Is preventive care worth the cost? Evidence from mandatory checkups in Japan*

Toshiaki Iizuka†
Katsuhiko Nishiyama‡
Brian Chen§
Karen Eggleston**

December 30, 2017

Abstract

Using unique individual-level panel data, we investigate whether preventive care triggered by health checkups is worth the cost. We exploit the fact that the health of individuals just below and above a clinical threshold is similar, whereas treatments differ according to the checkup signals they receive. We find that people respond to health signals by increasing medical care utilization. However, we find no evidence that additional care is cost effective; neither physical measures nor predicted risks of diabetes complications improve in the 3-5 years after the index checkup. For efficient use of medical resources, careful examination of cost effectiveness is essential.

Key words: health, prevention, cost-effectiveness, diabetes
JEL Classification: I10

---

* We are grateful to Ming-Jen Lin, Chie Hanaoka, Hyuncheol Kim and seminar and conference participants at the Korea-Taiwan-Japan Health Economics Associations Joint Conference, the Annual Conference of JHEA, iHEA Congress, EuHEA Conference, and Tokyo University of Science for helpful comments and suggestions. Financial support from the Kikawada Foundation is gratefully acknowledged. All remaining errors are our own.
† Graduate School of Economics, the University of Tokyo. toshi.iizuka@gmail.com (corresponding author)
‡ Graduate School of Economics, the University of Tokyo. katsu.nishiyama@gmail.com
§ University of South Carolina. BCHEN@mailbox.sc.edu
** Stanford University 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 are skeptical.[1]

While the importance of prevention is hard to deny, not all preventive services are cost-effective, so that attention should be given to assessing whether preventive care along different margins is worth its cost.[2] 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 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 examine whether the additional care triggered by a health signal is worth the cost, by comparing the additional medical spending to the value of any resulting improvement in health outcomes.

To identify the cost effectiveness or net value of preventive care, we apply a regression discontinuity (RD) design. We exploit the fact that individuals with health checkup results just below and above a threshold, e.g., a given level of fasting blood sugar (FBS), are similar in their underlying health status.[3] However, people with measured values just above the threshold may

---

[1] See for example discussion in Patel, Asch, and Volpp (2015).

[2] Cohen, Neumann and Weinstein (2008) conducted a systematic review of 599 peer-reviewed articles, 279 of preventive services, and found that while some preventive measures are cost-saving, most reviewed in the literature are not, and many have unfavorable cost-effectiveness ratios.

[3] For the closely related condition of hypertension, a review of the 2017 US clinical guidelines states "the exact cut points for each of these classifications are somewhat arbitrary…[with] strong epidemiologic evidence to support a generally linear association…" (Greenland and Peterson 2017).

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 approach builds on the pioneering work of Almond et al. (2010), who used the "very low birth weight" threshold for newborns to estimate the marginal returns to medical care for at-risk newborns.[4]

Using Japanese data provides several key advantages. First, we can construct unique individual-level panel data, which consist of medical claims, health survey information, and health checkup measurements. These data 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 and more than 95% of employees in large corporations, such as those in our sample, receive a checkup.[5] Typically, in other settings health-conscious people are more likely to obtain signals about their health by participating in health checkups or using wearable devices, and this sample selection is likely to bias estimation results. The mandatory health checkups in Japan alleviate this concern. Third, we have outcome variables suitable for examining the health and survival impacts of 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 study whether additional preventive care promotes health as measured by risk of salient medium- and longer-run health outcomes—such as stroke, heart disease, or death—that are a non-linear function of multiple risk factors and age. Moreover, by assuming a value of statistical life year, we can quantify the monetary value of any

[4] Other studies in the US, China and Korea have also used regression discontinuity to study response to health signals (Zhao et al. 2013, Kim et al. 2017, Oster 2017), without assessing the net value of any additional medical care triggered by the health signals. We discuss these studies and other related literature in section 2.1.

[5] Source: Special Survey on Industrial Safety and Health (MHLW 2012). As we note later, our data come from large corporate health insurers. According to the Industrial Safety and Health Act, any employer that fails to comply with the mandate will face a fine of up to 500 thousand JPY (approximately US$5,000).

improvement in health and survival. This is an advantage compared to only examining intermediate health measures such as FBS, glycated hemoglobin (HbA1c), or body mass index (BMI) that are more easily available but are also more difficult to interpret in isolation (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, and greater risk of premature mortality. 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 is also a major health problem in Japan. It constituted the third largest disease category in 2014, with a national prevalence rate of 7.7% that is increasing as the population ages, and more than 28% of Japan's adult population may have pre-diabetes or DM.[6] Japan's total healthcare expenditure on its 7.2 million people with confirmed diabetes aged 20-79 was the fifth highest in the world in 2017 (International Diabetes Federation 2017). 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 FBS), 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 (also 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,

[6] National Health and Nutrition Examination Survey (MHLW 2015) http://www.mhlw.go.jp/bunya/kenkou/eiyou/dl/h27-houkoku.pdf (P.165). According to biomarker results from the 2016 National Health and Nutrition Survey, over 10 million adults are strongly suspected of having diabetes, and an additional 10 million individuals are above the threshold for "pre-diabetes"(see report at http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/kekkagaiyou_7.pdf (P.8)).

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.

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 blood pressure) and for predicted risks of mortality and serious complications using the JJ Risk Engine. We further confirm these results by estimating a "fuzzy" version of the RD model that examines the causal effects of medical care utilization on health outcomes, recognizing that not everyone who passes the threshold may receive the signal. Thus, we conclude that there is no evidence that DM-related preventive medical care is cost effective (or even 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 to five years after a checkup). These results suggest that either the threshold or the way results are reported to patients 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 classified as a "diabetic type," there is only weak evidence that crossing the threshold increases medical care utilization and improves some health behaviors and health outcomes, and these results are far from robust across measures and specifications. Because we do not observe a clear increase in medical care utilization at this threshold, we are unable to assess the cost-effectiveness of preventive care at the "diabetic type" threshold.

The result that a signal of higher risk has weaker effects on medical care utilization is counter intuitive. Inspections of actual checkup reports revealed, however, 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. Furthermore, there might be an opportunity to significantly enhance the health benefits of Japan's investment in mandatory checkups if insights from behavioral science were used to enhance the clarity

and salience of check-up results, and/or follow-up monitoring was provided for those at the highest predicted risk.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss related literature and the institutional context of our study (e.g. 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 our graphical and econometric results and additional analyses, including the long-run effects of preventive care and various robustness checks. Section 7 discusses how these results compare to those of related studies and Section 8 concludes our paper, pointing out the general applicability of our approach to measuring cost-effectiveness of care around clinical thresholds.

2. Background

2.1. Related literature

Assessing the cost-effectiveness of health interventions has a venerable, if challenging, history (Garber 2000). Randomized controlled trials (RCTs) are the recognized gold standard for evidence, and organizations devoted to assessing cost-effectiveness of preventive care, such as the US Preventive Services Task Force, summarize evidence primarily from such trials. However, RCTs are not always possible and their results may have limited external validity. Economists have employed multiple techniques to assess cost-effectiveness outside the context of a RCT, including exploiting natural experiments (e.g. Finkelstein et al. 2012) and instrumental variables.[7] 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.

[7] 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.

These discontinuities allow them to conclude that medical care for at-risk newborns is cost effective around the threshold.

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 and management of patients with chronic diseases. 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.

Our study is also closely related to Zhao, Konishi and Glewwe (2013), Oster (2017), 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 has generally been less encouraging about modifying long-term behavior with individual health signals. For example, studying consumers in the US, Oster (2017) 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 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. Although closely related, our study differs from theirs in three 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 better health outcomes in the long run, taking into account the interaction of multiple risk factors. Third, our data are based on mandatory checkups in Japan, whereas participation in health screening is voluntary in Korea. Only around 66% of Koreans choose to participate in screening, and this self-selection into screening may bias estimation results. We discuss the differences between their results and ours in the final section.[8]

2.2. Mandatory health checkups in Japan

Japan's universal health coverage system includes insurance programs managed by employers for their employees, as well as insurance programs managed by municipalities. Although the benefit package and payment systems are uniform nationally, the requirements for annual health check-ups differ between the employed and unemployed. Since our data comes from insurance plans managed by corporations, the insurer in this case is the employer.

The most reliably enforced mandatory health check-ups are for employees aged 40-74 working in large firms, such as the insured individuals in our dataset. Under the Industrial Safety and Health Act of 1972, employers must provide, at the firm's expense, annual health check-ups to their employees, with oversight by the Labor Standards Inspection Office and penalties of up to 500,000 JPY (approximately US$5,000) for noncompliance. The employees are also mandated to have these employer-provided check-ups annually, although the individual-level requirement does not have associated legal penalties, just the possibility of employers taking disciplinary actions against those who do not comply. As noted above, more than 95% of employees in large corporations do comply and receive annual checkups.[9]

---

[8] In related work, Kim et al. (2015) find that Korea's national cancer screening program modestly decreased relative inequalities in cancer screening while increasing overall uptake of such screening.

[9] For the non-employed aged 40-74, insurers (usually municipalities) provide the annual health check-up, but there are no penalties for insurers if their insureds do not comply, and

These mandatory checkups must include a series of items, including measurement of height, weight, liver function, blood lipids, blood pressure, and blood sugar. Individuals receive a report of these measurements within one or two months after a checkup. Figure 1 shows an example of such a report. If any measured value exceeds a clinical 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; employers and insurers that conduct the checkups determine their own thresholds, presumably with reference to the clinical guidelines. Moreover, after receiving a health warning, such as "H", whether to visit a physician or not is up to the individual. The person has no obligation to follow up with a medical provider, and neither employers nor other insurers are 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 provided 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.

---

the compliance rate has been low. In April 2017 the Japanese Ministry of Health, Labor and Welfare proposed a plan to introduce financial incentives to increase the annual health check-up adherence rate among non-employed adult Japanese. See discussion in Council on Economic and Fiscal Policy, "Governance reform of Prevention, Health, Medical care and Long-term Care" on April 12th 2017 (in Japanese), available at http://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2017/0412/shiryo_04.pdf, and "Summary of health checkup in Japan" on February 2015 (in Japanese), available at http://www.mhlw.go.jp/file/05-Shingikai-10601000-Daijinkanboukouseikagakuka-Kouseikagakuka/0000104589.pdf; and the Ministry of Health, Labour and Welfare (MHLW) documents in the references.

2.3. Detection and management of DM and pre-diabetes

FBS, or the detection of elevated sugar levels in the blood after an overnight fast, has long been considered the "gold standard" for the diagnosis of diabetes, although the cost-effectiveness of general population screening depends on prevalence and other factors (Hoerger et al. 2004, 2007; Gillies et al. 2008, Sacks 2011). There are two FBS thresholds that could trigger preventive care for DM.[10] The Japan Diabetes Society (JDS), which issues treatment guidelines for DM, specifies that 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."[11] Individuals classified as "borderline type" have a higher rate of developing DM and are sometimes called "pre-diabetic" (Seino et al. 2010). A FBS value below 110 mg/dl is "normal type."[12] These JDS-specified FBS threshold values for DM diagnosis and elevated risk of DM have remained unchanged since before our study period.

Common and simple medical interventions are known to be effective at improving the health measures that we study. Aggressive treatment of hypertension and hypercholesterolemia is recommended for adults with diabetes to prevent cardiovascular complications and other severe sequelae. The standard recommended treatment is a combination of these medications with health behavior improvement such as appropriate physical activity, healthy eating, and quitting smoking.[13] Studies have shown that this multifactorial approach to treatment of diabetes can reduce the risk of diabetes complications (UK Prospective Diabetes Study Group 1998, Miller

[10] Although DM can also be diagnosed by examining blood sugar two hours after ingesting 75 grams of glucose, such measures are rarely available in regular checkups and thus we do not consider them in this study.

[11] We note that the borderline or pre-diabetes threshold is set at FBS=100 mg/dl in the United States and other countries.

[12] Starting in April 2013, JDS defines FBS values greater than or equal to 100 and less than 110 as "high normal." 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." In this paper, we focus on FBS but also look at HbA1c as an intermediate health outcome, as is common in clinical practice; we express HbA1c values based on the National Glycohemoglobin Standardization Program (NGSP) values.

[13] http://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/treatment/txc-20270050

and Dunstan 2004), although low adherence to treatment presents a significant barrier to its effectiveness (Vermeire et al. 2005).[14]

As noted, FBS between 110 and 125 mg/dL is considered pre-diabetes, or "borderline type" in Japan. Recommendations to prevent pre-diabetes from progressing to diabetes include healthy lifestyle choices—healthy eating, physical activity, weight loss, and smoking cessation—and on occasion, medications such as metformin. Evidence suggests that metformin and lifestyle interventions can delay the onset of type 2 diabetes by 3 and 11 years, and reduce the incidence of diabetes by 8% and 20%, respectively (Herman et al 2005).

2.4. Health signals from mandatory check-ups in Japan: Empirical distribution of thresholds used in checkup reports

As discussed in the previous section, there are two FBS thresholds (i.e., FBS=110, 126) that are most relevant for detecting pre-diabetes and 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 online to investigate what thresholds are typically used in actual checkup reports. We found posted on websites more than 50 checkup reports 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 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 web 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

[14] For example, a US study found that in 2009 only about 23% of adults over 40 with DM received all 4 recommended interventions (≥2 HbA1c tests, foot examination, dilated eye examination, and flu shot; Benjamin et al. 2017, p.e301).

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, few 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. Accordingly, we anticipate there may be little response at the FBS=126 mg/dl threshold, despite its clinical importance for detecting diabetes.[15]

We emphasize that we report the "empirical distribution" only for the purpose of interpreting empirical results. Neither do we use any number from the distribution to adjust our estimates, nor do we assume that the above distribution holds for the sample that we examine in this paper.

3. Empirical Model

3.1. Effects of Health Signals on Utilization, Health Behavior, and Health Outcomes

In this section, we examine whether health signals affect (i) medical care utilization, (ii) health behavior, and (iii) health outcomes, using an RD design. The primary purpose of the analysis is to document whether these outcomes indeed discretely change at the threshold. We postpone to Section 3.3 the discussion of the causal effect of preventive care on health outcomes.

In our analysis, we exploit the fact that health checkup thresholds (such as FBS=110 mg/dl and 126mg/dl) are arbitrarily determined and individuals with values just below and above a threshold are similar in their health status.

[15] 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.

However, only those whose values cross the threshold will receive a health signal, and this discontinuity identifies the effect of health signals on the three outcomes mentioned above.

The RD approach addresses the potential endogeneity between health signals and the outcomes of our interest. For example, if one simply regresses the amount of medical utilization on checkup values, the effect of the checkup value is likely to be biased because omitted variables, such as the person's unobserved health status, may be correlated with the checkup value. The RD approach addresses endogeneity because we compare individuals who are similar in all ways except that their checkup values are just above and below an arbitrary threshold.

One common concern when using an RD approach is manipulation of the running variable. In our case, blood sugar levels (our running variable) vary over time and it is difficult for individuals to precisely control those levels. Moreover, physicians and checkup takers do not know blood sugar levels on site and thus it is unlikely that they manipulate the measures in a precise manner. Also, we focus on non-DM patients who were not diagnosed as DM in our analysis; these people are typically neither aware of their blood sugar levels nor have any incentive to manipulate them. Nonetheless, we formally address this concern in Section 4 by performing "manipulation checks" as suggested by McCrary (2008).

In the case of the FBS=110 mg/dl threshold, we estimate the following local polynomial regression using a rectangular kernel:

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

where Y_{it+1} represents one of three types of variables, i.e., (i) medical care utilization, (ii) health behavior, and (iii) health outcome, for person i in year t+1. We discuss each dependent variable in detail in the following section.

$FBS110_{it}$ is a dummy variable that equals one if person i's FBS in year t is greater than or equal to 110 mg/dl, and zero otherwise. We define another threshold value, $FBS126_{it}$, in the same way. $f(FBS_{it} - 110)$ is a function that controls for the FBS level in year t. We experiment with linear, quadratic, and cubic polynomials with respect to $(FBS_{it} - 110)$, allowing their effects to differ before and after the threshold. Z_{it} is a vector of covariates that accounts for person i's demographics, including age, age squared, and gender.

$A_t$are year fixed effects. We estimate the model with and without Z_{it} and A_t. Qualitative results change little with or without covariates, as we report later. αs are parameters to be estimated. μ_{it} is the error term which we allow to be correlated over time.[16]

FBS values are available only in integer values. Because there is no clear way to determine the optimal bandwidth in the case of discrete running variable, we check the robustness of the results by experimenting with different bandwidths between 3 mg/dl and 10 mg/dl. The maximum width is 10 mg/dl in our case because there is another cutoff value of FBS=100mg/dl for "metabolic syndrome screening," as we discussed in Section 2.3.

3.2. Dependent variables

Medical Care Utilization. We use the following four variables to represent medical care utilization: i) Any DM visit: 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. In Japan, physicians must list all diseases of the patient (including suspected ones) and we observe them in our data. We determine that the visit is DM-related if DM is included as one of the diseases; ii) Number of DM visits: the total number of DM-related visits for person i within a year after a checkup in year t; and iii) OGTT examination: a dummy variable that equals one if person i takes an additional test to diagnose DM, called an oral glucose tolerance test (OGTT), within a year after a checkup in year t; iv) DM-related outpatient medical spending: medical spending on outpatient care (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 I defines all the variables used in this study.

Health Behavior. For health behavior, we create dummy variables for (i) walk or exercise regularly, (ii) smoke, (iii) drink every day, and (iv) eat after dinner, and use them as the dependent variables. These variables are taken from health surveys conducted at the time of the health checkups.

---

[16] We cluster standard errors by person because we sometimes observe the same individual in multiple years. Following the suggestion of Kolesar and Rothe (2017), we do NOT cluster standard errors by the running variable because doing so substantially reduces standard errors especially when the number of discrete supports or the bandwidth is small.

Health Outcomes. We use two types of health outcomes as our dependent variable. The first captures physical health measures, including blood sugar level (FBS and HbA1c), 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, since they interact in non-linear ways with each other and the patient's age when determining individual-specific health outcomes such as risks of stroke, heart attack, and mortality. Therefore, we apply a Japan-specific risk prediction model, the JJ Risk Engine (Tanaka et al. 2013) 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.[17] 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.[18] Appendix II describes how we implemented the JJ Risk Engine using our data.

The predicted risks of real clinical endpoints arise from a multistate model that follows the Markov renewal process applied to Japanese patient data from two clinical trials (Tanaka et al. 2013). The JJ risk scores would not be expected to be different if none of the input factors changed, but the predicted risks from more than one risk factor changing are not simple linear functions of the individual factors, and interact with the patient's age.

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

[17] We do not estimate risk of overt nephropathy or progression of retinopathy, because our sample includes patients not diagnosed with diabetes and therefore the default (diabetic) values from the risk engine are not appropriate for determining their 5-year risks of these complications. Moreover, estimating the cost-effectiveness of reduction in these risks involves population-specific measures of quality of life at different gradations of morbidity and is thus less straightforward and generalizable than estimating the value of change in risk of death, or the avoided treatment spending from any reduction in risk of a stroke.

[18] We are extremely grateful for Shiro Tanaka and co-authors for providing the computer code for this project.

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.

3.3. Cost Effectiveness of Preventive Medical care

The main objective of this paper is to quantify the cost effectiveness of preventive medical care triggered by health signals, by looking at the causal effects of medical care utilization on health outcomes.[19] To do so, we implement a "fuzzy" version of RD, where we use the status of surpassing a diagnosis threshold (such as FBS=110 mg/dl) as an instrumental variable for medical care utilization. A "fuzzy" RD is appropriate because not everyone who passes the threshold may receive the signal. People with measured values just above a threshold may receive more preventive care than those with values just below the threshold, and we quantify cost effectiveness by comparing the marginal costs of medical care with the value of any better health outcomes resulting from the additional medical care.

Specifically, in the case of the FBS=110 mg/dl threshold, we estimate the following fuzzy RD model, using local-polynomial regressions with a rectangular kernel. The first stage regression is as follows:

$$Utilization_{it} = \alpha_0 + \alpha_1 FBS110_{it} + f(FBS_{it} - 110) + \alpha_2 Z_{it} + A_t + \mu_{it}\ , \quad (2)$$

where $Utilization_{it}$ is the amount of medical care utilization of person i in year t. We use "Number of DM visits" and "DM-related outpatient medical spending" (including all DM-related outpatient care and medications) for $Utilization_{it}$, both of which are assumed to be sufficient statistics for DM-related medical utilization. The remaining variables are the same as in Equation (1).

[19] Please note that in Section 3.1, we quantified the overall effect of surpassing a threshold on health outcomes regardless of the channel, i.e., whether signals affect health behaviors or medical care. This section focuses on the effect of medical care on health outcomes.

The second stage regression is given by the following:

$$y_{it+1} = \beta_0 + \beta_1 Utilization_{it} + f(FBS_{it} - 110) + \beta_2 Z_{it} + A_t + \epsilon_{it}\,, \qquad (3)$$

where y_{it+1} captures a health outcome of person i in year t+1. We use both physical health measures and predicted risks of DM complications, as discussed in the previous section. The remaining variables are the same as in Equation (1). β_s are parameters to be estimated. β_1 is the RD coefficient, our main interest in this section.

The fuzzy RD approach is valid if our excluded instrument, $FBS110_{it}$, satisfies the following conditions. First, it is correlated with $Utilization_{it}$ and, second, it affects health outcomes, y_{it+1}, only through $Utilization_{it}$. The first condition is the same empirical association that we discussed in Section 3.1. For the second condition, a major concern is that the health signal (such as FBS>=110 mg/dl) 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. Note again that whether health signals affect health behaviors is exactly what we discussed in Section 3.1. To preview our results, we show that in Section 7.4, health signals have little effect on health behaviors, which supports the assumption required for identification.

4. Data

Our data consist of medical claims, health checkup measurements, and health survey responses. 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.[20] 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 and the employer provides data to JMDC. Individuals usually have a health checkup once annually; our data includes the year and month

[20] JMDC claims data have been used by a number of previous studies, including Iizuka (2012) and Fukushima et al. (2016).

of the checkup. A health survey – asking 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 primarily 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 the 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; and iv) the patient was 30～64 years old at the time of the checkup.[21]

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. FBS values are available only as integers and thus we use a bin size of one for the figures throughout the paper. Figure 3 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. We also performed the "manipulation test" proposed by McCrary (2008). As shown in Table 2, the test statistics become significant when the bandwidth is 4 mg/dl and 10 mg/dl for FBS=110 mg/dl and 10 mg/dl for FBS=126 mg/dl. However, the statistics also become significant at other "placebo" FBS values such as FBS=108 mg/dl or 112 mg/dl. Thus, our interpretation is that the test results do not necessarily suggest that the running variable is manipulated at the FBS=110 mg/dl or 126 mg/dl thresholds.

More than 297,000 observations are available around the "borderline type" signal, i.e., measured values which fall between FBS>=100 and FBS<=119. We have fewer observations around the "diabetic type" signal, but we still observe about 44,000 observations for the same bandwidth around FBS=126 mg/dl.

