NBER WORKING PAPER SERIES

FALSE ALARM? ESTIMATING THE MARGINAL VALUE OF HEALTH SIGNALS

Toshiaki Iizuka
Katsuhiko Nishiyama
Brian Chen
Karen Eggleston

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2017, Revised January 2020

This paper is a substantially revised version of the paper previously circulated as "Is Preventive Care Worth the Cost? Evidence from Mandatory Checkups in Japan." We are grateful to Philip Clarke, Martin Gaynor, Chie Hanaoka, Daniel Herrera, Hyuncheol Kim, Ming-Jen Lin, Nada Wasi, and seminar and conference participants at Stanford University, University of Southern California, Tokyo University of Science, AEA meetings, AWEHE, EuHEA Conference, iHEA Congress, IRDES-Dauphine Health Economics Workshop, JHEA Conference, and the Korea-Taiwan-Japan Health Economics Associations Joint Conference for helpful comments and suggestions. The Kikawada Foundation, Abe Fellowship Program, and JSPS Bilateral Open Partnership Joint Research Projects are gratefully acknowledged for their financial support. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

False Alarm? Estimating the Marginal Value of Health Signals
Toshiaki Iizuka, Katsuhiko Nishiyama, Brian Chen, and Karen Eggleston
NBER Working Paper No. 23413
May 2017, Revised January 2020
JEL No. I10

ABSTRACT

We investigate the marginal value of information in the context of health signals after checkups. Although underlying health status is similar for individuals just below and above a clinical threshold, treatments differ according to the checkup signals they receive. For the general population, whereas health warnings about diabetes increase healthcare utilization, health outcomes do not improve at the threshold. However, among high-risk individuals, outcomes do improve, and improved health is worth its cost. These results indicate that the marginal value of health information depends on setting appropriate thresholds for health warnings and targeting individuals most likely to benefit from follow-up medical care.

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

Katsuhiko Nishiyama
University of North Carolina
at Chapel Hill.
katsu.nishiyama@gmail.com

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

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

An online appendix is available at http://www.nber.org/data-appendix/w23413

1. Introduction

In recent years, it has become far easier to obtain information about one's own health. In addition to traditional channels, including health checkups, workplace wellness programs, and cancer screening, wearable and portable devices are gaining popularity, allowing people to monitor their health more easily. Large pharmacy chains such as Walgreens and CVS also offer health screening in many retail locations. Advocates suggest that such health information will lead to appropriate preventive care, improve health outcomes, and may ultimately reduce medical spending. For example, the US Affordable Care Act provides incentives for wellness programs, which include free health screening, on the presumption that preventive care improves health and wellbeing. However, others are skeptical about such effects[1] and evidence is mixed at best (Baicker, Cutler and Song 2010; Jones, Molitor, and Reif 2019; Song and Baicker 2019). In fact, even if health information encourages visits to physicians, such visits may result in wasteful resource use if signals increase low-value care.

The aim of this paper is to investigate the marginal value of information in the context of health signals that people receive after routine checkups, focusing on risk for diabetes mellitus (DM). We first look at whether checkup results indicating risk of developing DM affect individuals' medical care utilization, health behaviors, and health outcomes. We then examine whether the additional care triggered by exceeding a clinical threshold is worth its cost, by comparing the additional medical spending to the value of any resulting improvement in health outcomes up to five years after the check-up.

We apply a regression discontinuity (RD) design to assess the incremental value of receiving a signal of being in poor health. We exploit the fact that individuals with health

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

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. Because the clinical thresholds in Japan were largely adopted from international standards based on western populations and have changed over time (as discussed below), the specific threshold within a range of FBS is arbitrary. However, people with measured values just above the threshold, indicating poor health, may receive more preventive care – such as further diagnostic tests and diabetes-related physician visits – compared to those with values just below the threshold. This additional care may lead to better health outcomes for the individuals just above the threshold, compared to those just below the threshold. By comparing the cost of care and health outcomes of these people, we can assess the efficiency of marginal 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.

