*: 1 % confidence level, \*\*: 5 % confidence level, \*: 10 % confidence level.

Table 5: Heterogeneous responses to health signals

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Quadratic Window= 5mg/dl	Quadratic Window= 10mg/dl	Quadratic Window= 5mg/dl	Quadratic Window= 10mg/dl	Quadratic Window= 5mg/dl	Quadratic Window= 10mg/dl	Quadratic Window= 5mg/dl	Quadratic Window= 10mg/dl
FBS110	0.056*** (0.008)	0.054*** (0.005)	0.047*** (0.007)	0.047*** (0.004)	0.057*** (0.008)	0.055*** (0.005)	0.068*** (0.008)	0.065*** (0.006)
BADEAT	-0.004 (0.002)	-0.001 (0.001)					-0.003 (0.002)	-0.000 (0.001)
BADEAT*FBS110	-0.016*** (0.005)	-0.015*** (0.004)					-0.014*** (0.005)	-0.012*** (0.004)
SMOKE			-0.005** (0.002)	-0.005*** (0.001)			-0.004 (0.003)	-0.005*** (0.001)
SMOKE*FBS110			-0.013*** (0.005)	-0.013*** (0.004)			-0.009* (0.005)	-0.008** (0.004)
DRINK					-0.012*** (0.003)	-0.013*** (0.001)	-0.013*** (0.003)	-0.014*** (0.001)
DRINK*FBS110					-0.018*** (0.005)	-0.015*** (0.004)	-0.015*** (0.005)	-0.014*** (0.004)
Constant	-0.150*** (0.039)	-0.059** (0.024)	-0.169*** (0.031)	-0.081*** (0.019)	-0.177*** (0.035)	-0.056*** (0.017)	-0.102*** (0.031)	-0.062** (0.024)
Observations	94,011	232,871	104,674	258,492	98,304	243,111	93,462	231,479
R-squared	0.018	0.025	0.017	0.024	0.018	0.025	0.019	0.026

Note: This table shows the results from Equation (1). The dependent variable is “DM diagnosis.” Only the coefficients for the FBS threshold dummy variable, individual characteristics, and their interaction term (our main interest) are reported. Standard errors, corrected for clustering at the person level, are in parentheses.

\*\*\*: 1 % confidence level, \*\*: 5 % confidence level, \*: 10 % confidence level.