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, female has no apparent discontinuity at the thresholds, but age may exhibit

[21] In rare cases, people take a checkup more than once in a year. In such cases, we include the first checkup in our analysis.

a small jump at the FBS=110 mg/dl threshold. To check these observations, we estimated local-linear regressions using our preferred specification. It turns out that there is a small jump for the former at the threshold as reported in Table A1. To alleviate the concern, we control for the covariates in our regressions and find that, with or without covariates, the results are almost identical. Additionally, Table A1 shows that our dependent variables are reasonably balanced just before and after the thresholds, although the large number of observations results in a small number of coefficients being significant.

In our data, we observe individuals only if they are working in 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 to the threshold values, where *Attrition* equals one if the person disappears from our data within 12 months after a checkup and zero otherwise. As shown in Figure 5, there is no apparent discontinuity at the thresholds, indicating that attrition is not likely to be related to the cutoff values. We also estimated a local-linear regression with our preferred specification that uses a rectangular kernel with a bandwidth of 5 mg/dl with covariates and found that the dummy variable for the threshold is not statistically significant at either the FBS=110 mg/dl or 126 mg/dl thresholds (not reported).[22]

Barreca et al. (2011) suggests that when there is a non-random heaping in the data, we need to be careful about constructing the data set and they propose a donut-hole RD in such a case. In our case, our running variable, FBS, is available only in integers and as shown in Figure 3 there is no heaping in our running variable.

5. Empirical Results

5.1. Effects of Health Signals on Utilization, Health Behavior, and Health Outcomes

In this section, we report how health signals affect (i) medical care utilization, (ii) health behavior, and (iii) health outcomes. We show the results for the "borderline type" signal, followed by those for the "diabetic type" signal.

[22] The results are available from the authors upon request.

5.1.1. Effects of the "borderline type" signal (FBS>=110 mg/dl)

Medical Care Utilization: Figure 6 presents the effects on utilization of the FBS>=110 signal, above which the person is considered a "borderline type." Figure 6 clearly shows that all measures of medical care 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 visits per year. Similarly, the use of an oral glucose tolerance test (OGTT), an additional test to diagnose DM, increases approximately six times, from 0.1% to 0.6%. DM-related outpatient medical spending also increases by around 2,000 JPY (approximately US $20) per year per person. Considering that only 5% of people additionally respond to the signal, medical spending increases approximately by 40,000 JPY (or $400) for those who responded to the signal.

In Panel A of Table 3, we report corresponding local-linear regression results from our preferred specification. To save space, we only report the coefficients for the threshold dummy variables. Consistent with Figure 6, we find that all four measures of utilization significantly increase at the threshold, with magnitudes similar to those shown in Figure 6. For example, the probability of visiting a doctor for DM at least once within a year after a checkup (Any DM visit) increases by about 5 percentage points and the number of DM visits increases by 0.2 per year. All of these estimates are significant at the one percent level.

Figure A1 in the Appendix checks the robustness of the results by experimenting with different bandwidths from 3 mg/dl to 10 mg/dl and by using local linear, quadratic, and cubic polynomials. The vertical lines in the figure indicate the 95% confidence interval for each estimate. As shown, the results are robust regardless of bandwidths and polynomials. Figure A2 reports the results without covariates. The results are virtually identical to those with covariates, providing further confidence on our empirical findings.

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 (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.

Health Behavior. In Figure 7, we report results for health behavior. In stark contrast to Figure 6, we observe little effect of the health signal on health behavior. Although the figure leaves open the possibility that "walk or exercise" and "smoke" may have improved slightly at the threshold, local linear regressions reported in Panel B of Table 3 indicate that in fact none of the coefficients are 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 conclude that there is no evidence that health signals affect health behaviors.

The result that people do not alter health behaviors after receiving a health warning may not be surprising in the current context. As previously mentioned, when a physical health measure exceeds a DM diagnosis threshold such as FBS>=110 mg/dl, a checkup report usually recommends a visit to a physician and/or re-test of the measure that is outside the normal range, but typically does not recommend or provide information regarding lifestyle changes. The check-up report (see Figure 1) is rather cryptic and does not explain the significance of any given measure or the combined level of risk that an individual may have, given their age and other risk factors. 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 "borderline type" DM.

At the same time, however, the result is worrisome from the perspective of preventing DM. 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 evidence of the lack of effectiveness of preventive care at this margin.

In Figure A3, we check the robustness of the results by altering bandwidths and polynomials. Figure A4 further checks the results without covariates. The results are robust to these checks. Therefore, there is little

evidence that health signals affect health behavior at this threshold, unlike the case of medical care utilization.

Health Outcomes. Figures 8 and 9 show the impact of the FBS>=110 health signal on health outcomes. In contrast to the impact on medical care utilization, we observe virtually no discontinuities in health outcomes at the threshold, whether measured by intermediate health outcomes (Figure 8) or by predicted 5-year risks of mortality and significant DM complications (Figure 9). Estimation results reported in Panel C of Tables 3 confirm these observations. In particular, none of the estimated coefficients are significantly negative. As show in Figures A5—A8, these results stay the same even when we use different bandwidths, polynomials, and with and without covariates. All of these results indicate that whereas the signal clearly increases physician office visits and medical care utilization, there is no evidence that the additional utilization leads to better health or survival.

We note that although the estimated coefficients are not significant, they are relatively precisely estimated as indicated by the standard errors reported in Table 3. For example, the point estimate of the RD dummy for risk of stroke is -0.028 with a standard error of 0.059, implying the confidence interval ranges from -0.146 to 0.090. These potential effects are much smaller than the average risk of stroke at FBS=110 mg/dl, which is approximately 3.5 (see Figure 9). Similarly, for risk of CHD and Non CV mortality, corresponding confidence intervals range from -0.088 to 0.040 and -0.031 to 0.41, respectively. They are again much smaller than the average risks for these conditions at FBS=110 mg/dl, which are approximately 2.8 and 1.5, respectively. Thus, the impact of the "borderline type" signal on health outcomes is likely to be small even if it is non-zero.

5.1.2. Effects of the "diabetic type" signal (FBS>=126 mg/dl)

Medical Care Utilization: Figures 10 shows the effects on medical care utilization of the FBS=126 mg/dl signal, above which the person is considered a "diabetic type." As shown in Figure 10, medical care utilization also appears to increase at this threshold, but the impacts are less clear than those found for the FBS>=110 mg/dl threshold shown in Figure 6. For example, the probability of having at least one DM visit (Any DM visit) appears to increase at the threshold but only around 4 percentage points, slightly smaller than

the 5-percentage-point increase found at the FBS>=110 threshold. The use of OGTT also appears to increase at the threshold, although we cannot be definitive without a regression analysis. The number of DM visits and DM-related outpatient medical expenditure may also slightly increase at the threshold.

Local-linear regression results reported in Table 3 show that the probability if a DM visit and the use of OGTT examination significantly increase by 4 and 0.5 percentage points, respectively. However, the number of DM visits and DM-related outpatient medical expenditure do not significantly increase at this threshold. As before, we also estimated the model using different bandwidths, polynomials, and with and without covariates. The results are robust as shown in Figure A9 (with covariates) and Figure A10 (without covariates). Thus, medical care utilization does increase in some models, but the results are not robust across measures and specifications at the "diabetic type" threshold.

At first glance, it appears counter-intuitive to find weaker responses at FBS>=126 mg/dl than at FBS>=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 "diabetic type" signal because very few checkup reports adopt FBS=126 mg/dl as a 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, and conveyed in a clear and understandable way to individuals.

Health Behavior. Turning to the effects on health behaviors, Figure 11 indicates that the signal continues to have little effect on any measure of health behavior, although the probability of "walk or exercise" and "drink every day" may be somewhat affected at the threshold. Panel B of Table 3 reports local-linear regression results that use the preferred specification. It shows that "walk or exercise" increases at the threshold. However, all other health behaviors are little affected by the signal at this margin. As before, we also performed robustness checks using different bandwidths and

polynomials and with and without covariates. Figures A11 and A12 indicate that the RD coefficient for "walk or exercise" becomes significant for local-linear models with relatively wider bandwidths. For other variables, we continue to find no significant effect. Thus, we conclude that, unlike the "borderline type" signal that had no effect at all on any health behavior, we have weak evidence that the "diabetic type" signal affects some types of health behavior but the result is not robust across behaviors or empirical specifications.

Health Outcomes. Figures 12 and 13 show the results for health outcomes. In Figure 12, intermediate health outcomes such as BMI and FBS are smooth around the FBS=126 mg/dl threshold and there is no clear evidence that the signal improves health outcomes. In Figure 13, "risk for stroke" and "risk for non-CV mortality" appear to decrease somewhat at the threshold. However, as we report in Panel C of Table 3, results from local-linear regressions do not indicate that the FBS=126 mg/dl signal significantly affects either of these outcomes even at the 10 percent confidence level. As before, we performed robustness checks using different bandwidths and polynomials and with and without covariates. As reported in Figures A13 and A14, as well as Figures A15 and A16, there is no evidence that the "diabetic type" signal affects health outcomes, whether they are measured by intermediate health outcomes or by predicted risks of mortality and DM complications.

5.2. Cost Effectiveness of Preventive Medical Care

In this section, we examine the cost effectiveness of preventive medical care by implementing a fuzzy RD approach.

5.2.1. First-stage results

For the first-stage regression, we regress two medical care utilization variables, i) Number of DM visits and ii) DM-related outpatient medical spending, on FBS thresholds. Because the first-stage regression results are the same as the ones reported in Section 5.1.1, we only briefly review the results.

As reported in Panel A of Table 3, we found that the Number of DM visits, the first dependent variable, significantly increases at the FBS=110 mg/dl threshold. We also observed a significant increase in DM-related outpatient

medical spending at this threshold. These results indicate that our instrumental variable (the FBS>=110 mg/dl signal) is correlated with the endogenous variable (medical care utilization). In contrast, for the FBS=126 mg/dl threshold, we did not find evidence that the "diabetic type" health signal increases both dependent variables (see Panel A of Table 3). For this reason, we focus on the FBS=110 mg/dl threshold to study the cost effectiveness of preventive medical care.

As discussed in Section 3.1, one concern regarding our "fuzzy" RD approach is that after receiving a health signal, individuals may alter health behavior, which in turn may affect health outcomes. If so, this makes it difficult to identify the impact of medical care utilization on health outcomes. For the FBS>=110 mg/dl threshold, however, this is not a serious concern in our case. As reported in Panel B of Table 3 and Figure A3, the "borderline type" signal has virtually no effect on health behaviors, which provides confidence about the second qualification for the instrument, that the health signal affects health outcomes only through medical care utilization.

5.2.2. Second-stage results

Table 4 reports the results from the second-stage regressions. 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. We only report the results for the "borderline type" threshold because in Section 5.1.2 we did not find that the two medical care utilization measures increase at the "diabetic type" threshold.[23] Panel A of Table 4 shows that none of the coefficients are negative and statistically significant. Thus, we find no empirical evidence that the significant increase in DM-related medical care around the threshold improves health outcomes. Robustness check results with various bandwidths and polynomials reported in Appendix A17 and A18 further confirm this observation. Moreover, while we only look at short-run effects in this section, as we report later in Section 6.2, we also do not find evidence of improvements in long-run health outcomes. Therefore, the medical care triggered by the FBS>=110 mg/dl threshold does not appear to be effective,

[23] Nonetheless, we also estimated the regressions for FBS>=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.

much less cost effective.

Note that we have more than 55,000 observations in our regressions even when we use a narrow window width of 5 mg/dl. Moreover, the first-stage F-statistics are bigger than 20 for local-linear regressions, 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.

6. Additional analysis

6.1. Effects on first-time signal receivers

One might expect that some people are not health conscious and they may routinely ignore health warning signals. 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 analyses by excluding individuals who received the "borderline type" signal in the previous year.

In Figure A19, we report the impact of "borderline type" signal of FBS=110 mg/dl on 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 those reported in Figure 6. For example, as shown in Figure A19, the probability of visiting a doctor for DM increases by approximately 6 percentage points among those who did not receive a signal the previous year, slightly higher than the 5-percentage-point increase found in Figure 6.

The results for health outcomes reported in Figures A20 and A21 are similar to our previous results; there is no evidence that the "borderline type" signal affects health outcomes. These results are confirmed by local-linear regressions (not reported). Thus, as expected, individuals who did not receive the warning last year respond more to the signal; however, the additional medical care utilization still does not seem to improve health outcomes. These results provide further confidence in our conclusion that preventive DM care around the "borderline threshold" of FBS=110 mg/dl is neither effective nor cost effective.

We now turn to the results for the "diabetic type" threshold. Regarding medical care utilization, as shown in Figure A22, we continue to find that the

probability of a DM visit increases at the threshold but other utilization variables are little affected in this case. Turning to health outcomes shown in Figures A23 and A24, a notable departure from our previous analyses is that we find some evidence that the "diabetic type" signal reduces FBS values about 4 mg/dl (which is about a 4% reduction from the mean at FBS=125) when the individual did not receive a signal in the previous year; for the majority of local-linear regressions and some quadratic regressions, the RD coefficients are significantly negative at the 5% confidence level. For all other health outcomes, however, the effects are not distinguishable from zero. Thus, the "diabetic type" signal appears to have some influence on health outcomes for those who did not exceed the threshold in the previous year, although the results are not robust across bandwidths and alternative health outcomes.

6.2. Longer-run effects on health outcomes

So far, we have looked at short-run effects of health signals and found no robust 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 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. We use the full sample for this analysis. Figure A25 shows the effects on intermediate health outcomes three years after a checkup: we find no apparent discontinuity at the threshold. The results for predicted risks of mortality and complications are similar, as shown in Figure A26. The results for five years after a checkup are similar, as reported in Figures A27 and A28 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.

6.3. 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 III 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 A29, there is no clear evidence that the "borderline type" and "diabetic type" signals affect the WHO risk measure with different bandwidths and polynomials. These results provide additional support for our finding that the health signals have little effect on health outcomes.

7. Discussion

There are some differences in results between our study and the most closely related one, Kim et al. (2017) on health screening in Korea. First, at the lower "pre-diabetes" threshold, utilization clearly increases in Japan but not in Korea. In both countries, at this threshold there is no intervention such as a phone call from a physician or nurse urging a visit. One reason for the differences in response may be that the multiple thresholds in the Korean checkup report—with only a "yellow light" and no free secondary exam at the lower threshold, compared to a "red light" and offer of a free second exam at the 126 threshold—lead individuals to infer that the "pre-diabetes" signal is not serious enough to make a physician visit. On the other hand, because most Japanese reports only have the "pre-diabetes" threshold, people may take it more seriously and visit physician offices.[24] In addition, differences in the samples may contribute to the differing results. For example, the Korean sample has many older, unemployed or retired individuals, whereas our Japan sample includes only working-age adults. Also, whereas health checkups are mandatory in Japan, the Korean sample consists of individuals who chose to take a checkup. Additionally, our sample has more than four times the number of observations around the "pre-diabetes" threshold as the Korean sample does, and this might have made the Japan estimate at that

[24] Alternatively, the Korean estimates may be biased downwards because as the authors note the secondary examination after a screening is not counted as an outpatient visit in Korea.

threshold more precise.[25]

At the higher threshold corresponding to "diabetic type," the Korean study has more convincing evidence on the effect of the threshold on the use of DM medications and intermediate health outcomes (such as BMI and waist circumference), although the evidence is still not robust across different measures. For example, FBS values are little affected by the "diabetic type" signal in the Korean study and they do not find evidence that the signal affects any health behavior.[26] We find that health outcomes improve in some specifications as discussed in Section 6.1; but our results are generally less clear cut at this threshold, probably because few checkup reports adopt 126 mg/dl as a threshold and not many people receive that signal.

Despite these differences, consistent economic insights flow from the results of both studies: namely, that individuals only respond to a clear signal (FBS=110 for Japan's "high" report, FBS=126 for Korea's "red light" report), and such information is more conducive to health improvement when combined with a follow-up intervention (as in Korea for the higher threshold)—although even then, the effects attenuate over time.

8. Conclusion

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 its 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 preventive health signals and increase their probability of visiting a physician. For example, we estimate that medical spending increases approximately 40,000 JPY (or $400) per year

[25] To analyze the effect of the signal on the "number of outpatient visits for DM," we have 120 thousand observations at the FBS=110 threshold with a 5 mg/dl bandwidth, whereas the Korean study has 52 thousand observations at the FBS=100 threshold with a bandwidth twice as wide, 10 mg/dl.

[26] The authors speculate that the reductions in BMI and waist circumference stem from behavior change (reduction in calorie intake); however, since unexplained significant weight loss is itself a symptom of diabetes, the evidence is not so clear in this case. (The Korean form for the secondary exam includes a warning "if you have any specific problems (e.g. excessive weight loss), please talk to your physician"; and the authors do not find any weight loss for non-diabetes health signals such as from the obesity screening.)

for those who respond to the FBS ≧ 110 pre-diabetes signal. This result confirms that health signals can potentially help prevent chronic disease by bringing people to physicians' offices and leading to follow-on care. However, this result also implies that if thresholds do not reflect the cost-effectiveness of preventive care, they may exacerbate 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, importantly, we do not find evidence that additional medical care triggered by health signals is effective (much less cost effective) at the "borderline type" threshold. We do find substantial increases in DM-related medical care utilization. However, health outcomes did not improve either for physical measures (biomarkers) or for predicted risks of mortality and serious DM complications. Thus, we find no evidence that additional preventive medical care triggered by the health signal is worth its cost. Our results suggest that the current "pre-diabetes" threshold may need to be reexamined from the perspective of cost effectiveness.

Our third main finding is the lack of response at the higher threshold that corresponds to "diabetic type." With no robust evidence that medical care utilization increases at this threshold, we could not assess cost effectiveness of such utilization. One of the reasons why the impact of the signal is weaker may be that few people actually receive this signal because employers rarely adopt this threshold in their checkup reports. Accordingly, individuals may not be informed that their check-up results imply they are a higher-risk "diabetic type." When multiple diagnosis thresholds are meaningful for a condition, it is important that individuals receive separate health signals according to their level of risk, with simple and eye-catching reports that do not bury important health information in fine print or with unintelligible jargon.

These results reinforce previous findings in the literature that people do respond to health signals under some circumstances, resulting in greater use of preventive services or other medical care (Zhao et al. 2013, Kim et al. 2017), but medium- and long-run outcomes often do not improve (Oster 2017, Kim et al. 2017). We contribute to the literature by not only estimating the effect of health signals on utilization and health behavior but also analyzing cost effectiveness of preventive care. Clinicians have pointed out the potential for

overtreatment when a large fraction of the population is classified as "borderline type" (Yudkin and Montori 2014). Our results are also consistent with such a concern.

Our results for both clinical thresholds suggest that a "traffic light" system of reporting or other use of behavioral insights for the check-up reports might significantly enhance the effectiveness (and cost-effectiveness) of Japan's existing investments in mandatory check-ups. Predicted risks—such as those for stroke and mortality that we estimate with a Japanese-specific risk model—are increasingly used for translating the health implications of multiple sub-optimal risk factors into understandable language for patients, and thus might be important to include in check-up reports to improve the framing and salience of clinical measurements and thereby enhance the cost-effectiveness of mandatory checkups. Such risk prediction models are currently used in the Korean national screening program reports as well as in the US for categorizing the severity and appropriate management of the most common chronic disease, hypertension (Whelton et al. 2017).

More generally, there are a large number of diagnosis thresholds for multiple conditions 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.

Appendix I: Definitions of variables

Outcome variables	Definition
■ Any DM Visit	▪ The rate of a diagnosis of diabetes (or at least rule-out diagnostic testing for diabetes) within 1 year of the index checkup ▪ "diagnosis" is defined by the following 2 conditions: (1) The individual has a claim that includes DM as one of the conditions and (2) Consultation days > 0. ▪ "DM" are diseases categorized as E10~E14 in ICD10 code.
■ Number of DM Visits	▪ Consultation days for DM within 1 year of the checkup
■ OGTT examination	▪ Rate of conducting an oral glucose tolerance test (OGTT) within 1 year of the checkup
■ DM-related outpatient medical spending	▪ "DM-related outpatient medical spending" is 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 "DM drug" as one of drugs) ▪ "DM drugs" are drugs categorized as A10 in ATC code. Note:1 point = 10 YEN
■ Walk or Exercise ■ Smoke ■ Drink everyday ■ Eat after dinner	▪ Self-report of health habits as measured in the health survey associated with the checkup a year after the index check-up ▪ "Walk or Exercise" = 1 if the individual reports exercising enough to work up a sweat for 30 minutes or more per day & 2~7 days per week. ("Walk or Exercise" = 0 otherwise.) ▪ "Drink everyday" = 1 if he/she drinks every day. ("Drink everyday" = 0 otherwise.) ▪ "Smoke" = 1 if he/she has a habit of smoking. ("Smoke" = 0 otherwise.) ▪ "Eat after dinner" = 1 if he/she eats a midnight snack 3 days or more per week. ("Eat after dinner" = 0 otherwise.)
■ FBS / HbA1c / BMI / SBP	▪ The values of FBS / HbA1c / BMI / SBP measured at the checkup a year after the index checkup
■ Risk for Stroke / Risk for CHD / Risk for non CV mortality	▪ 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 ▪ These predicted risks are calculated from the JJRE equations as developed by Tanaka et al. (Diabetes Care 2013)

Appendix II: 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_{1c}, 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, non-cardiovascular 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 non-cardiovascular mortality in our data. We do not estimate risk of overt nephropathy or progression of retinopathy, 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.[1]

[1] However, to run the JJRE SAS model requires inputting values for all risk factors

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 calculated as current year – min(FIRSTDX). Patients without two separate non-suspect (i.e. suspicion

and outputting all five predicted risks; therefore, we utilize the JJRE default values for the risk factors that are missing in the JMDC data.

flag[2]=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>=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 formula, weight (kg) / [height (m)]2 after converting height in centimeters to meters.

[2] 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).

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).

Appendix III. 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.

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 >= 40%

TheWHO/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 >= 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 = 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 —> 4
>=5 to < 6 —> 5
>= 6 to < 7 —> 6
>= 7 to < 8 —> 7
>= 8 —> 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.

Almond, Douglass, Joseph J. Doyle, Jr., Amanda E. Kowalski, and Heidi Williams. "The Role of Hospital Heterogeneity In Measuring Marginal Returns To Medical Care: A Reply To Barreca, Guldi, Lindo, And Waddell." *Quarterly Journal of Economics* 126 (2011): 2125-2131.

American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2010;33(Suppl. 1):S62–S69.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance," *The Quarterly Journal of Economics* 1623 (2015): 1667.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. "Heaping-Induced Bias in Regression-Discontinuity Designs." *Economic Inquiry* 54.1 (2016): 268-293.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, Glen R. Waddell (2011) "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification," *The Quarterly Journal of Economics*, 126, 2117-2123.

Benjamin, Emelia J., Michael J. Blaha, Stephanie E. Chiuve, Mary Cushman, Sandeep R. Das, Rajat Deo, Sarah D. de Ferranti et al. "Heart disease and stroke statistics—2017 update: a report from the American Heart Association." *Circulation* 135, no. 10 (2017): e146-e603.

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.

Cohen, J. T., Neumann, P. J., and Weinstein, M. C. (2008). Does preventive care save money? Health economics and the presidential candidates. *New England Journal of Medicine*, 2008(358), 661-663.

Doyle, Joseph, John Graves, and Jonathan Gruber. *Uncovering Waste in US Healthcare*. No. w21050. National Bureau of Economic Research, 2015.

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.