We use Japanese data for this study, which 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, the mandatory health checkups in Japan alleviate the concern about self-selection intro screening, because more than 95% of employees in large corporations, such as those in our sample, receive annual checkups.[2] 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 (e.g. Jones, Molitor, and Reif 2019; Myerson

[2] Special Survey on Industrial Safety and Health (MHLW 2012).

et al. 2018). Third, we have outcome measures suitable for examining the health and survival impacts of prevention. We use biomarkers from checkups as well as apply a population-specific risk prediction model to our data to predict the 5-year risk of mortality for each individual. These measures allow us to study whether additional preventive care triggered by a checkup leads to better health outcomes. Moreover, by assuming a value of a statistical life year, we can quantify the monetary value of any improvement in survival. This is an advantage compared to only examining intermediate health measures such as FBS that are more easily available but miss the non-linear interaction of several risk factors (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. DM is 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.[3]

Unlike genetic disease, where information avoidance may be optimal (Oster, Shoulson, and Dorsey 2013), DM is a chronic condition that can be prevented in the "pre-diabetes" stage by early intervention to reduce lifestyle risk factors (such as smoking, unhealthy diet, sedentary lifestyle, and obesity). Once diagnosed, DM can be managed with therapies that combine lifestyle change with medication to minimize likelihood of serious complications and improve survival. DM and pre-diabetes can be detected by elevated blood sugar levels (i.e., as measured by FBS), a

[3] National Health and Nutrition Examination Survey (MHLW 2015) http://www.mhlw.go.jp/bunya/kenkou/eiyou/dl/h27-houkoku.pdf (P.165).

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. Since health checkups are designed to catch disease early enough for preventive measures to be effective, our empirical strategy and results focus on individuals on the margin of clinical thresholds for pre-diabetes and DM; these "marginal" individuals constitute more than half of our sample of over 600,000 working-age adults.

We have three main findings. First, at a relatively low diagnosis threshold that corresponds to pre-diabetes in Japan (also called "borderline type"), we find strong evidence that surpassing the threshold increases medical care utilization as measured by DM-related physician visits and DM-related outpatient expenditures, including spending on medications. This finding indicates that some 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 pre-diabetes threshold, we find no evidence that the additional care improves health outcomes as measured by a variety of biomarkers such as FBS, body mass index (BMI), and blood pressure. Thus, 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 the threshold may need to be reexamined from the perspective of efficiency.

Third, at a higher diagnostic threshold above which the person is classified as a diabetic type we continue to find no evidence for the general population that crossing the threshold improves health outcomes. However, if we specifically focus on high-risk individuals with elevated blood pressure and high cholesterol, we find evidence that the exceeding the diabetic type threshold

improves health outcomes, namely HbA1c and FBS levels. This result suggests that targeting high-risk individuals after a checkup is more likely to be a fruitful approach. In fact, a further analysis that utilizes a risk prediction model indicates that the cost-per-life saved due to reduced blood sugar at this margin is comparable to the conventional estimate of the value of a statistical life.

These results indicate that health signals can improve welfare but only if thresholds are set appropriately and follow-up medical care targets those at highest risk. Given that relatively few respond to the checkup results, the benefit of checkups could also be increased if, for example, insights from behavioral science were used to enhance the clarity and salience of check-up results for high risks, and/or follow-up monitoring was provided for those at highest predicted risk. More generally, our results suggest that consumer response to information and seeking of expert advice is not necessarily efficient, and the burden of proof is on showing that policies and programs that emphasize preventative care can improve health outcomes cost effectively.

This paper is related to the strand of literature that examines the value of information. Dranove et al. (2003), for example, examine the impacts of cardiac healthcare report cards introduced in New York and Pennsylvania and find that, in contrast to common belief, more information provided by report cards reduced welfare, especially for the sick. Kolstad (2013) finds that surgeons improved quality after report cards were introduced, focusing on the marginal value of information on the supply side when suppliers are intrinsically motivated. Jin and Leslie (2003) study the impact of restaurant hygiene report cards in California and find that restaurant hygiene quality improved after its introduction.[4] Like Handel and Kolstad (2017), the focus of

[4] Dranove and Jin (2010) summarize the literature on information about quality, finding mixed evidence about the welfare implications of consumer and provider response to quality disclosure in healthcare.

our study is individual health information, access to which has become dramatically easier in recent years.

Second, our study contributes to the literature on marginal returns to medical care. As previously indicated, a notable pioneer in using clinical thresholds to further this line of research is Almond et al. (2010). Focusing on the "very low birth weight" threshold for newborns, they find that those whose birth weights are just below the threshold receive more medical care and experience lower one-year mortality rates, compared to newborns with birth weight just above the threshold. These discontinuities allow them to conclude that medical care for at-risk newborns is cost effective around the threshold. We contribute to this literature by providing empirical evidence on the value of preventive care triggered by health checkup results. While many governments and organizations emphasize the importance of preventive care, not all such care can improve welfare. False alarms may lead to low-value diagnostic procedures, some of which may even increase risk of complications.