Gillies, C. L., Lambert, P. C., Abrams, K. R., Sutton, A. J., Cooper, N. J., Hsu, R. T., ... & Khunti, K. (2008). Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. BMJ 336(7654), 1180-1185.

Greenland, Philip, and Eric Peterson. “The New 2017 ACC/AHA Guidelines “Up the Pressure” on Diagnosis and Treatment of Hypertension,” JAMA Published online November 20, 2017.

Herman, W. H., Hoerger, T. J., Brandle, M., Hicks, K., Sorensen, S., Zhang, P., ... & Ratner, R. E. (2005). The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. Annals of internal medicine, 142(5), 323-332.

Hoerger, T. J., Harris, R., Hicks, K. A., Donahue, K., Sorensen, S., & Engelgau, M. (2004). Screening for type 2 diabetes mellitus: a cost-effectiveness analysis. *Annals of Internal Medicine*, 140(9), 689-699.

Hoerger, T. J., Hicks, K. A., Sorensen, S. W., Herman, W. H., Ratner, R. E., Ackermann, R. T., ... & Engelgau, M. M. (2007). Cost-effectiveness of screening for pre-diabetes among overweight and obese US adults. *Diabetes care*, 30(11), 2874-2879.

Iizuka, T. 2012. “Physician Agency and Adoption of Generic Pharmaceuticals.” *American Economic Review*. 102(6):2826–58.

International Diabetes Federation (2017). IDF diabetes atlas 8th edition. Brussels:International Diabetes Federation. Available at https://www.idf.org/e-library/epidemiologyresearch/diabetes-atlas.html

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.

Kim, S., Kwon, S. and Subramanian, S.V., 2015. Has the National Cancer Screening Program reduced income inequalities in screening attendance in South Korea? *Cancer Causes & Control*, *26*(11), pp.1617-1625.

Kolesar, Michal, and Rothe, Christoph. 2017. “Inference in Regression Discontinuity Design with a Discrete Running Variable” Working paper.

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.

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

Ministry of Health, Labour and Welfare (MHLW), Government of Japan, "Health checkups for employees" (in Japanese) http://www.mhlw.go.jp/file/06-Seisakujouhou-11200000-Roudoukijunkyoku/0000103900.pdf. and "Health checkups for part-time employees" (in Japanese) http://www.mhlw.go.jp/bunya/koyoukintou/pamphlet/pdf/150330-1.pdf.

Miller, Y. D., and Dunstan, D. W. (2004). The effectiveness of physical activity interventions for the treatment of overweight and obesity and type 2 diabetes. *Journal of Science and Medicine in Sport*, 7(1), 52-59.

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

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.

Oster, Emily. 2017. "Diabetes and Diet: Purchasing Behavior Change in Response to Health Information." *American Economic Journal: Applied Economics*, forthcoming.

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.

Sacks, D. B. (2011). A1C Versus Glucose Testing: A Comparison. *Diabetes Care*, 34(2), 518–523. http://doi.org/10.2337/dc10-1546.

Seino, Yutaka, 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.

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.

UK Prospective Diabetes Study (UKPDS) Group. "Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33)." *The Lancet* 352.9131 (1998): 837-853.

Vermeire, E. I., Wens, J., Van Royen, P., Biot, Y., Hearnshaw, H., & Lindenmeyer, A. (2005). Interventions for improving adherence to treatment recommendations in people with type 2 diabetes mellitus. The Cochrane Library.

Whelton PK, Carey RM, Aronow WS, et al. 2017. ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines [published online November 13, 2017]. *Hypertension*. doi:10.1161/HYP.0000000000000065

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 (accessed October 11, 2016).

Yudkin, John S, and Victor M Montori, "The epidemic of pre-diabetes: the medicine and the politics." *BMJ* 2014; 349 doi: https://doi.org/10.1136/bmj.g4485 (Published 15 July 2014).

Zhao, Meng, Yoshifumi Konishi and Paul Glewwe, "Does Information on Health Status Lead to a Healthier Lifestyle? Evidence from China on the Effect of Hypertension Diagnosis on Food Consumption" *Journal of Health Economics* (2013) 32(2), p367-385

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

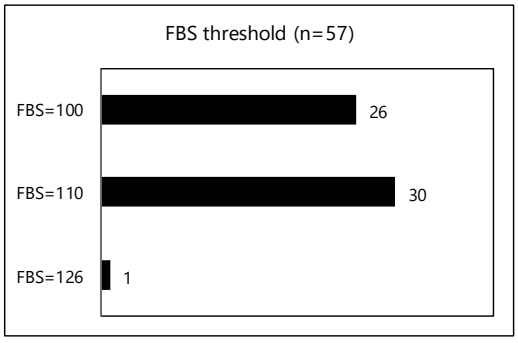
		2012/04/20	2011/04/10	2010/04/15	2009/04/18	Reference(or Normal)
Age		47	46	45	44	
Physical Examination	Height(cm)	171.8	171.9	171.8	171.8	
	Weight(kg)	65.5	66.7	65.2	60.1	
	BMI	22.2	22.6	22.1	20.4	
	Waist Circumference	72.4	72.8	71.3	70.5	
Eye sight	Without glasses(R/L)	-	-	-	-	
	With glasses(R/L)	0.9/1.0	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	H 112	108	104	H 115	≦109(mg/dL)
	HbA1c(NGSP)	H 5.9	5.5	5.2	H 6.0	≦5.8(%)
Anemia test	RBC	470	465	480	472	≧400,≦539(10^4/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	(−)	(−)	(−)	(−)	
…	…	…	…	…	…	…

Evaluation	
The following items are out of normal range.	
Test Items	Comments
FBS	Re-testing required. You may have a re-examination at a medical institution. Lifestyle advice will also be given by a physician, a nurse, or a dietician.

Physician's Signature	**********	Office/Institutions	*** clinic

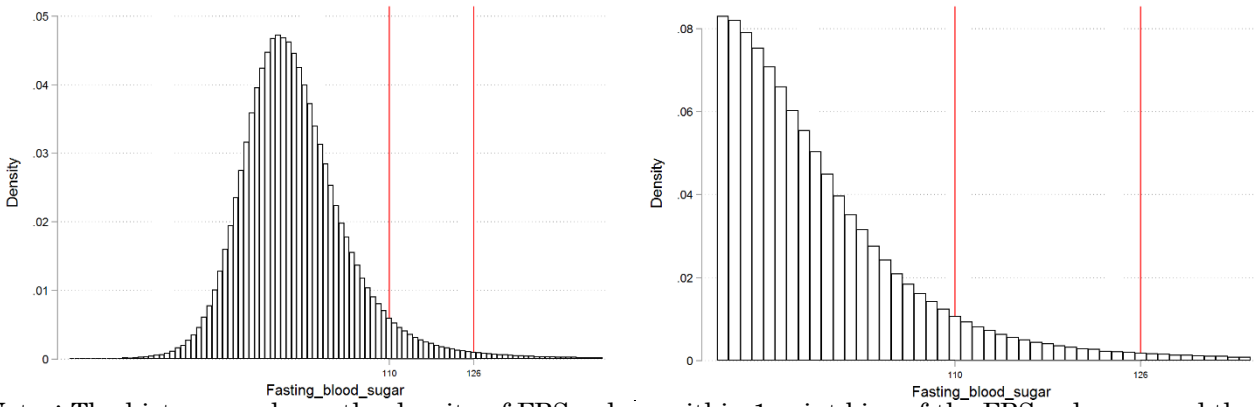
Notes: This figure shows a typical checkup report that employees receive. The original language is in Japanese.

Figure 2. Distribution of threshold values used in checkup reports



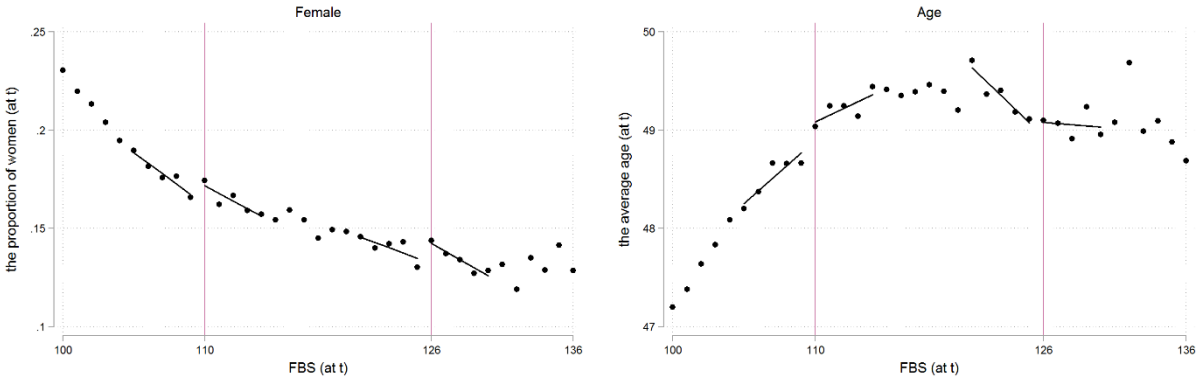
Notes: This figure shows the "empirical distribution" of threshold values used in checkup reports obtained from the authors' internet search.

Figure 3. Distribution of FBS values



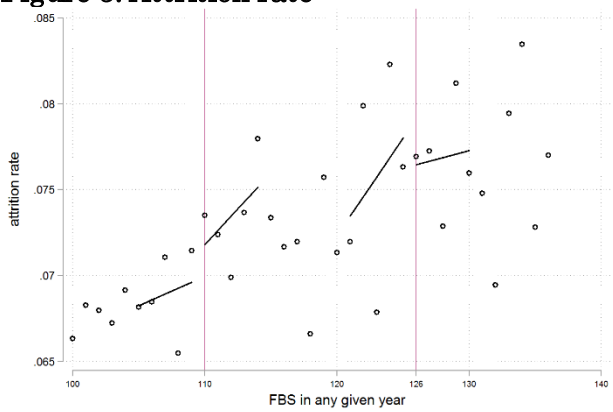
Notes: The histogram shows the density of FBS values within 1 point bins of the FBS value around the "borderline type (FBS=110 mg/dl)" and "diabetic type (FBS=126 mg/dl)" thresholds.

Figure 4. Covariates Balance



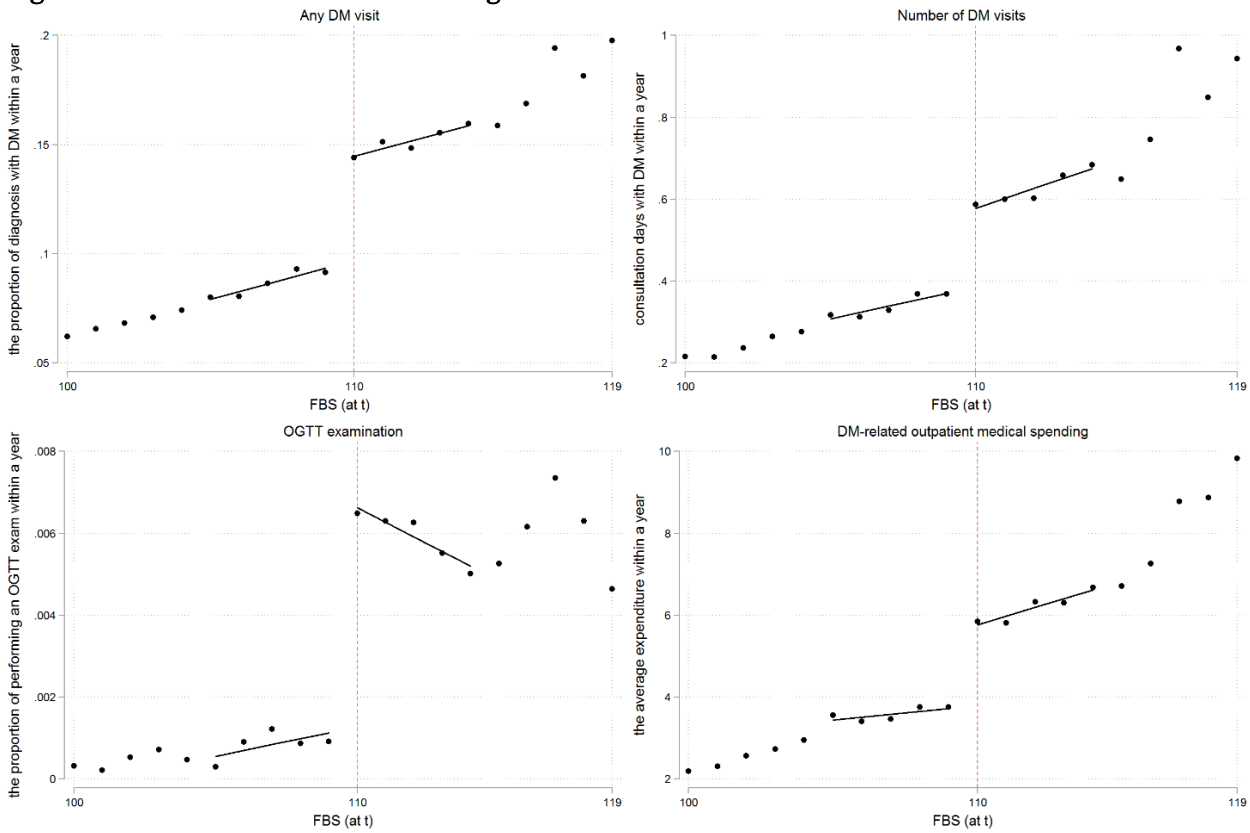
Notes: The scatter plot shows the mean of the covariate within 1 point bins of the FBS value. The vertical lines indicate the two threshold values for DM diagnosis. We fit the values using a linear function within 5 mg/dl of FBS values before and after the thresholds.

Figure 5. Attrition rate



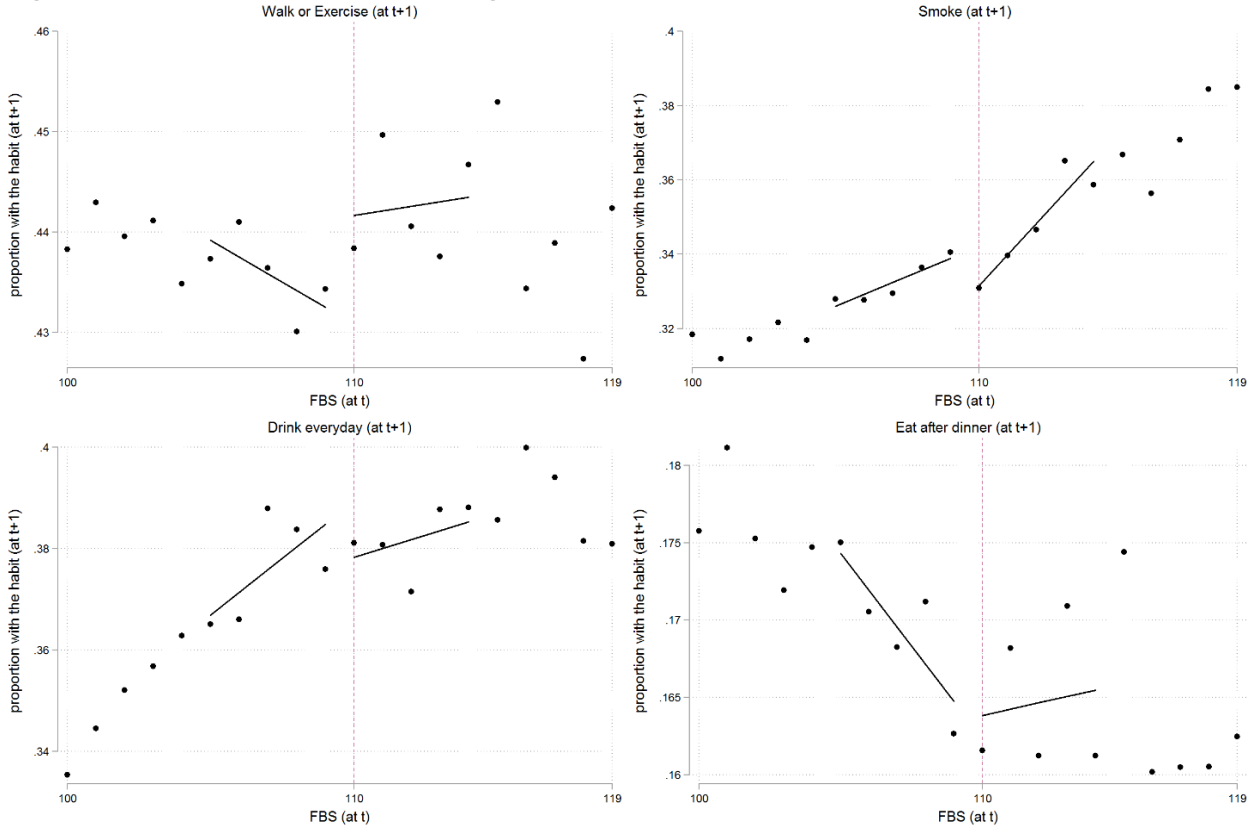
Notes: The scatter plot shows the mean of attrition within 12 months after a checkup, using 1 point bins of the FBS value. The vertical lines indicate the two threshold values for DM diagnosis. We fit the values using a linear function within 5 mg/dl of FBS values before and after the thresholds.

Figure 6. Effects of the FBS≧110 signal on medical care utilization



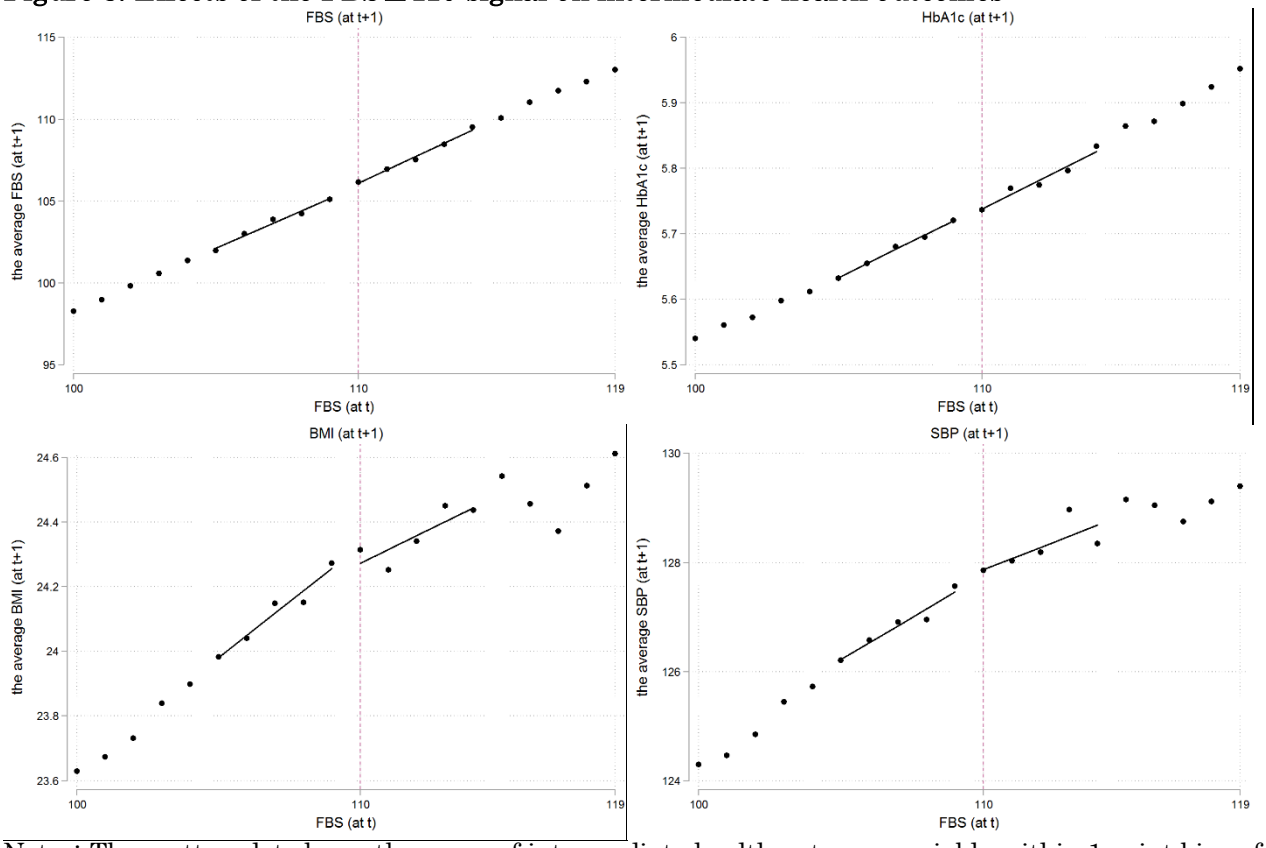
Notes: The scatter plot shows the mean of medical care utilization variable within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 7. Effects of the FBS≧110 signal on health behavior



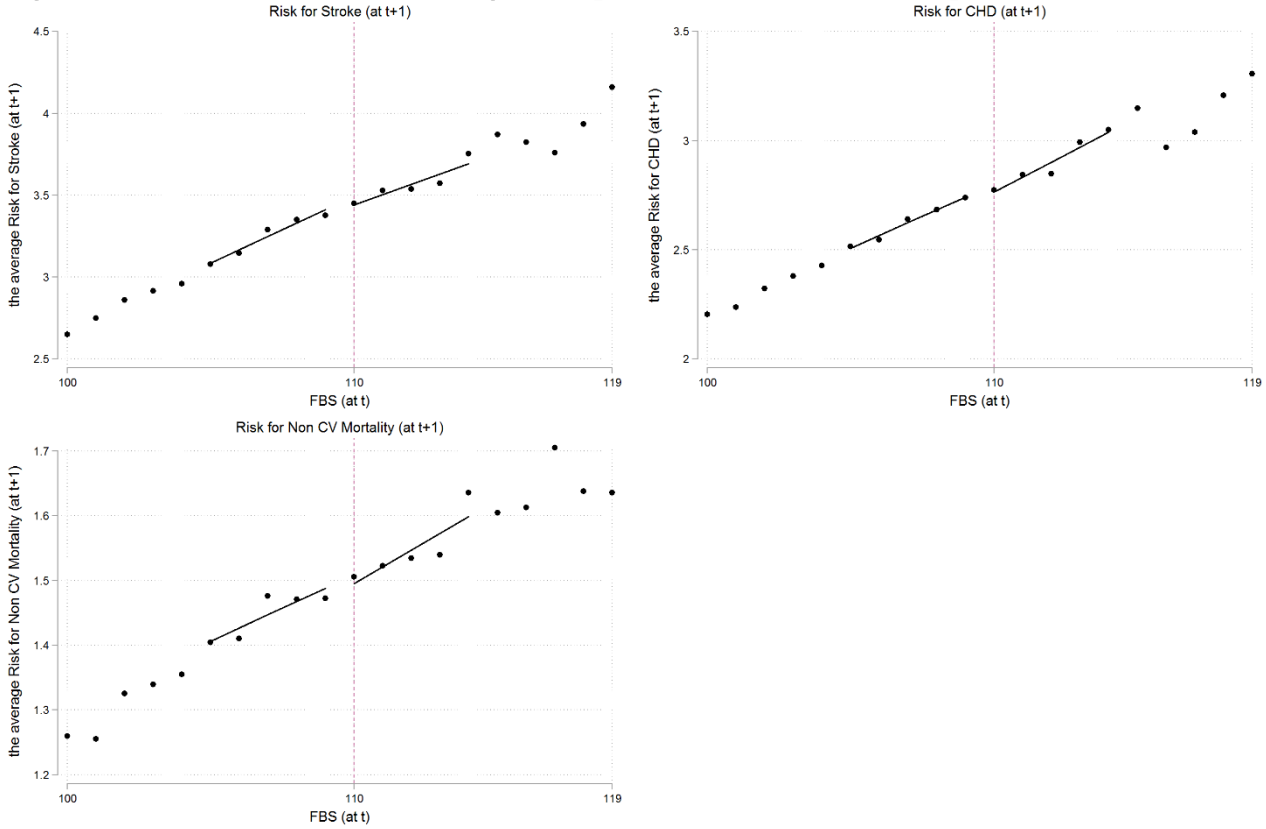
Notes: The scatter plot shows the mean of health behavior variable within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 8. Effects of the FBS≧110 signal on intermediate health outcomes