Third, this study is related to the literature that examines behavioral responses to health condition diagnosis (Zhao, Konishi and Glewwe 2013; Oster 2018; Jones, Molitor, and Reif 2019; Kim, Lee and Lim 2019; Alalouf, Miller and Wherry 2019).[5] The most closely related study is that of Kim, Lee and Lim (2019), who examine screening for diabetes, obesity, and hyperlipidemia under the National Health Screening Program in Korea.[6] Our study differs from theirs in important ways. First, we study the increase in medical expenditures triggered by health checkups and the value of lives extended due to crossing the threshold. Kim, Lee and Lim (2019)

[5] For discussion of the cost-effectiveness literature, see Cohen, Neumann and Weinstein (2008).

[6] Among other related studies, Oster (2018) 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. She suggests that relatively modest "sin taxes" (e.g. on sugary sodas) or subsidies of healthy foods might be more effective than individual health signals. Jones, Molitor, and Reif (2019) provide evidence that workplace wellness programs often engender strong selection but fail to spur behavioral change or improve health outcomes.

do not study medical spending as an outcome nor whether the additional utilization induced by moving above the threshold is cost effective (the effect of moving above the threshold on spending divided by the effect of moving above the threshold on health).[7] Second, 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 impact estimation results (Jones, Molitor, and Reif 2019).[8] Third, our sample has a much larger number of observations near the threshold,[9] which might explain the differences in results; for example, while we find clear increases in utilization at the pre-diabetes threshold, there was no significant effect in the Korean sample.[10]

This paper also contributes to the debate over the costs and benefits of health screening initiatives. Recently, the value of annual physicals has received renewed attention and our study provides empirical evidence on this debate (e.g., Mehrotra and Prochzka, 2015; Goroll, 2015; Rubin, 2019). Also, our results speak to the debate over the cost effectiveness of wellness programs, which has become an $8 billion industry in the United States (Baicker, Cutler and Song 2010; Jones, Molitor, and Reif 2019; Song and Baicker 2019).

The remainder of the paper is organized as follows. In Section 2, we briefly discuss 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 Section 5 we report our graphical and econometric results and additional analyses,

---

[7] Indeed, they suggest that screening information should be combined with more medical intervention, whereas we evaluate that suggestion and show that the induced medical spending is only efficient among high-risk individuals (with three health measures out of the normal range).

[8] Screening participants in Korea are more likely to have employee insurance, higher income, and lower medical expenditure, compared to those eligible who did not participate in screening (Kim, Lee and Lim, p.5).

[9] We have about five times more observations around the pre-diabetes threshold than the Korean sample.

[10] Appendix III provides a more detailed comparison of our short- and long-term results (for both the general population and for high risk individuals) with the results of Kim et al. (2019) for Korea, as well as some discussion of Alalouf et al.'s (2019) application of similar methods to US data.

including the long-run effects of preventive care and various robustness checks. Section 6 concludes our paper, pointing out the general applicability of our approach to measuring cost-effectiveness of care around clinical thresholds.

2. Mandatory Health Checkups and Management of DM in Japan

Mandatory checkup. In Japan, 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$4,500) 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 employer disciplinary actions for non-compliance. As noted above, more than 95% of employees in large corporations do comply and receive annual checkups.

Checkup report. Within a month or two after a checkup, individuals are mailed a report of the index checkup results, including measured values of height, weight, liver function, blood lipids, blood pressure, and blood sugar, among others. 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.

Thresholds for DM diagnosis. FBS, or sugar 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. The Japan Diabetes Society (JDS) 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 pre-diabetes or borderline type. Individuals classified as pre-diabetes have a higher rate of developing DM (Seino et al. 2010). An FBS value below 110 mg/dl is "normal type." These JDS-specified FBS threshold values for DM diagnosis and elevated risk of DM have remained unchanged since before our study period (Haneda et al. 2018), although the exact values are somewhat arbitrary, having been adopted from western standards and changed before our study period (see discussion below).

Physician office visits. After receiving a health warning, such as "H", the decision whether to visit a physician is entirely up to the individual. Employers are neither mandated nor have the legal right to make an employee visit a physician. If the individual chooses to visit a physician, the physician may learn about the checkup results if the individual shares those results with the physician; Japan does not have an electronic health record network where physicians can check previous medical records including checkup reports. How to treat the patient is also up to the physician, although JDS has some treatment guidelines as we discuss below. Fees for the visit are covered by public health insurance, which pays physicians on a fee-for-service basis.[11] Our claims data shows all the treatments provided along with their associated diagnoses. The physician must record the diagnosis codes of the health conditions for which the visit is made (even for a rule-out diagnosis), and this information in the claims data allows us to identify DM-related physician visits.

Management of DM. Common and simple preventive actions are known to be effective at improving the health measures that we study. As noted, FBS between 110 and 125 mg/dL is

[11] Patients pay 30% of the fees (except for the elderly and children, whose coinsurance rate is lower; but our data only include working-age adults).

considered pre-diabetes 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). The recommendations at the higher DM threshold usually begin with similar lifestyle modifications; if glycemic control targets are not achieved with lifestyle modification alone, guidelines recommend first-line medication treatment such as oral hypoglycemic agents (Haneda et al. 2018, p.664). 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 and Dunstan 2004).

Distribution of thresholds. As discussed before, 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 provide 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. To get an idea about what thresholds are typically used in actual checkup reports, we report results from a survey conducted by the National Federation of Industrial Health Organizations, which shows the distribution of the "normal range" adopted by more than 300 firms that conduct laboratory testing on behalf of employers.[12]

Figure 2 shows the distribution of the upper bound of the "normal range" reported in the survey. It shows that FBS values of 109 mg/dl and 99 mg/dl are the most common, which implies

[12] The members of the Federation perform testing for over 46 million people per year. The survey is available from http://www.zeneiren.or.jp/cgi-bin/pdfdata/20151014132841.pdf (in Japanese) (accessed September 7, 2020).

that 110 mg/dl and 100 mg/dl are the most common starting values for "out of normal range." The 110 mg/dl threshold makes sense because it corresponds to the pre-diabetes threshold as discussed before. The 100 mg/dl threshold is used in metabolic syndrome screening. However, it is not a threshold for DM diagnosis and not intended to trigger preventive DM care. For this reason, subsequent analyses do not focus on the 100 mg/dl threshold. Importantly, the diabetic threshold (126 mg/dl) does not appear in the distribution reported in Figure 2. This omission probably arises because the survey asks organizations to report the values used to define the "normal range" of FBS, and none considered 126 mg/dl a normal value. Nevertheless, it is still possible that employers send a separate signal at 126 mg/dl, notifying individuals that they are diabetic.

We emphasize that we report the distribution in Figure 2 only for the purpose of motivating our empirical analysis. We do not use any number from Figure 2 to adjust our estimates, nor do we assume that the above distribution holds for the sample that we examine.

We also clarify that in our data we observe the exact FBS values, but not whether individuals received a signal that the FBS value is above a threshold.[13] However, the data are still suited to answer our question, i.e., whether additional spending triggered by a health screening is worth its cost at a commonly reported threshold. This is because we can estimate the effect of moving above the threshold on both spending and health outcomes, and then, by dividing the former (additional spending at the threshold) by the latter (improvement in health outcomes at the threshold), we can examine the cost effectiveness of care at that threshold.

[13] Kim, Lee and Lim (2019) similarly do not observe the actual screening reports that individuals receive, but deduce the signal from the FBS values and national clinical thresholds. The Korean checkup report states "reference range may differ by health checkup institutions"; see Figure 1 in Kim, Lee and Lim (2019).

3. Empirical Framework

3.1. Identification and Estimation

We examine whether health signals affect (i) medical care utilization, (ii) health behavior, and (iii) health outcomes, using an RD design. We exploit the fact that exact checkup thresholds (such as FBS=110 mg/dl and 126mg/dl) are not scientifically optimized thresholds, and that individuals with values just below and above a threshold are similar in their health status. 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 two thresholds are arbitrary as they reflect historical decisions on exact values about which many experts disagree. This arbitrariness is especially true in our setting because thresholds adopted in Japan were primarily chosen to be consistent with western standards, although optimal values are likely to differ for Asian populations.[14]