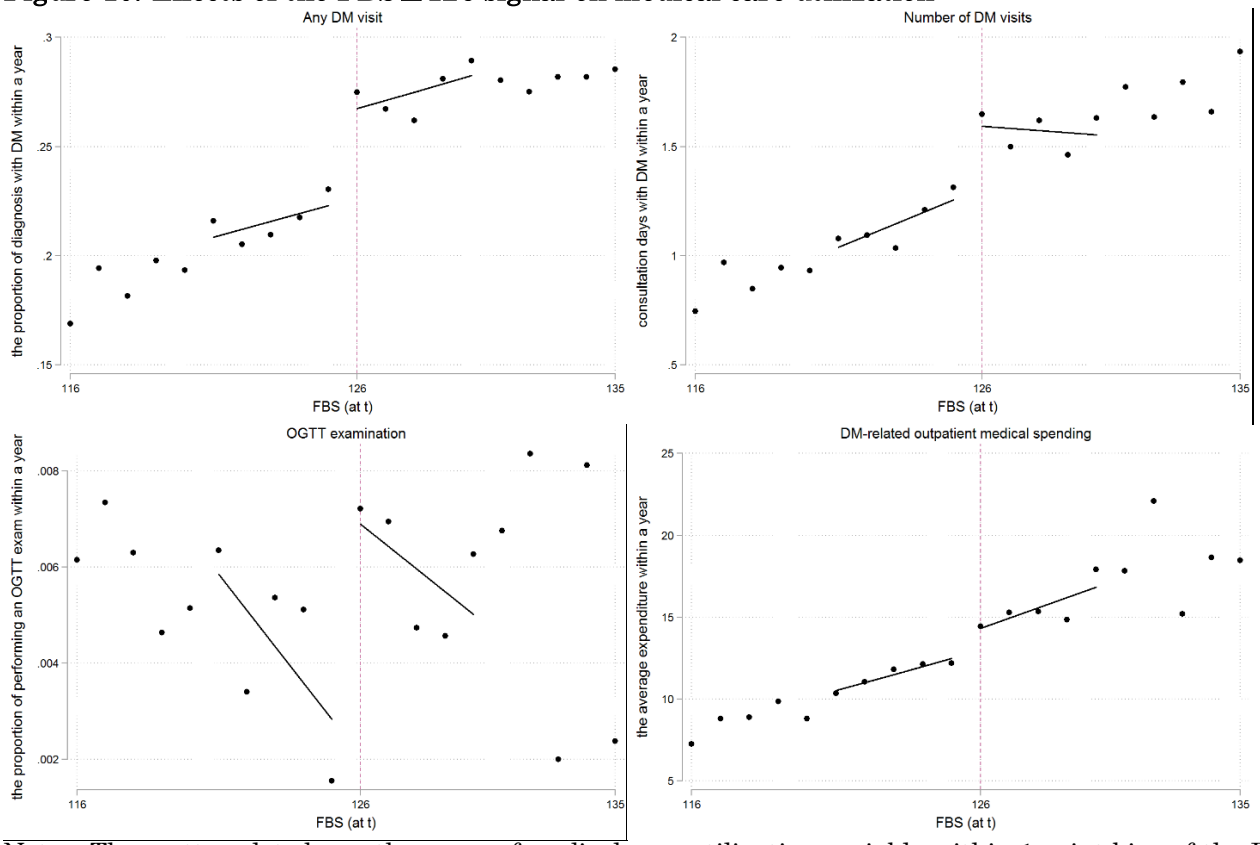
Notes: The scatter plot shows the mean of intermediate health outcome variable within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 9. Effects of the FBS≧110 signal on predicted risks of DM complications



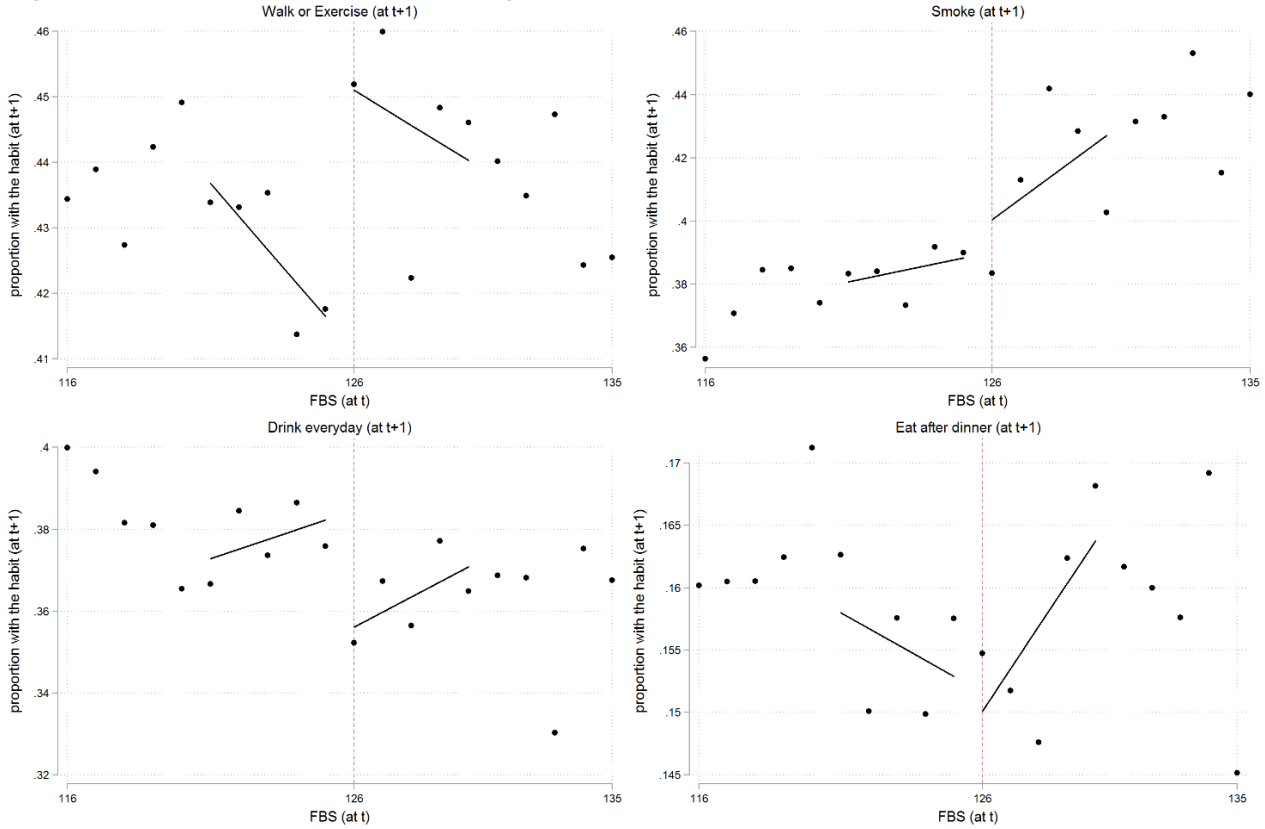
Notes: The scatter plot shows the mean of predicted risk of DM complication within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 10. Effects of the FBS≧126 signal on medical care utilization



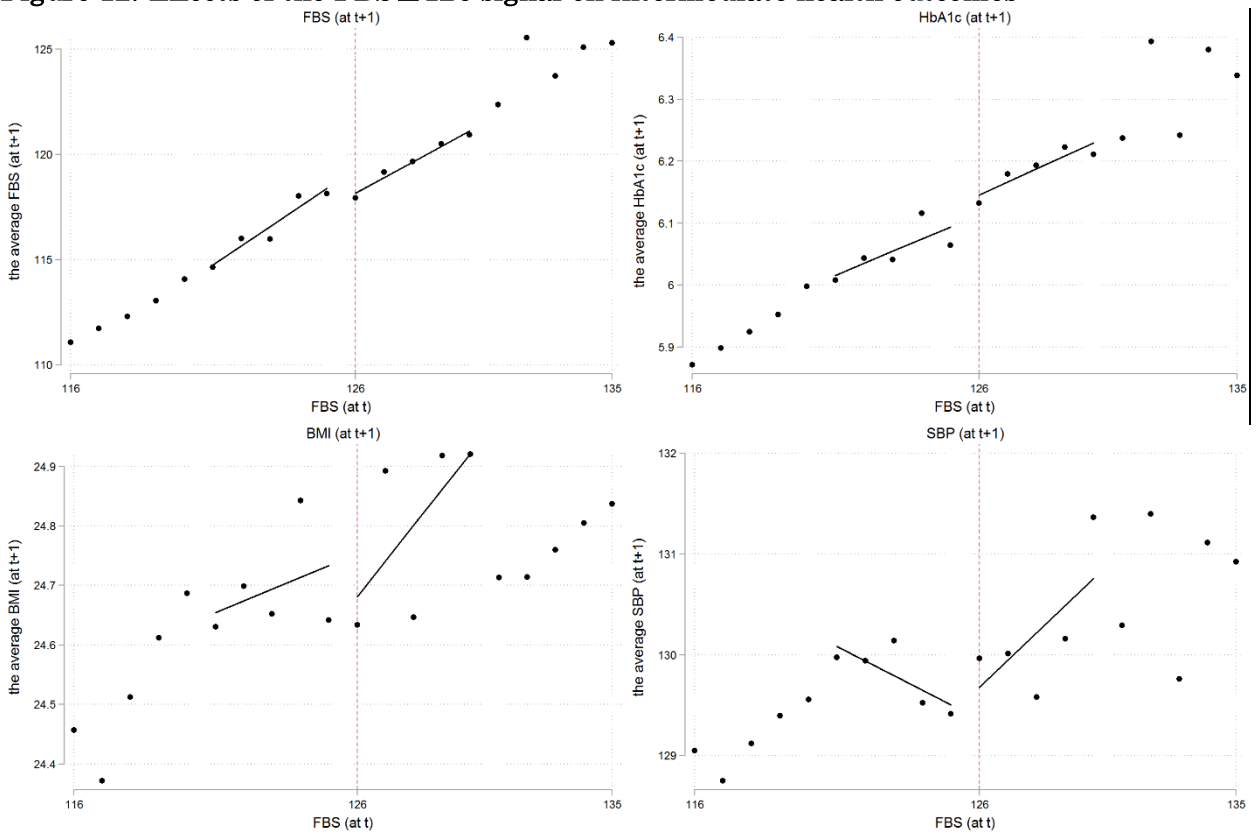
Notes: The scatter plot shows the mean of medical care utilization variable within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 11. Effects of the FBS≧126 signal on health behavior



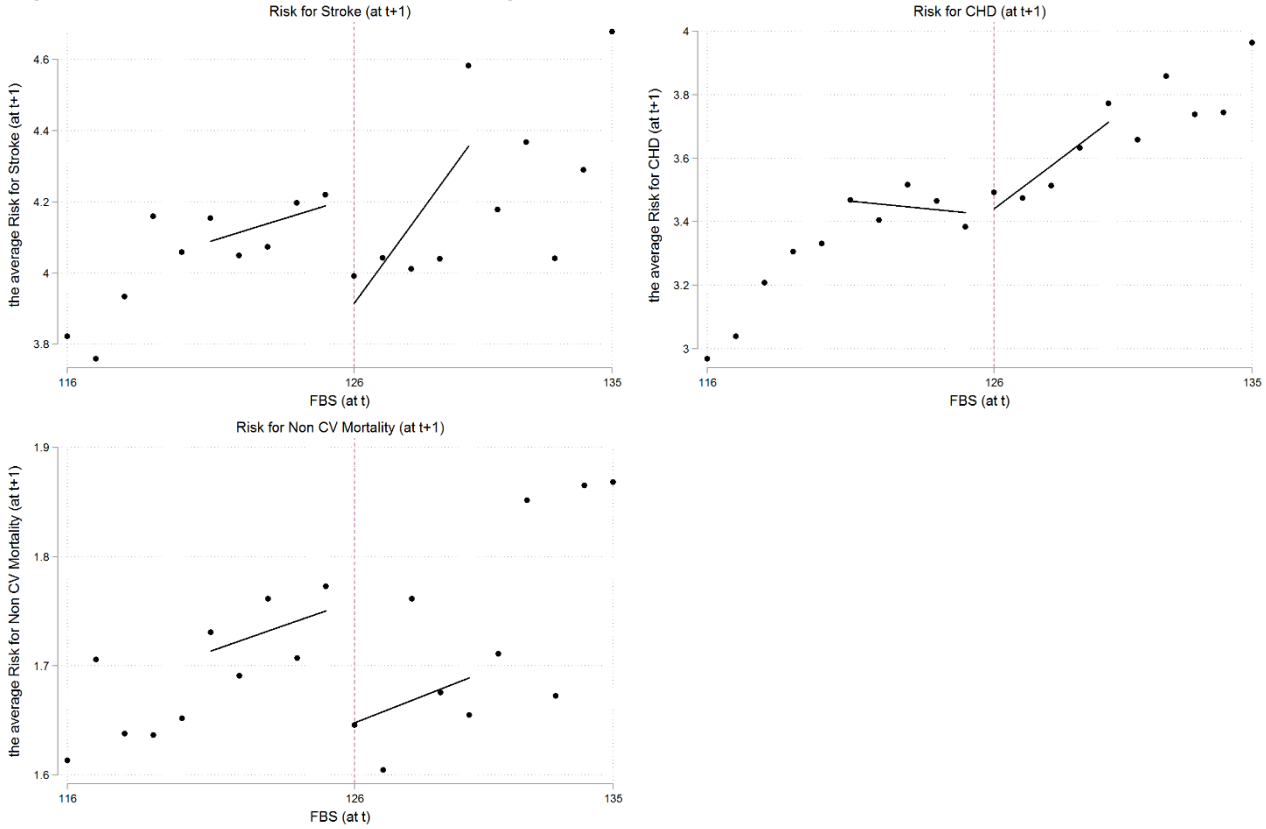
Notes: The scatter plot shows the mean of health behavior variable within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 12. Effects of the FBS≧126 signal on intermediate health outcomes



Notes: The scatter plot shows the mean of intermediate health outcome variable within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 13. Effects of the FBS≧126 signal on predicted risks of DM complications



Notes: The scatter plot shows the mean of predicted risk of DM complication within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Table 1. Summary statistics

	N	mean	std.err.
【running variable (at t)】			
FBS (mg/dl)	1,741,209	92.9	14.2
【covariates】			
Age	1,741,209	45.050	8.270
Female	1,741,209	0.363	0.480
【medical care utilization (between t and t+1)】			
Any DM visit	1,741,209	0.059	0.236
Number of DM Visits	1,741,209	0.240	1.805
OGTT examination	1,741,209	0.001	0.024
DM-related outpatient medical spending (in 1,000 JPY)	1,741,209	2.63	25.57
【health behavior (at t+1)】			
Walk or Exercise	967,757	0.439	0.496
Smoke	1,146,643	0.303	0.459
Drink everyday	1,073,649	0.267	0.442
Eat after dinner	940,030	0.182	0.385
【intermediate health outcomes (at t+1)】			
FBS	1,141,578	93.3	14.1
HbA1c	1,015,687	5.5	0.5
SBP	1,209,467	120.5	15.5
BMI	1,209,582	22.8	3.4
【predicted risks of DM complications (at t+1)】			
Risk for Stroke (%)	496,742	2.47	3.08
Risk for CHD (%)	496,742	2.04	1.89
Risk for Non CV Mortality (%)	496,742	1.17	1.28
WHO risk (0~4)	725,257	0.03	0.24

Table 2. McCrary test

Cutoff	Bandwidth				
	2	4	6	8	10
105	-1.171	-1.548	-1.439	0.282	1.394
108	0.491	1.805	**3.362**	**5.215**	**8.763**
110	-1.424	**-2.176**	-1.471	1.358	**5.841**
112	0.226	1.647	**2.431**	**4.348**	**9.186**
115	0.531	0.400	1.609	**4.584**	**9.040**
120	0.426	0.099	0.639	**2.069**	**4.913**
125	-1.200	0.241	0.561	1.323	**2.901**
126	1.010	0.585	0.706	1.141	**2.658**
130	1.143	1.497	**2.084**	**2.290**	**3.013**

Note: This table provides tvalues of the McCrary test. Bold numbers indicate tvalues greater than 2.

Table 3. Effects of "borderline type" and "diabetic type" signal

	FBS at 110 mg/dl			FBS at 126 mg/dl		
	coefficient	mean at 109	obs.	coefficient	mean at 125	obs.
Panel A: medical care utilization (at t+1)						
Any DM visit	0.047***	0.091	120,735	0.040***	0.230	19,241
	(0.004)			(0.013)		
Number of DM visits	0.188***	0.368	120,735	0.264*	1.312	19,241
	(0.027)			(0.148)		
OGTT examination	0.005***	0.001	120,735	0.005**	0.002	19,241
	(0.001)			(0.002)		
DM-related outpatient medical spending	1.933***	3.753	120,735	1.225	12.169	19,241
	(0.333)			(1.228)		
Panel B: health behavior (at t+1)						
Walk or Exercise	0.010	0.434	66,870	0.041**	0.418	10,461
	(0.008)			(0.020)		
Smoke	-0.006	0.341	80,640	0.013	0.390	13,322
	(0.007)			(0.017)		
Drink every day	-0.010	0.376	75,048	-0.019	0.376	12,201
	(0.007)			(0.017)		
Eat after dinner	0.001	0.163	64,924	-0.003	0.158	10,100
	(0.006)			(0.015)		
Panel C: health outcomes (at t+1)						
FBS	0.163	105	80,074	-1.018	118	12,920
	(0.183)			(0.789)		
HbA1c	-0.004	5.7	73,907	0.034	6.1	13,045
	(0.007)			(0.029)		
BMI	-0.036	24.3	85,306	-0.077	24.6	14,085
	(0.051)			(0.134)		
SBP	0.094	128	85,288	0.396	129	14,081
	(0.215)			(0.555)		
Risk for Stroke	-0.028	3.376	55,475	-0.163	4.220	9,171
	(0.059)			(0.170)		
Risk for CHD	-0.024	2.737	55,475	0.118	3.383	9,171
	(0.032)			(0.104)		
Risk for Non CV mortality	0.005	1.472	55,475	-0.035	1.773	9,171
	(0.018)			(0.051)		

Note: This table shows the results from local-linear regression using a rectangular kernel with a 5 mg/dl bandwidth and covariates. Only the coefficients for the RD dummies are reported. Standard errors, corrected for clustering at the person level, are in parentheses. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.

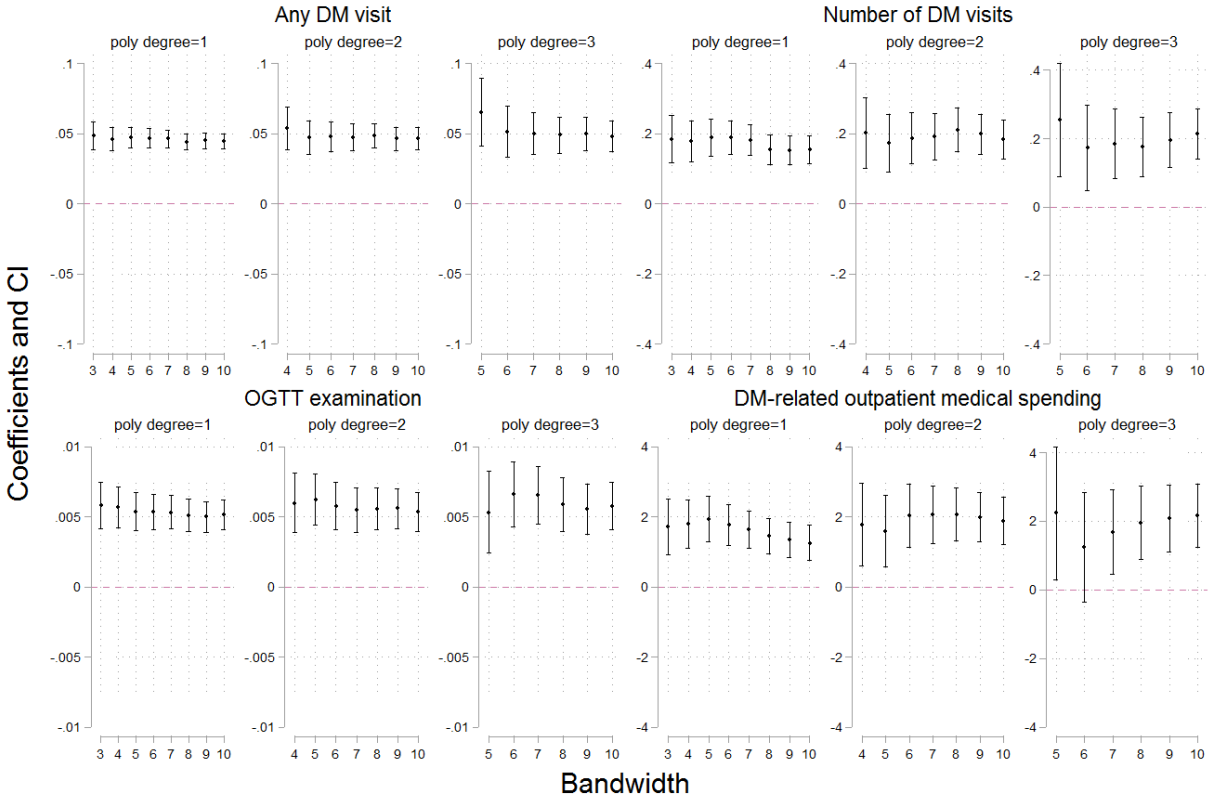
Table 4. Results from second-stage regressions

	FBS at 110 mg/dl	
	coefficient	obs.
Panel A: Endogenous variable = Number of DM visits		
FBS (at t+1)	0.763	80,074
	(0.863)	
HbA1c (at t+1)	-0.020	73,907
	(0.034)	
BMI (at t+1)	-0.164	85,306
	(0.238)	
SBP (at t+1)	0.433	85,288
	(0.987)	
Risk for stroke (at t+1)	-0.109	55,475
	(0.234)	
Risk for CHD (at t+1)	-0.096	55,475
	(0.128)	
Non-CV mortality (at t+1)	0.020	55,475
	(0.071)	
Panel B: Endogenous variable = DM-related outpatient medical spending		
FBS (at t+1)	0.062	80,074
	(0.070)	
HbA1c (at t+1)	-0.002	73,907
	(0.003)	
BMI (at t+1)	-0.015	85,306
	(0.021)	
SBP (at t+1)	0.038	85,288
	(0.088)	
Risk for stroke (at t+1)	-0.010	55,475
	(0.022)	
Risk for CHD (at t+1)	-0.009	55,475
	(0.012)	
Non-CV mortality (at t+1)	0.002	55,475
	(0.007)	

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.