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

[14] These two thresholds originated in a 1965 World Health Organization expert committee suggesting a threshold of ≥130 mg/dl as diagnostic of diabetes, and <110 mg/dl as non-diabetes (Bloomgarden 2008). In 1997, the American Diabetic Association lowered the diagnostic criterion from ≥130 to ≥126 mg/dl based on findings of increased microvascular risk at this threshold, but did not change the threshold for non-diabetes (Gabir et al. 2000). In 1999, Japan adopted the same threshold of ≥126 mg/dl for diabetes, and <110 for non-diabetes (Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes et al. 2010). A study based on over 13,000 Japanese subjects, however, suggests that a fasting glucose level of 124.3 corresponded to 75 gram oral glucose tolerance test values (200 mg/dl), which is a reliable criterion for diagnosing diabetes (Ito et al. 2000). Subsequent reports stated that Japan chose the 126 (rather than 124.3 or other values) and 110 thresholds to be consistent with international standards. More recently, additional studies have continued to challenge the 110 mg/dl threshold (Yudkin and Montori 2014; Yudkin 2016; Barry et al. 2017; Chatterjee, Khunti, and Davies 2017). These historical choices signify that the exact thresholds of 126 and 110 are arbitrary and do not represent optimal threshold values for DM treatment specific to Japan.

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 individuals who have not previous been 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 pre-diabetes (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 and quadratic 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 the

covariates, as we report later. The αs are parameters to be estimated. The error term μ_{it} is allowed to be correlated over time.[15]

FBS values are available only in integer values. Because there is no clear way to determine the optimal bandwidth in the case of a 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.

3.2. 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 the patient's medical conditions (including suspected ones) related to a given visit, and we observe this list in our data. We categorize the visit as DM-related if DM is included as one of the conditions; 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.

[15] We cluster standard errors by individual 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 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.

Health Outcomes. We use four biomarkers or physical health measures that are most closely related to DM -- i.e., blood sugar level (FBS and HbA1c), BMI, and systolic blood pressure (SBP) -- as our dependent variables. 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 the value of preventive care. In Section 5.5, we introduce a "risk engine," which allows us to calculate to what extent health signals reduce risk of all-cause mortality, if at all.

3.3. Standardized Treatment Effect and Multiple Hypothesis Testing

As we discussed in the previous section, we consider many outcome variables in our estimation. We address this issue in two ways. First, we summarize the effect of a health signal in each of the three domains (medical care utilization, health behavior, and health outcomes) by estimating the average standardized treatment effect. The average standardized treatment is useful in understanding whether there is a global pattern of a health signal on each outcome domain. It can be calculated as follows:

$$\tau = \frac{1}{k}\sum_{k}\frac{\xi_k}{\sigma_k}$$

where ξ_k is the treatment effect of the kth outcome of interest (e.g., the coefficient for the FBS110 dummy variable in the DM-related outpatient medical spending equation) and σ_k is the

standard deviation of the *k*th outcome in the control group.[16] Following Kling and Liebman (2004) and Finkelstein et al. (2012), we stack all individual outcomes within a domain and estimate a seemingly unrelated regression (SUR) that allows error terms to be correlated across outcomes. We further expand the SUR system to incorporate the covariance between ξ_k and σ_k, which allows us to calculate the standard error of τ using the delta method (Kling and Liebman, 2004).

Second, because we have many outcomes, some of the coefficients may become statistically significant due to random chance. To address this issue, we report p-values that account for multiple hypothesis testing within each domain. Each outcome is considered as part of a family of hypotheses that a threshold value has no effect on any of the outcomes within each domain. The adjusted p-values correspond to the probability of rejecting the null hypothesis of no effect for all members of the family given the observed effects. We report the family-wise error rate adjusted p-values based on Westfall and Young (1993) in addition to the significance of coefficients viewed in isolation.[17]

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 JMDC Inc. (JMDC). Individuals with region-based insurers, such as retirees, are not included in the data. As of April 2014, the JMDC claims data

[16] We use individuals with FBS=109 and FBS=125 as the control groups for the pre-diabetes and diabetic-type thresholds, respectively. For high risks, we have many fewer individuals around the thresholds and thus we use individuals with FBS values between 107 and 109 and between 123 and 125 as the control groups for the two respective thresholds.