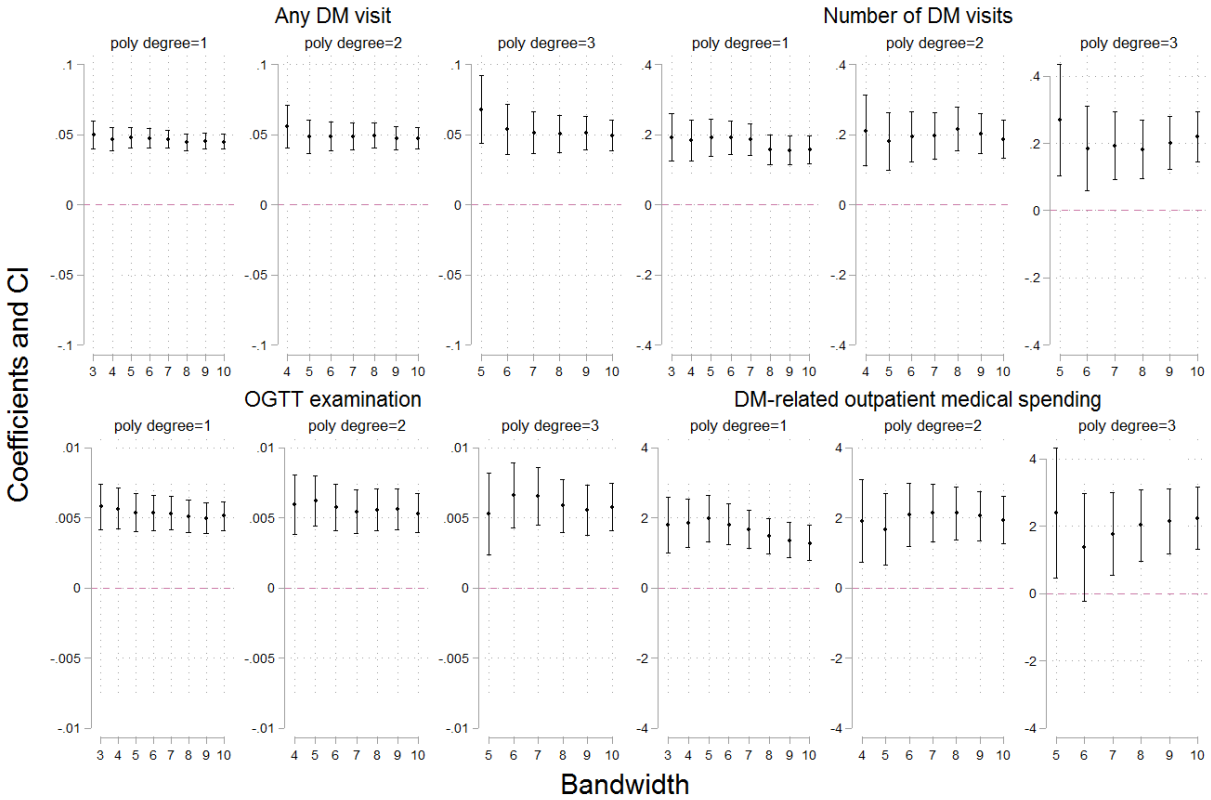
Online Appendix (Not for Publication)

Figure A1. Robustness of the effects of the FBS≧110 signal on medical care utilization



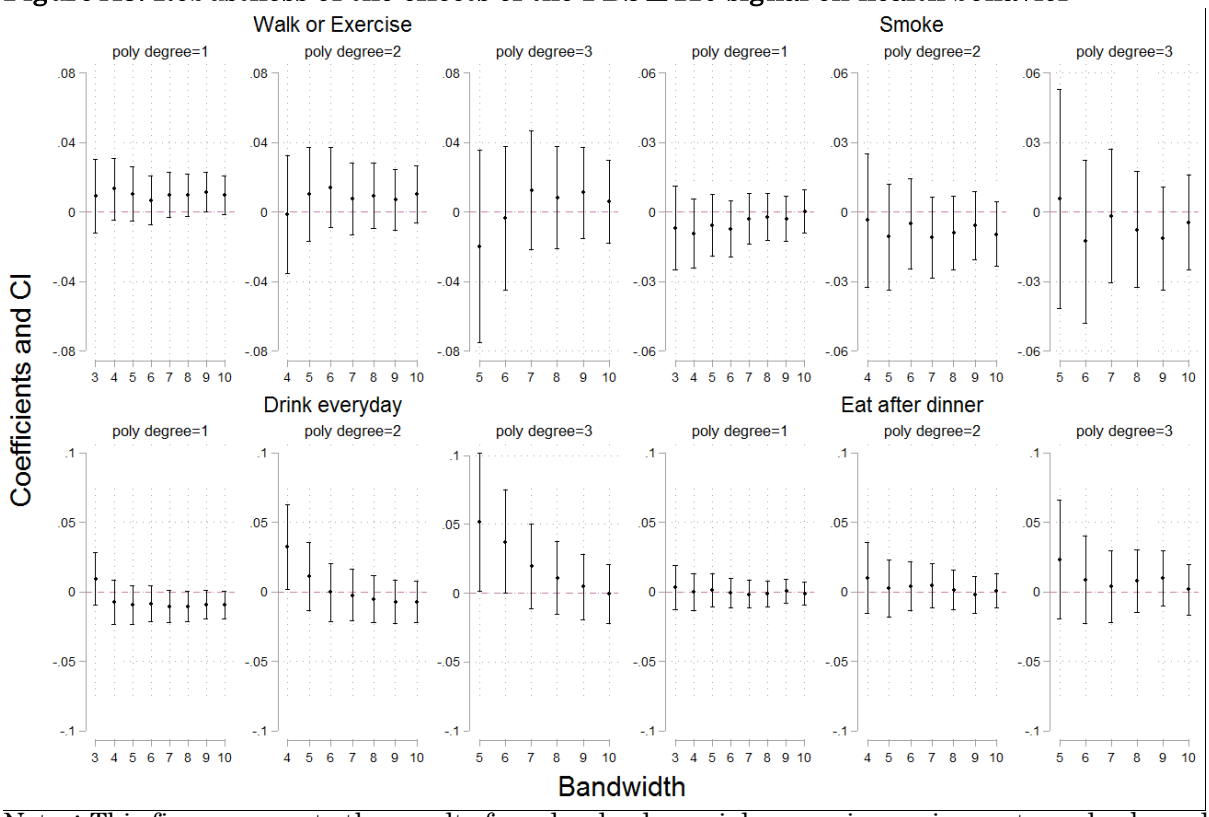
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A2. Robustness the effects of the FBS≧110 signal on medical care utilization (estimated without covariates)



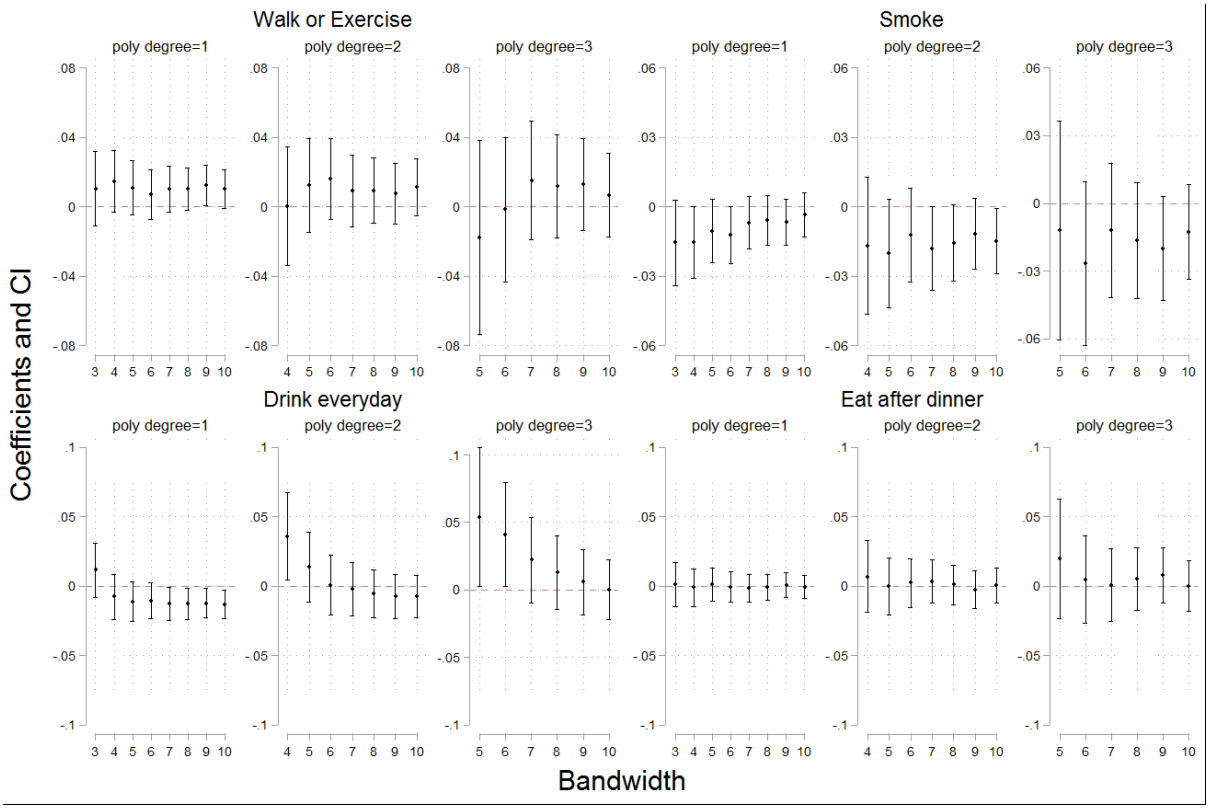
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A3. Robustness of the effects of the FBS≧110 signal on health behavior



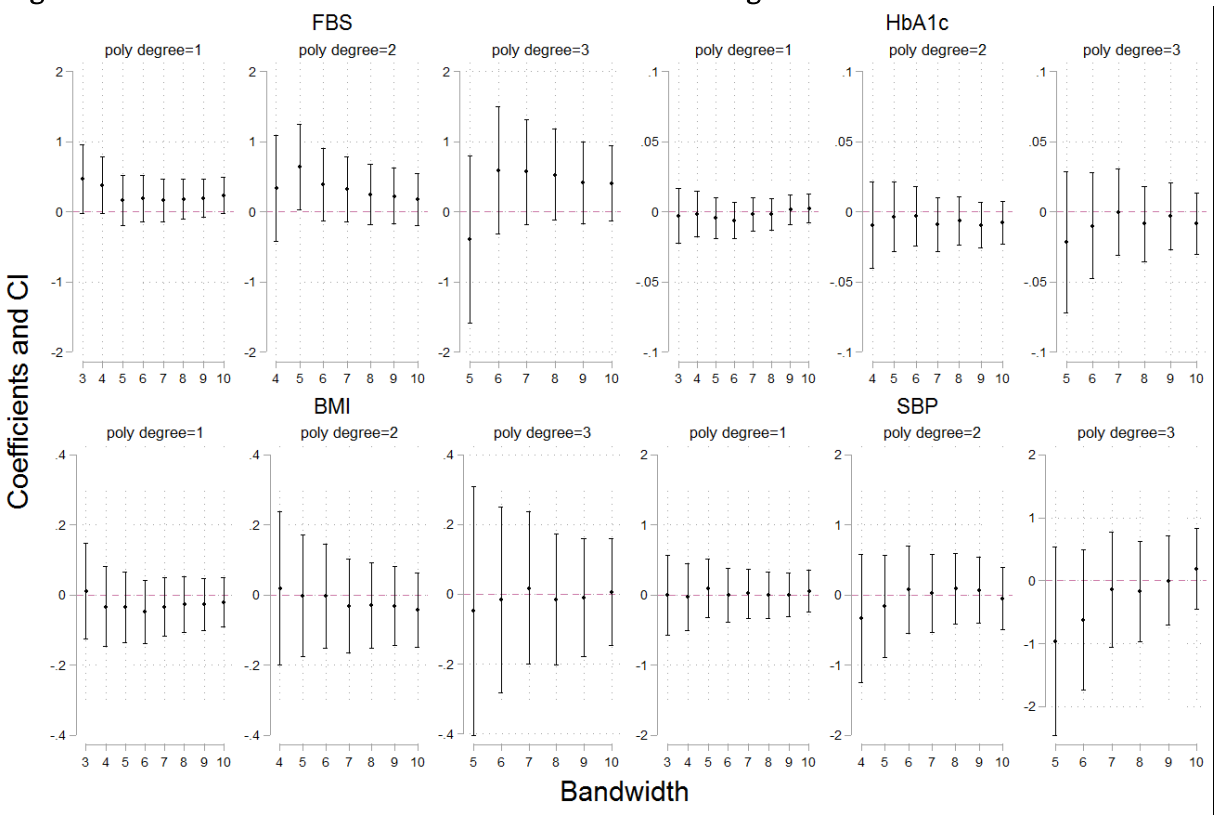
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A4. Robustness of the effects of the FBS≧110 signal on health behavior (estimated without covariates)



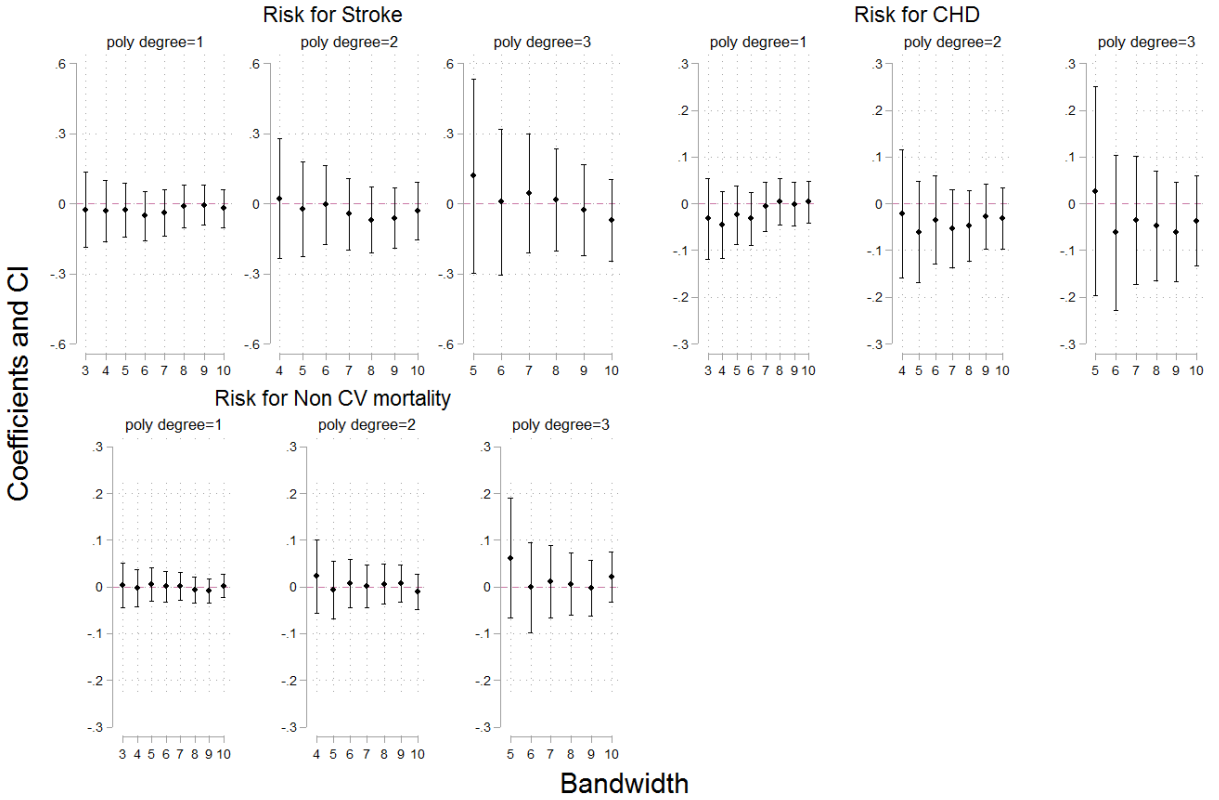
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A5. Robustness of the effects of the FBS≧110 signal on intermediate health outcomes



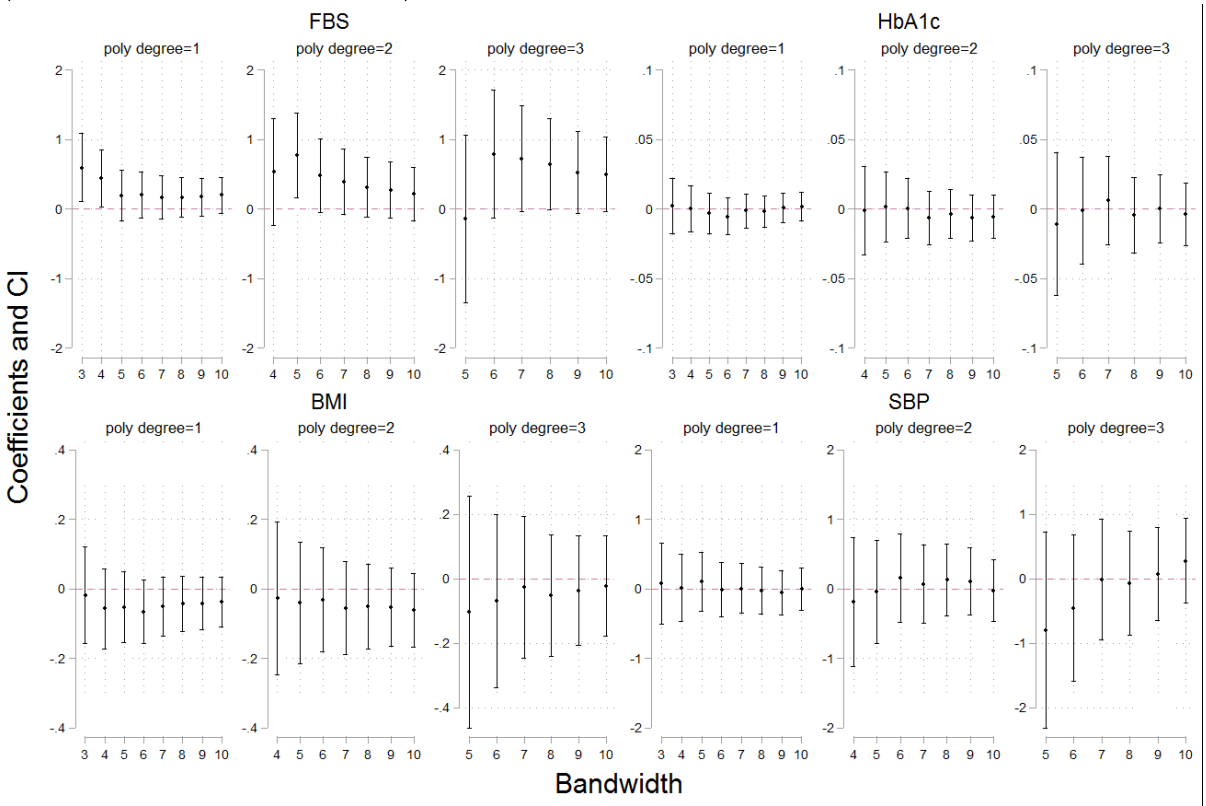
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A6. Robustness of the effects of the FBS≧110 signal on predicted risks of DM complications



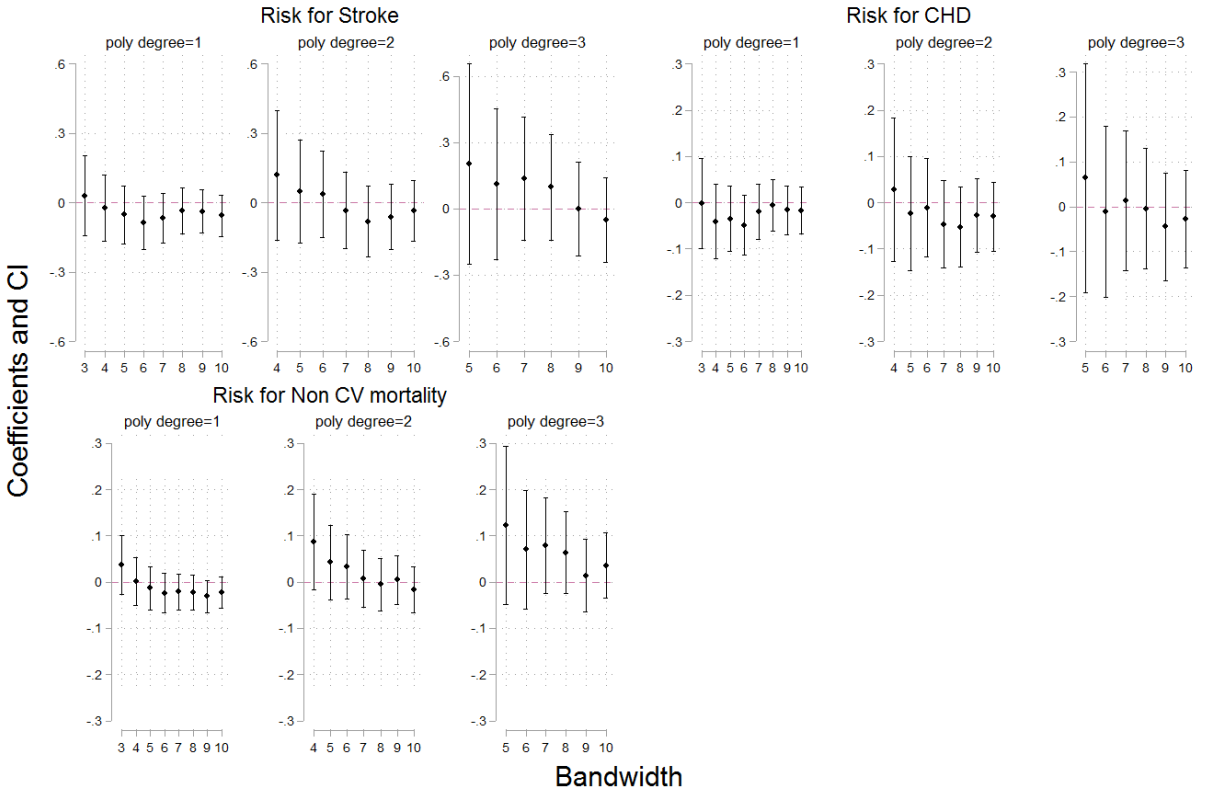
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A7. Robustness of the effects of the FBS≧110 signal on intermediate health outcomes (estimated without covariates)



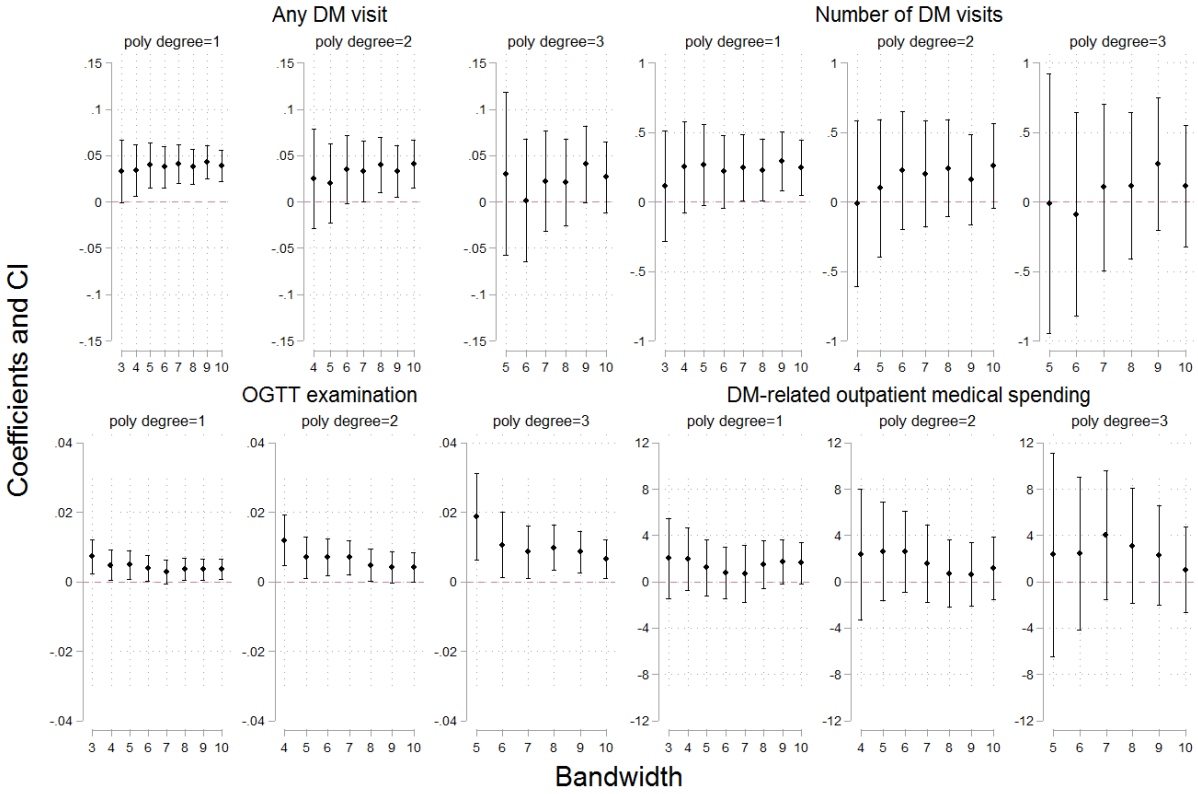
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A8. Robustness of the effects of the FBS≧110 signal on predicted risks of DM complications (estimated without covariates)



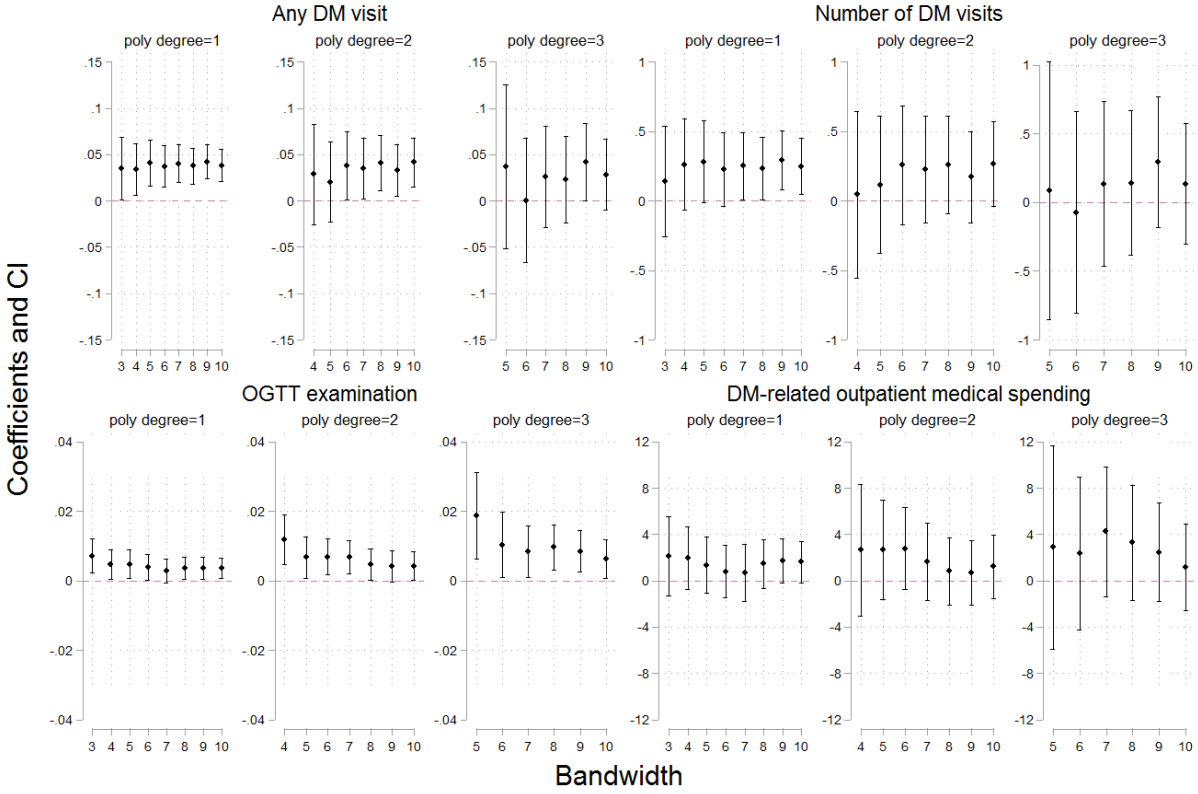
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A9. Robustness of the effects of the FBS≧126 signal on medical care utilization



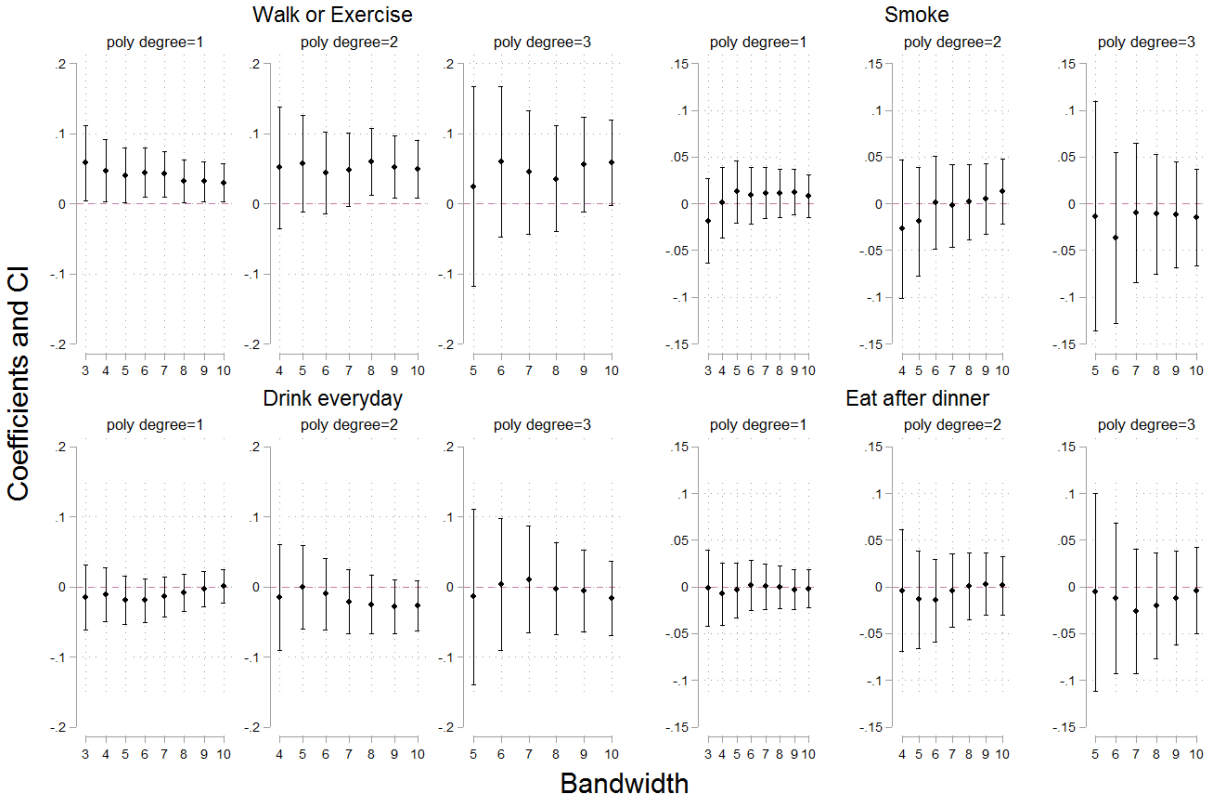
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A10. Robustness of the effects of the FBS≧126 signal on medical care utilization (estimated without covariates)



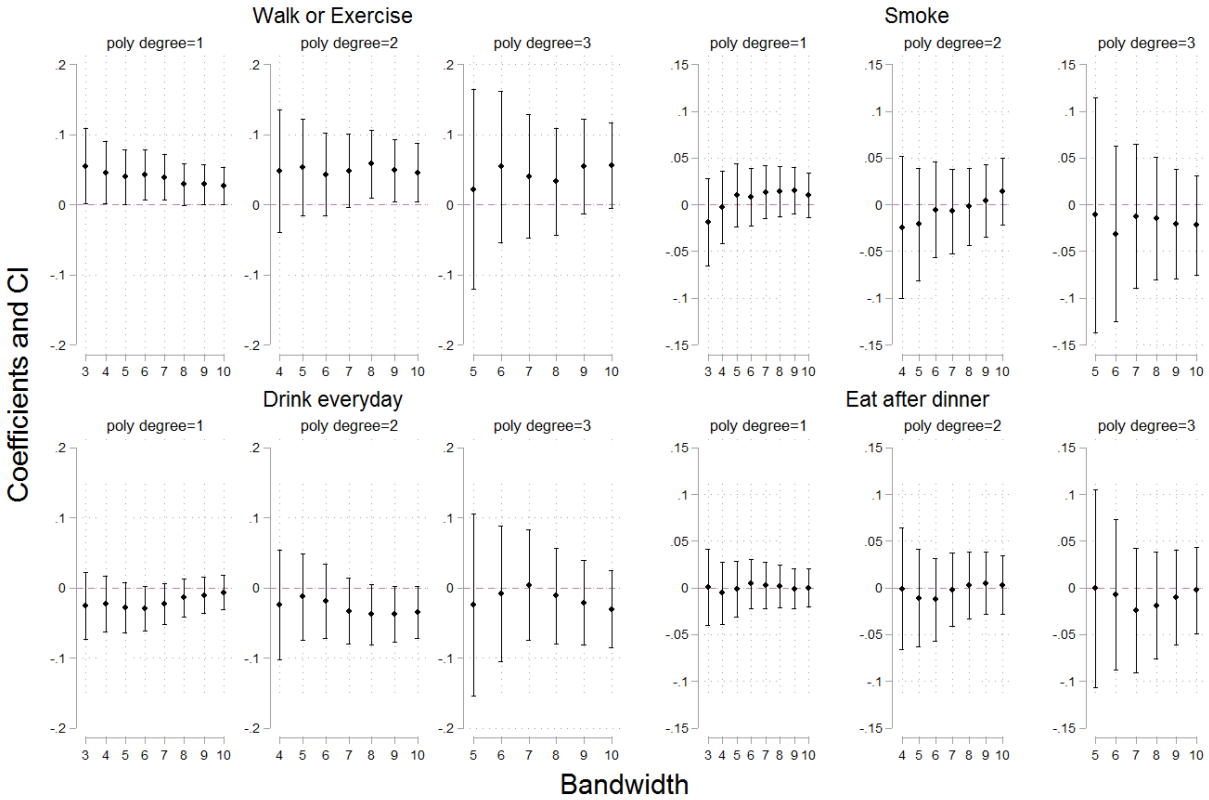
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A11. Robustness of the effects of the FBS≧126 signal on health behavior



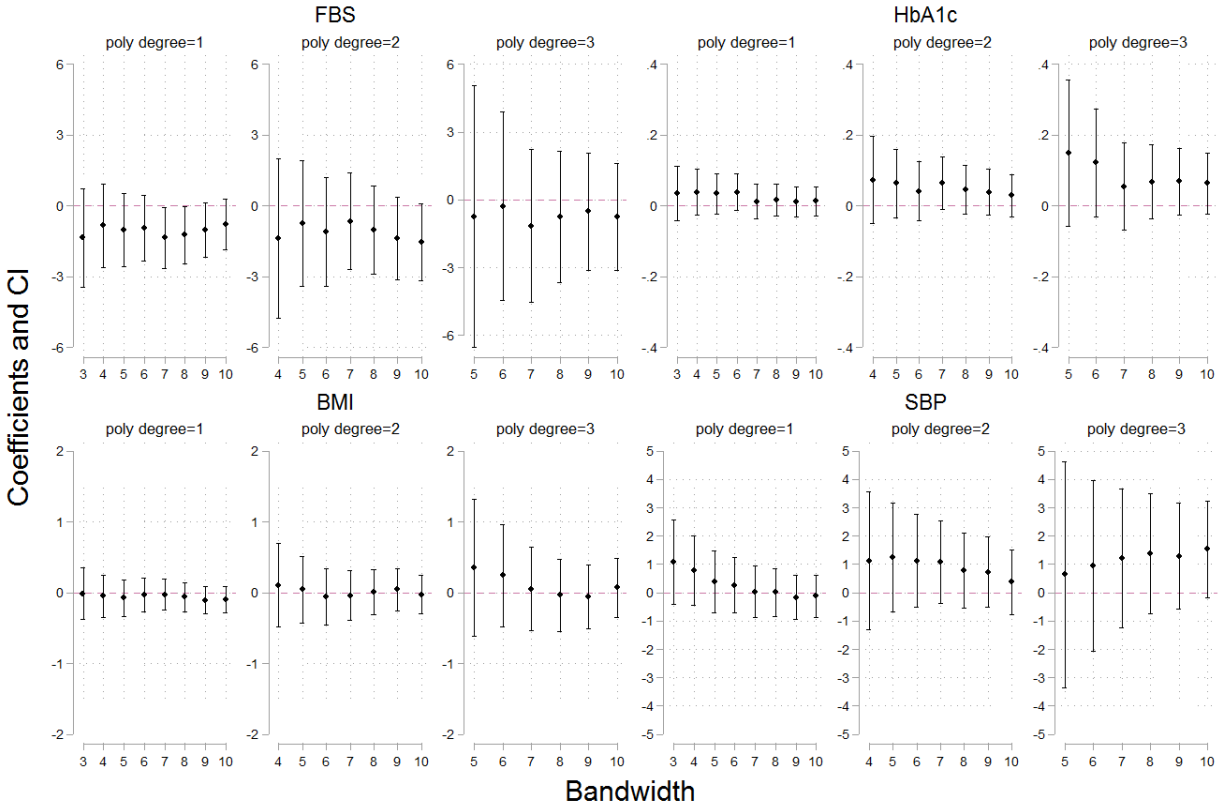
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A12. Robustness of the effects of the FBS≧126 signal on health behavior (estimated without covariates)



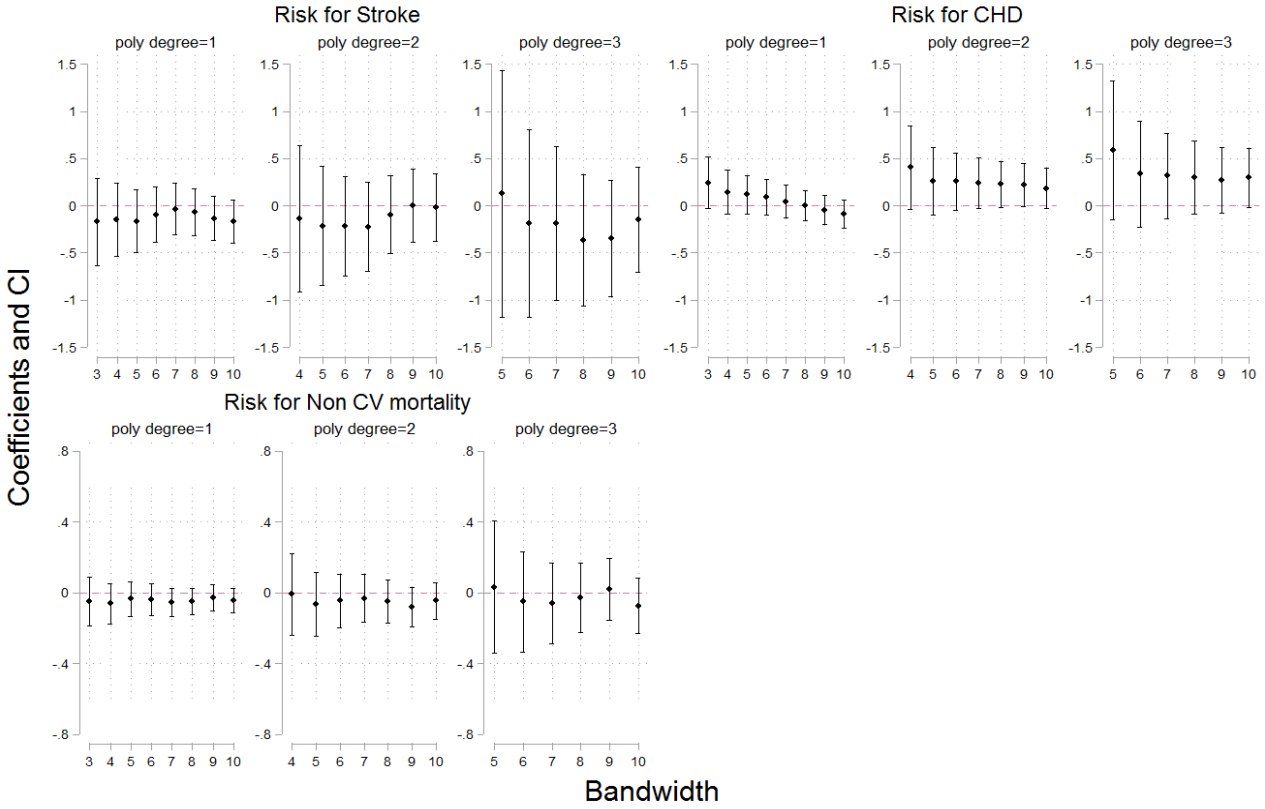
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A13. Robustness of the effects of the FBS≧126 signal on intermediate health outcomes



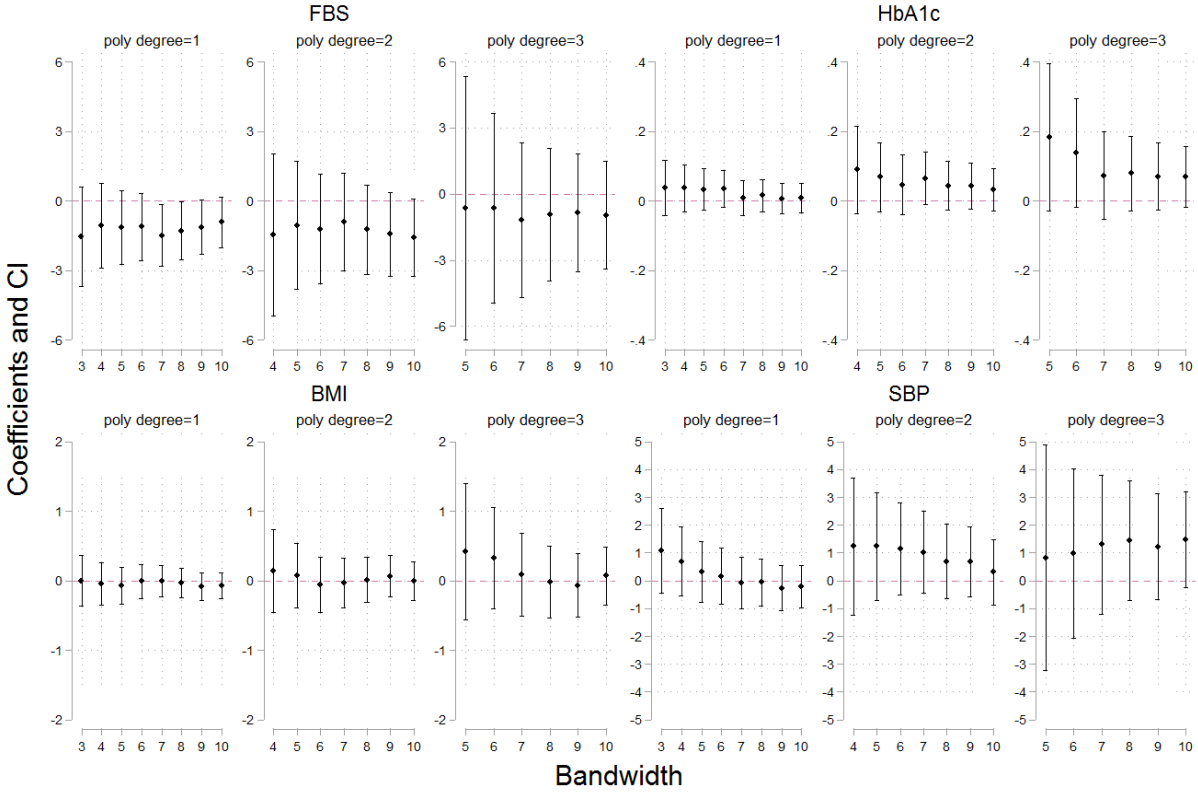
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A14. Robustness of the effects of the FBS≧126 signal on predicted risks of complications



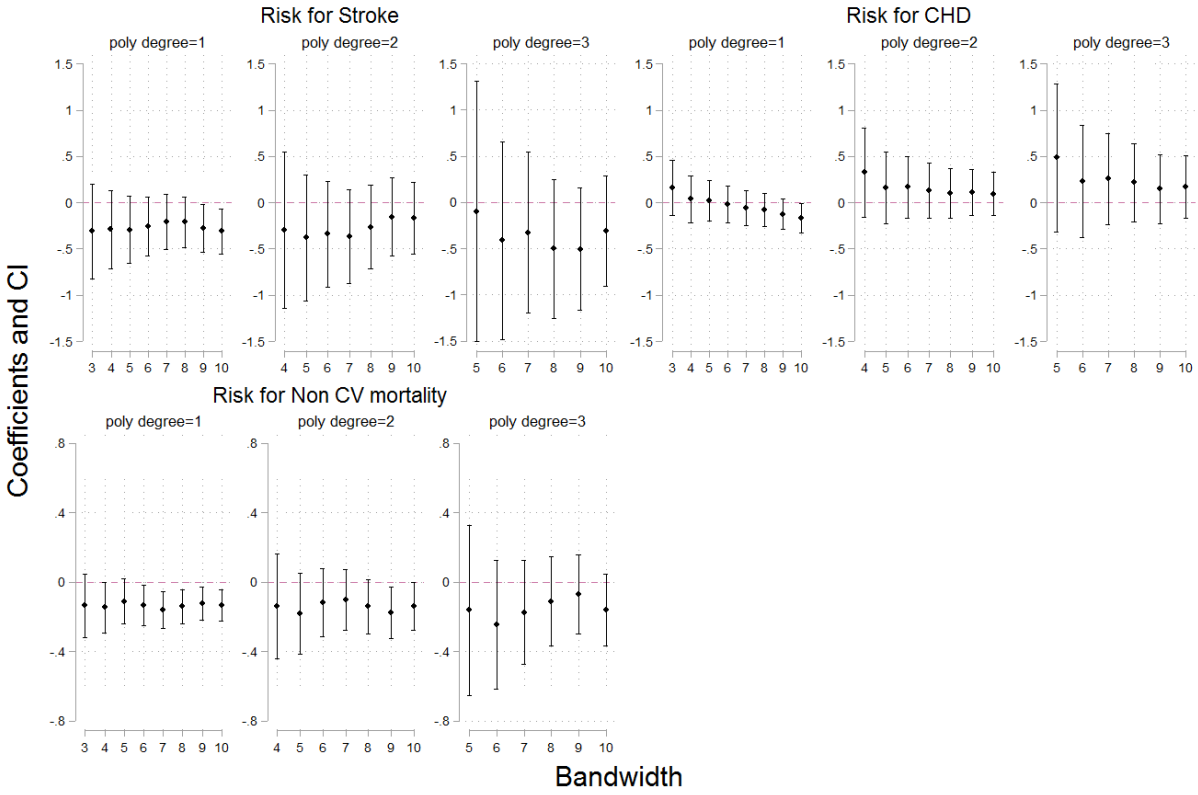
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A15. Robustness of the effects of the FBS≧126 signal on intermediate health outcomes (estimated without covariates)



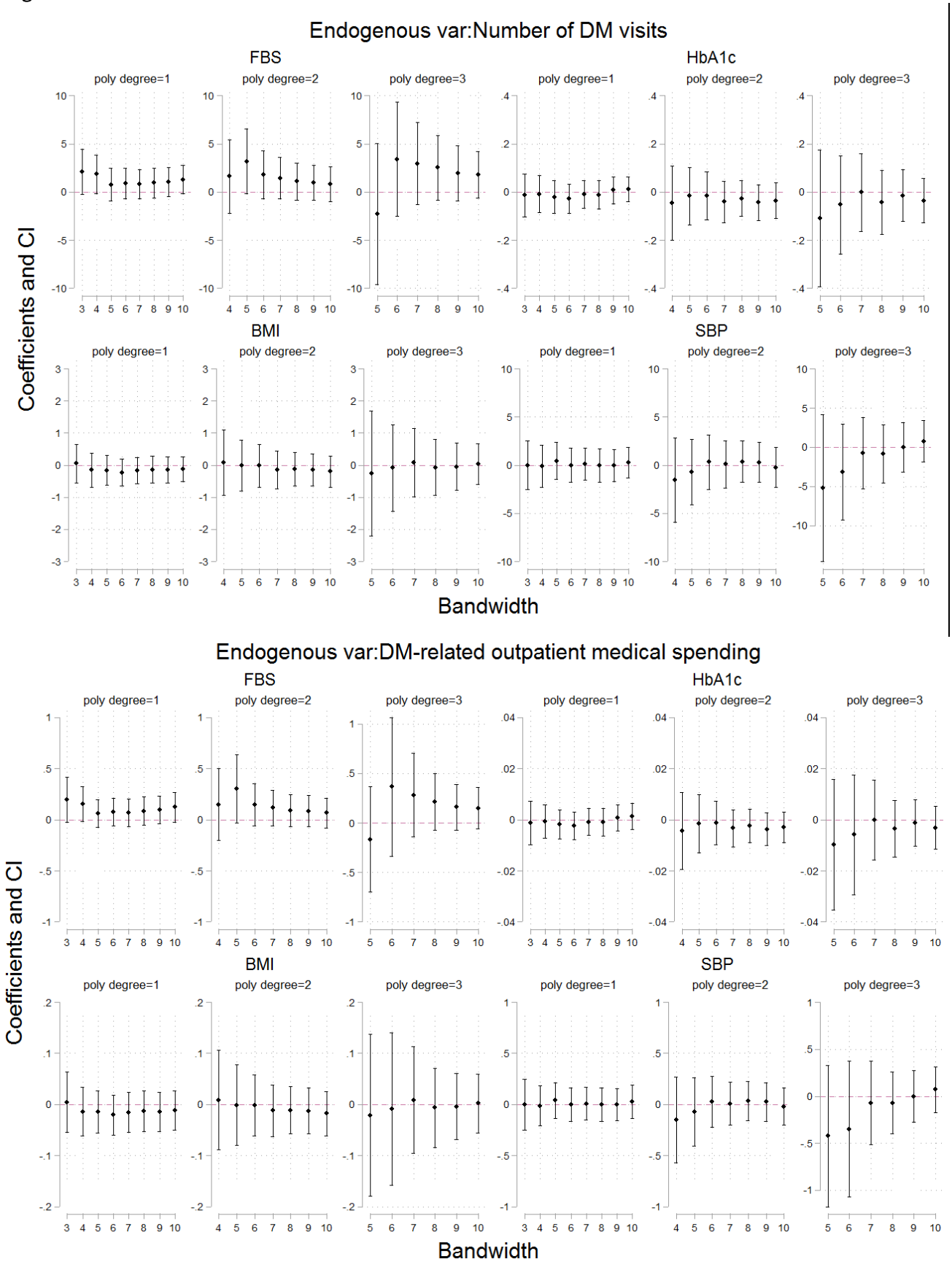
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A16. Robustness of the effects of FBS≧126 signal on predicted risks of DM complications (estimated without covariates)



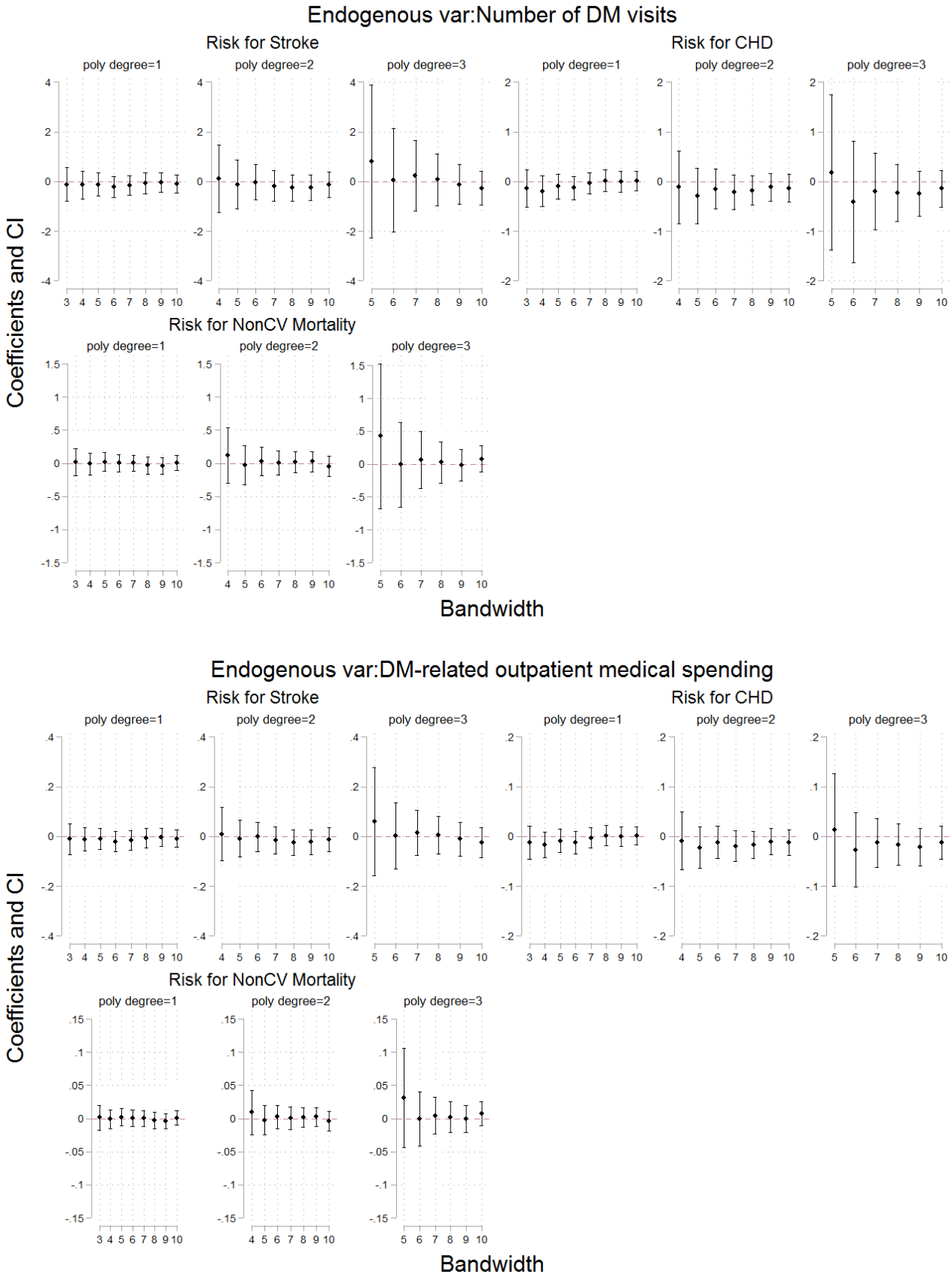
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and without covariates. Only the coefficients and 95% CIs for the FBS ≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A17. Robustness of the effects of additional DM care on intermediate health outcomes



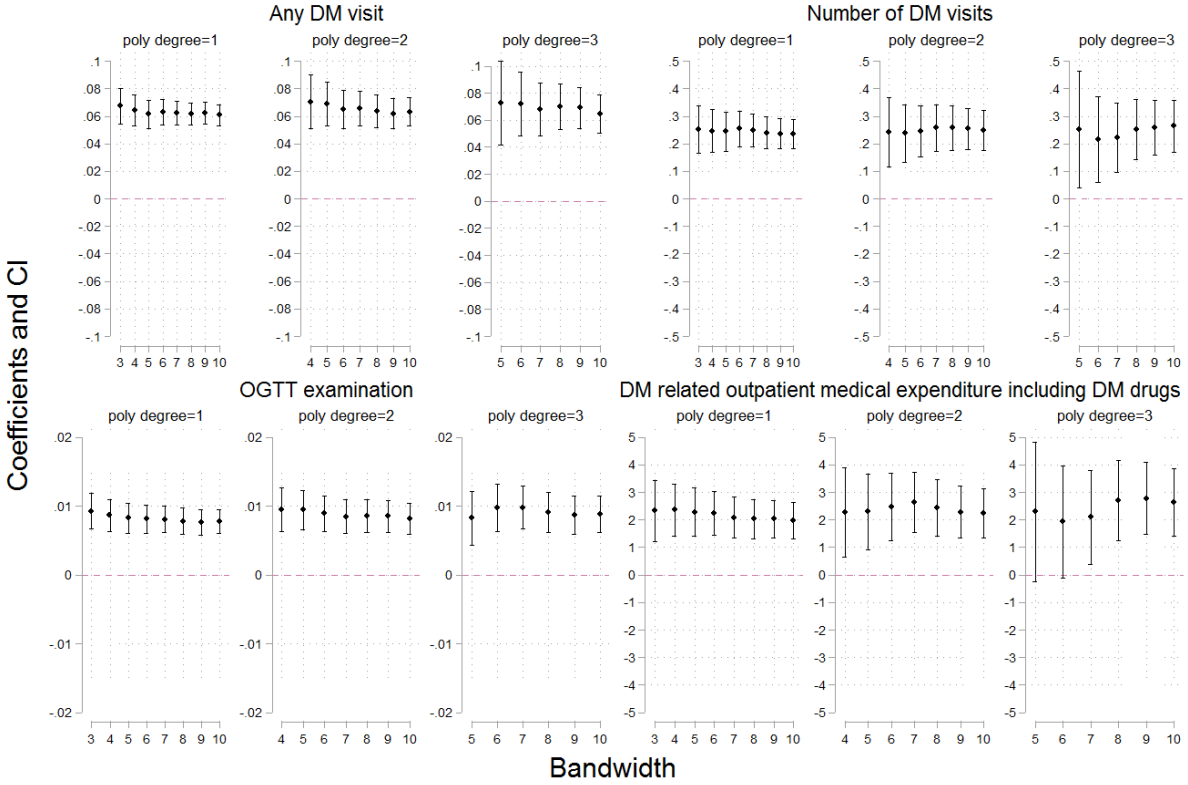
Notes: This figure presents the results from Equation (3). Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level

Figure A18. Robustness of the effects of additional DM care on predicted risks of DM complications



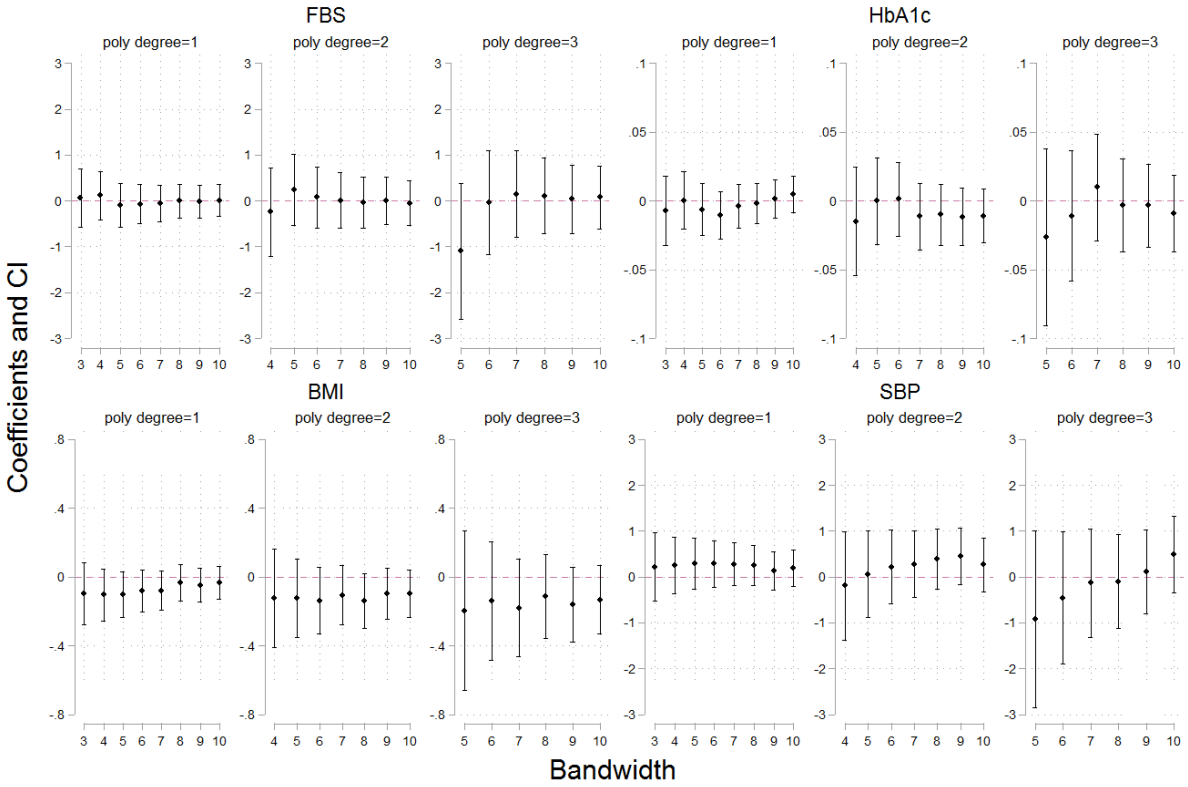
Notes: This figure presents the results from Equation (3). Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A19. Robustness of the effects of the FBS≧110 signal on medical care utilization (for those who did not exceed the threshold in the previous year)



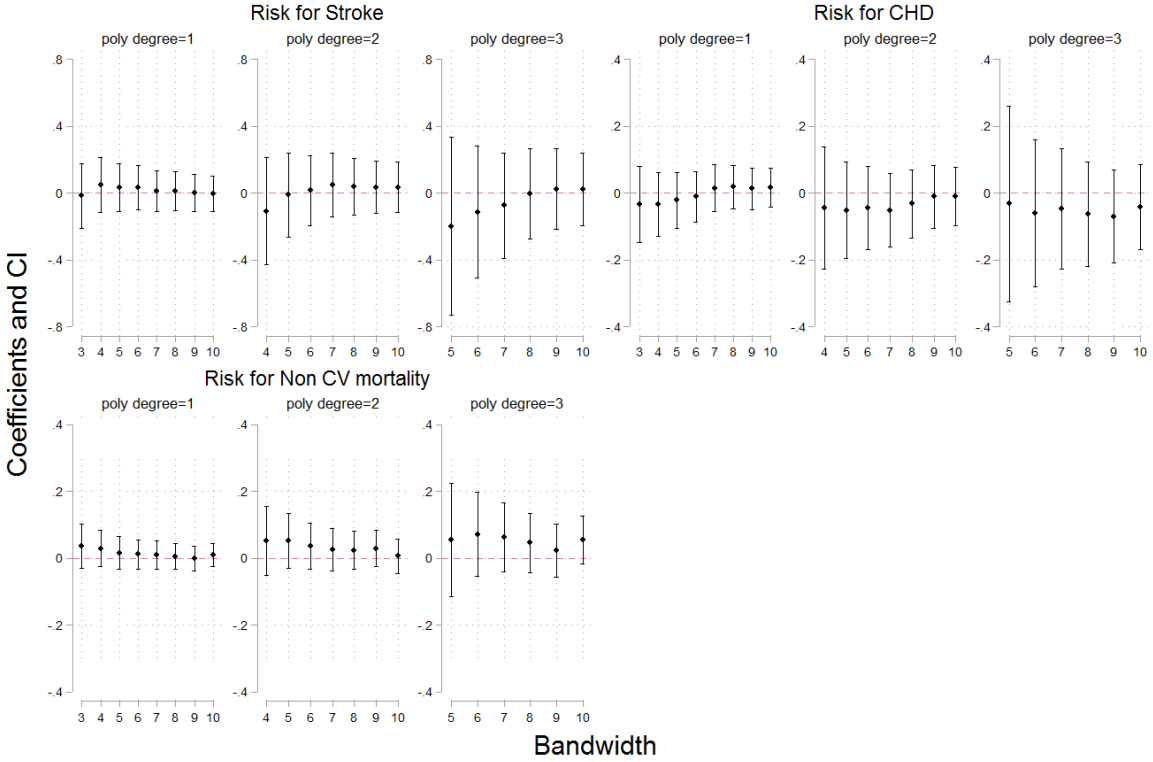
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A20. Robustness of the effects of the FBS≧110 signal on intermediate health outcomes (for those who did not exceed the threshold in the previous year)



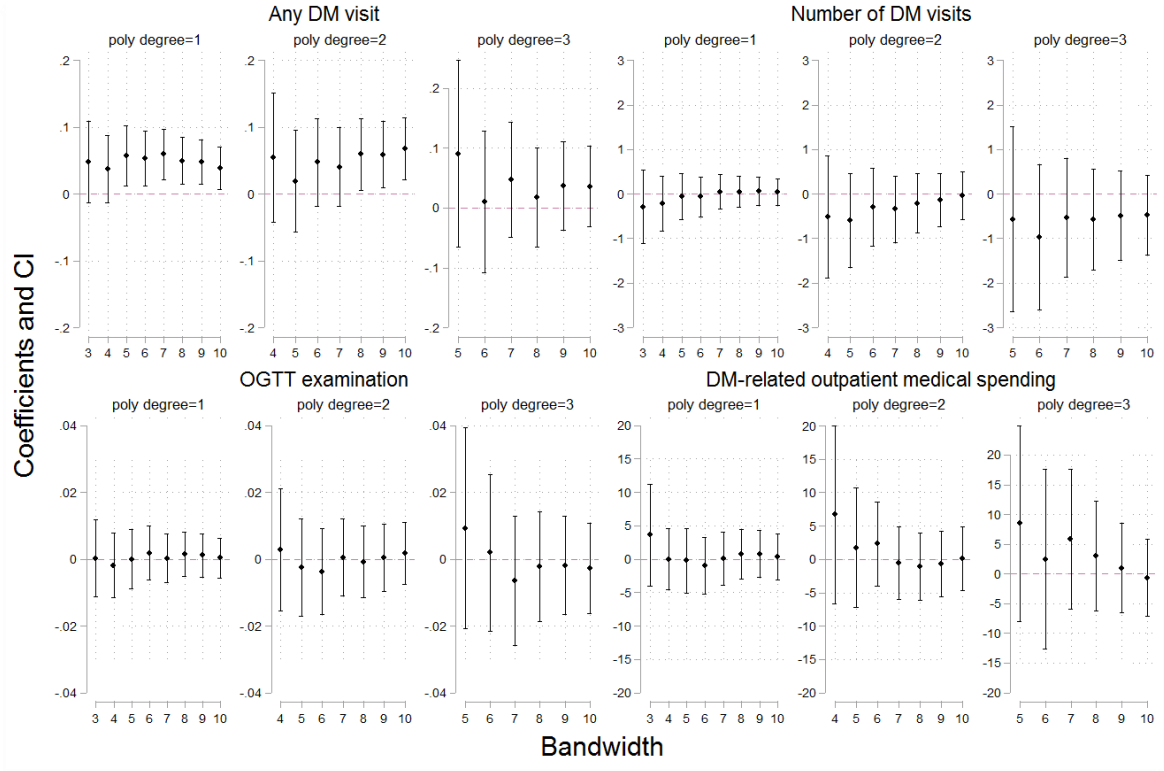
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A21. Robustness of the effects of the FBS≧110 signal on predicted risks of DM complications (for those who did not exceed the threshold in the previous year)



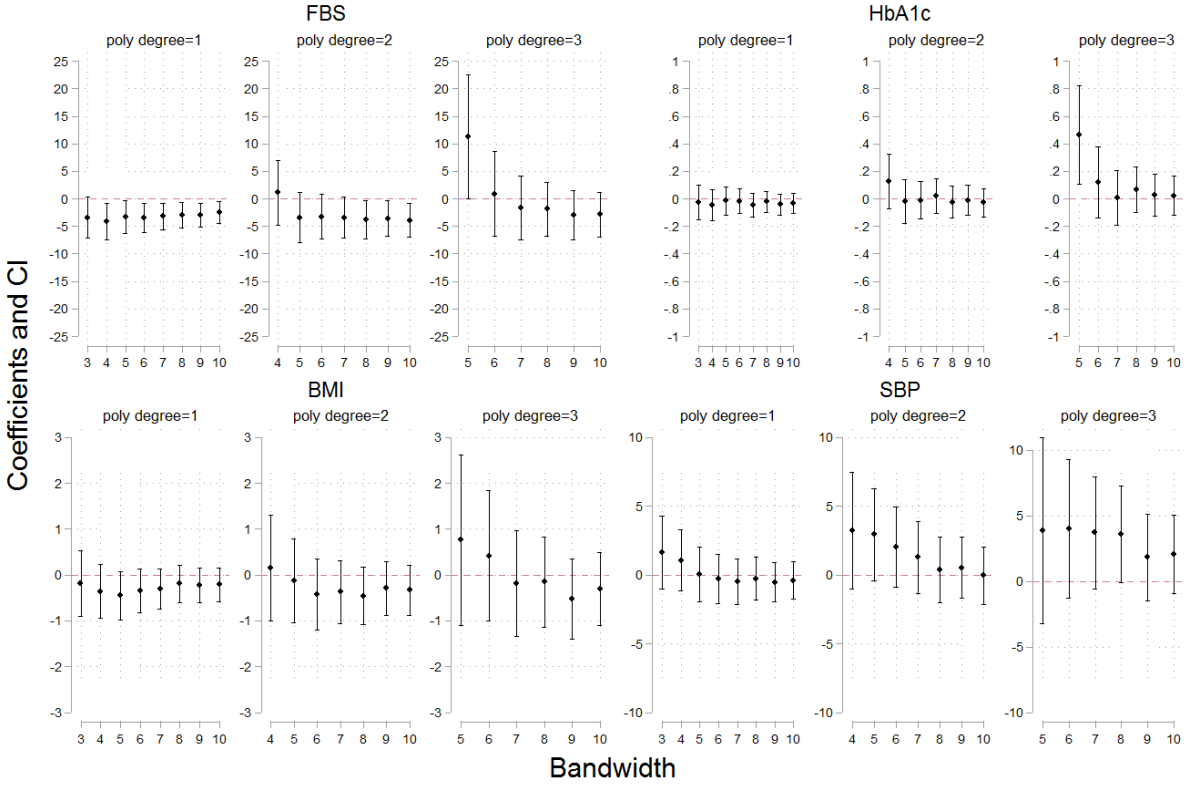
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A22. Robustness of the effects of the FBS≧126 signal on medical care utilization (for those who did not exceed the threshold in the previous year)



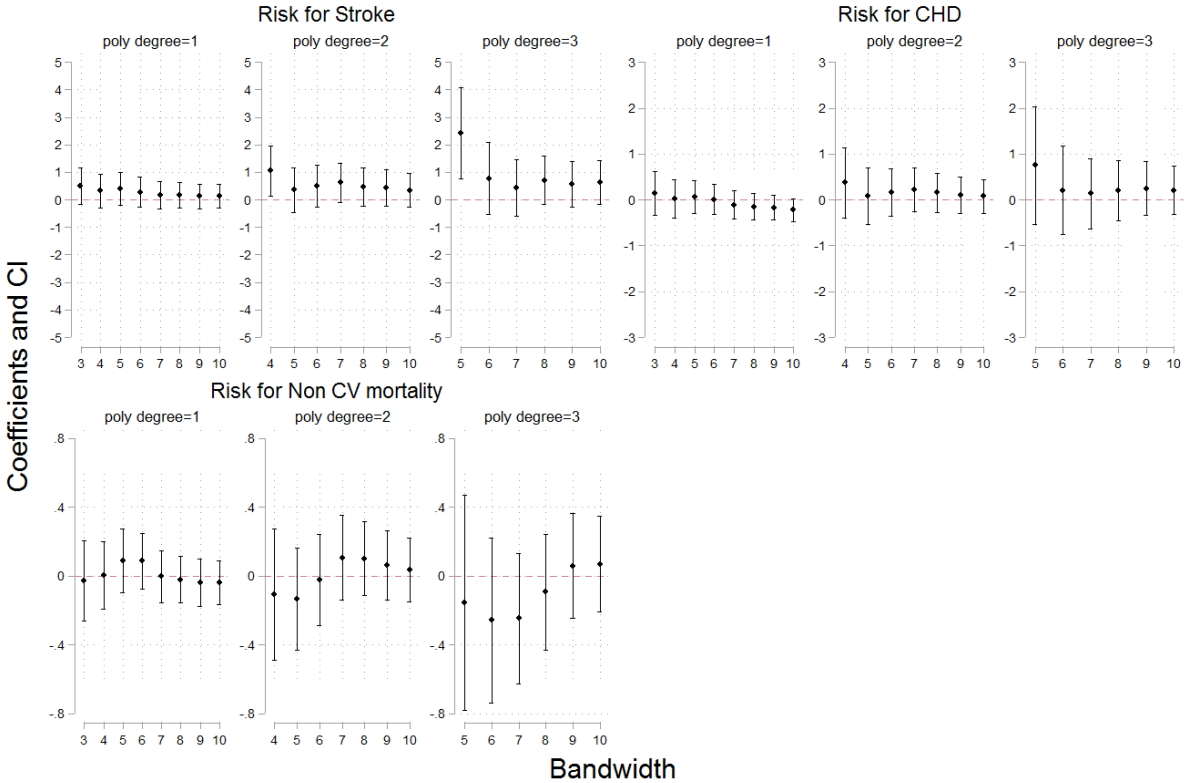
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A23. Robustness of the effects of the FBS≧126 signal on intermediate health outcomes (for those who did not exceed the threshold in the previous year)



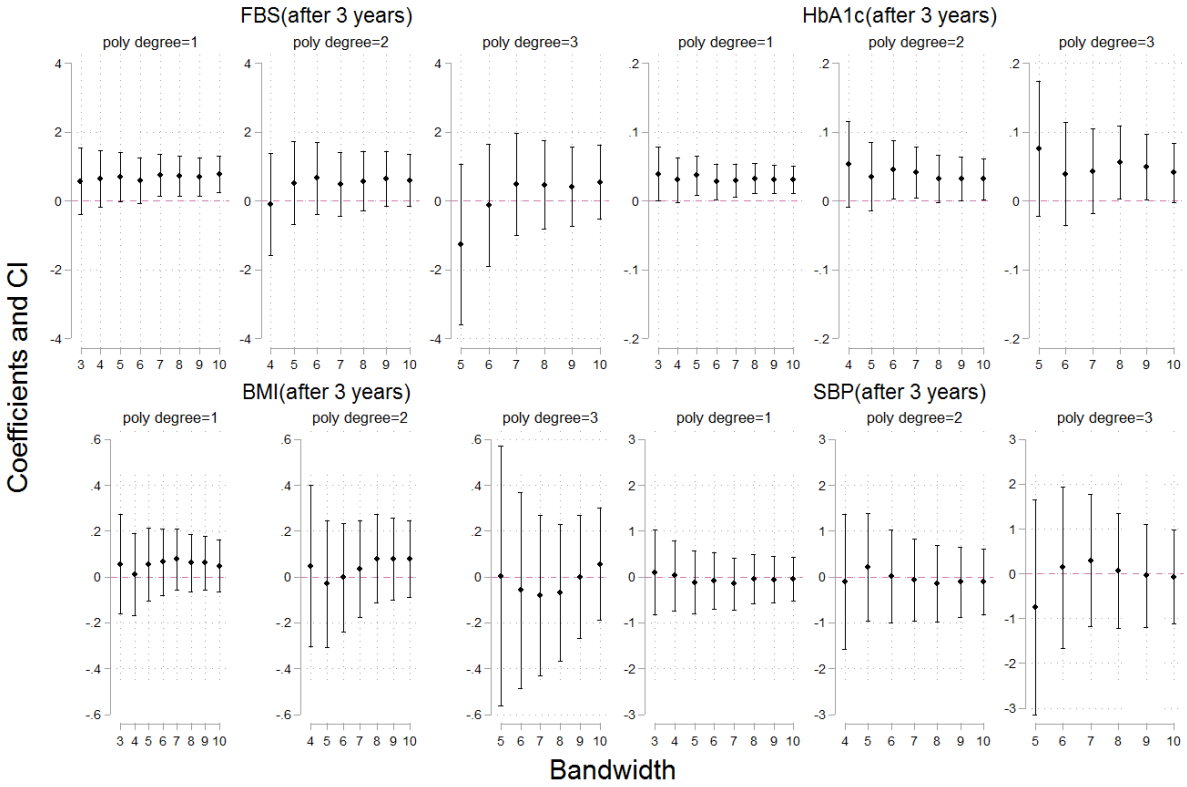
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A24. Robustness of the effects of the FBS≧126 signal on predicted risks of DM complications (for those who did not exceed the threshold in the previous year)



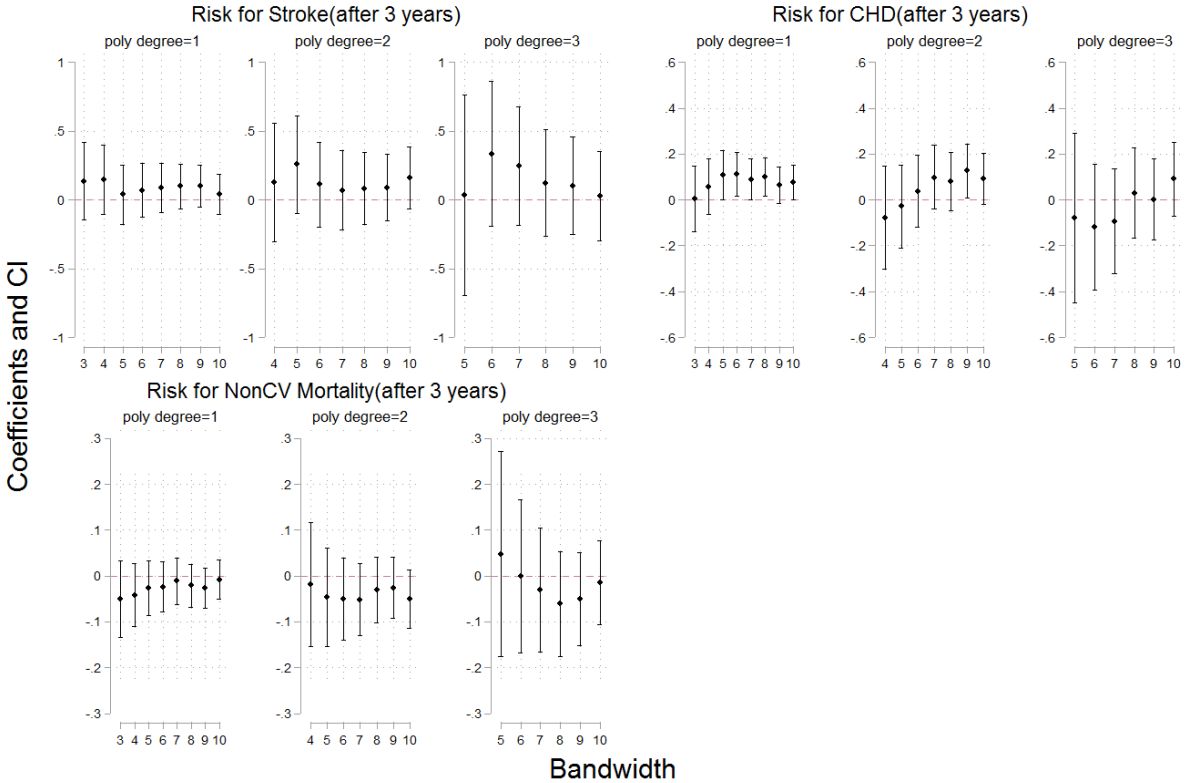
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and covariates. Only the coefficients and 95% CIs for the FBS≧126 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A25. Longer-run effects of the FBS≧110 signal on intermediate health outcomes (after 3 years)



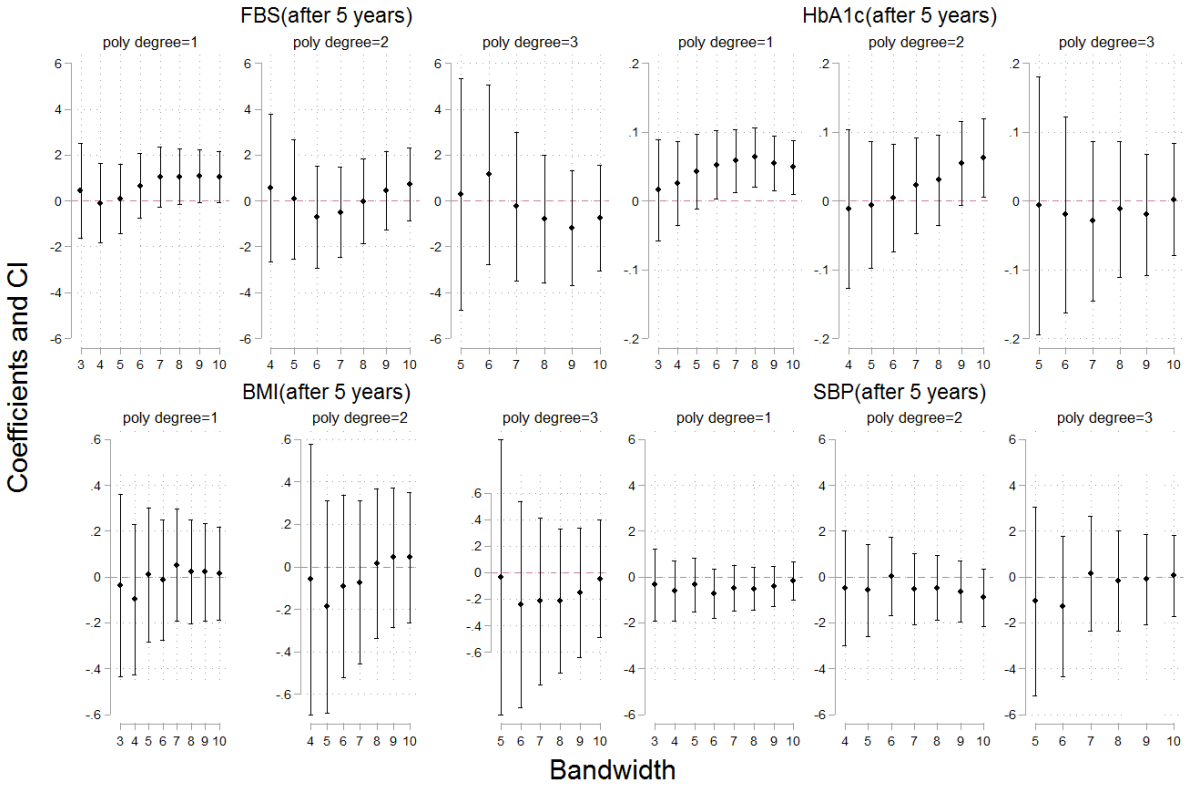
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and with covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A26. Longer-run effects of the FBS≧110 signal on predicted risk of DM complications (after 3 years)



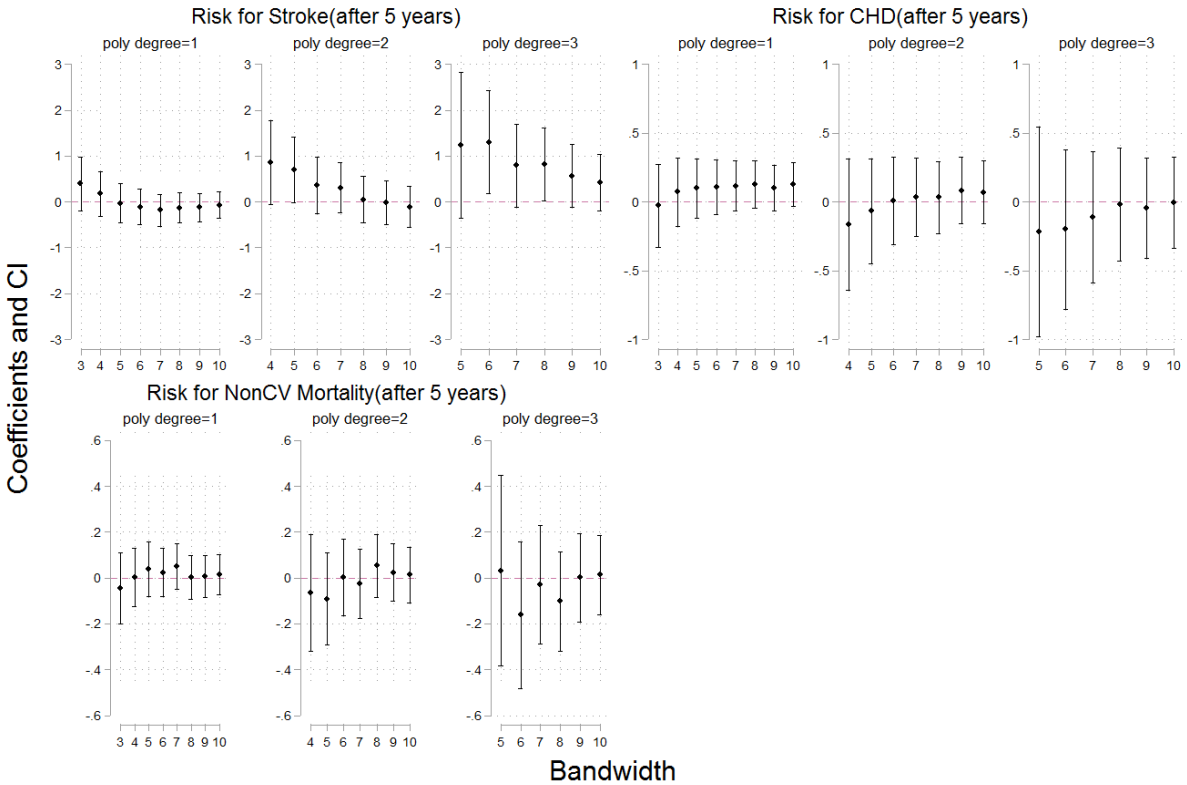
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and with covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A27. Longer-run effects of the FBS≧110 signal on intermediate health outcomes (after 5 years)



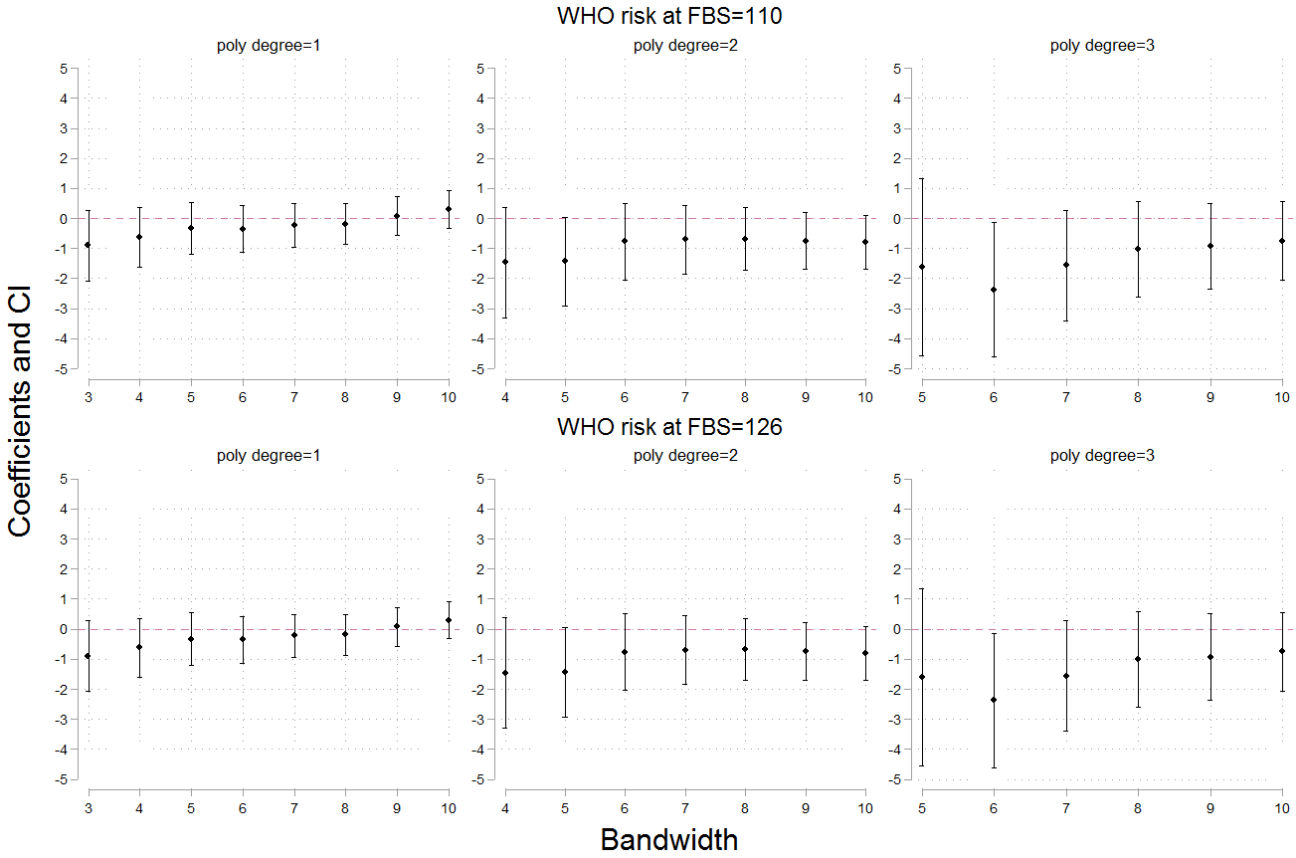
Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and with covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A28. Longer-run effects of FBS≧110 signal on predicted risk of complications (after 5 years)



Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and with covariates. Only the coefficients and 95% CIs for the FBS≧110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Figure A29. Robustness of the effects of FBS signals on an alternative measure of predicted risk of DM complications



Notes: This figure presents the results from local polynomial regressions using rectangular kernel with a bandwidth of 5 mg/dl and with covariates. Only the coefficients and 95% CIs for the FBS ≧ 110 mg/dl threshold are reported. Standard errors, corrected for clustering at the person level.

Table A1. Covariates Balance Test

	FBS at 110 mg/dl			FBS at 126 mg/dl		
	coefficient	mean at 109	obs.	coefficient	mean at 125	obs.
Covariates (at t)						
Age	0.157	48.7	120,735	0.164	49.1	19,241
	(0.097)			(0.252)		
Female	0.009**	0.166	120,735	0.009	0.130	19,241
	(0.005)			(0.010)		
Walk or Exercise	0.015**	0.433	91,769	-0.004	0.425	13,798
	(0.007)			(0.017)		
Smoke	-0.005	0.331	107,684	-0.005	0.397	16,891
	(0.006)			(0.015)		
Drink every day	-0.017***	0.385	100,989	-0.035**	0.390	15,635
	(0.006)			(0.015)		
Eat after dinner	0.001	0.169	88,704	-0.008	0.170	13,254
	(0.005)			(0.013)		
HbA1c	0.002	5.7	102,383	0.051***	6.0	17,653
	(0.005)			(0.019)		
BMI	0.022	24.3	120,673	-0.005	25.1	19,208
	(0.044)			(0.120)		
SBP	-0.060	128	120,650	0.699	130	19,210
	(0.185)			(0.493)		
Risk for Stroke	0.079	3.483	74,745	0.031	4.550	11,984
	(0.048)			(0.127)		
Risk for CHD	0.001	2.730	74,745	0.093	3.710	11,984
	(0.026)			(0.088)		
Risk for Non-CV mortality	-0.002	1.610	74,745	0.033	1.990	11,984
	(0.015)			(0.039)		

Note: This table shows the results from local-linear regressions using a rectangular kernel with a 5 mg/dl bandwidth and covariates. Only the coefficients for the RD dummies are reported. Standard errors, corrected for clustering at the person level, are in parentheses. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.