[17] We calculate the family-wise p-values using a Stata module WYOUNG written by Julian Reif (Reif 2017).

base covers 1.6 million members.[18] 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 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 signal-induced preventive care on health outcomes and medical spending. 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 individual was not diagnosed with DM during the 6 months before the checkup; ii) we have data for the individual at least 6 months before the checkup; iii) we have data for the individual at least 12 months after the checkup; and iv) the individual was 30-64 years old at the time of the checkup.[19]

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 A1, the test statistics become significant when the bandwidth is 4 mg/dl and 10 mg/dl for FBS=110 mg/dl

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

[19] In rare cases, individuals have a checkup more than once in a year. In such cases, we include the first checkup in our analysis.

and 10 mg/dl for FBS=126 mg/dl. However, the statistics also become significant at many 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 pre-diabetes threshold, 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 A1, female has no apparent discontinuity at the thresholds, but age may exhibit a small jump at the FBS=110 mg/dl threshold, although its magnitude is small (approximately 0.2 years). To examine the continuity more rigorously, we conduct a permutation test recently proposed by Canay and Kamat (2018).[20] We find that the null hypothesis of the continuity of the distribution of these covariates is not rejected at both thresholds for both variables (see the first two rows of Table A2). Table A2 shows that our dependent variables are also balanced in most cases.[21] The new test has better small sample properties and is more powerful than existing procedures, including local polynomial regressions (Canay and Kamat, 2018). Nonetheless, we later estimate models with and without covariates and find that the results are almost identical regardless of the specification.

[20] We conduct this test using a Stata module RDPERM written by the authors.

[21] Although a small number of coefficients are significant, the rejection may occur because of multiple hypothesis testing. In fact, if we examine the joint null hypothesis of the continuity of all the variables at the threshold, it is not rejected, as reported in the last row of the table.

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 A2 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 A2, 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).

Barreca et al. (2011) suggest that when there is non-random heaping in the data, we need to be careful about constructing the data set; 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 Crossing the Pre-diabetes Threshold (FBS=110 mg/dl)

In this section, we report how crossing the pre-diabetes threshold affects (i) medical care utilization, (ii) health behavior, and (iii) health outcomes.

Medical Care Utilization. Figure 4 presents the effects on utilization of crossing the FBS=110 threshold, above which the person is considered to suffer from pre-diabetes. Figure 4 clearly shows that all measures of medical care utilization significantly increase at the threshold. In particular, the probability of visiting a physician for DM (Any DM visit) increases about 5

percentage points (from 10% to 15%) at the threshold,[22] 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, which includes spending on DM medications,[23] also increases by around 2,000 JPY (US$18) per year per person. Considering that only 5% of people additionally respond to crossing the threshold, medical spending increases approximately by 40,000 JPY (or US$360) for those who do receive the signal and respond.

In Panel A of Table 2, 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 4, we find that all four measures of utilization significantly increase at the threshold, with magnitudes similar to those shown in Figure 4. 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 these estimates are significant at the one percent level.

Figure A3 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 and quadratic 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 A4 reports the results without covariates. The results are virtually identical to those with covariates, providing further confidence on our empirical findings.

---

[22] This implies that a lower bound on the probability of people received the signal is 5%/90%*100=5.6%.

[23] DM medications are medications categorized as ATC code A10. Please see Appendix I for more about the definitions of variables. As we discuss later, we find no evidence that the health signal increases the use of DM medications.

One potential concern regarding our approach is that the patient might have visited the doctor and learned about elevated blood sugar even without the health signal from a health checkup. This would cause a spurious result which cannot be attributed to the effect of the health signal. We believe, however, that this is not likely in our context because i) undiagnosed DM is asymptomatic and individuals generally do not become aware until symptoms manifest; ii) there are no routine "wellness visits" to a physician in Japan, other than the mandatory health checkup, that could detect high blood sugar; and iii) primary care physicians rarely order blood sugar tests for patients seeking care for other diseases because such diagnostic testing is not covered by insurance. Nevertheless, we estimated the same model using total visits and total medical spending as the dependent variables. As reported in Panel A of Table 2, we find that the estimated coefficients are all positive but they are generally not statistically significant. For example, the coefficient for "total number of visits" is 0.272, which is comparable to the coefficient (0.188) for "number of DM visits" reported in Table 2 but the former is significant only at the 10% confidence level. The result is less significant probably because there is more variability in the data when we consider total utilization.

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 4). One reason for the low response rate may be that only half of those who exceed the threshold possibly receive a warning signal of pre-diabetes, as Figure 2 indicates. Moreover, people may discount the clinical importance of the pre-diabetes signal even when they receive it.

In addition to the six utilization measures, we also separately examine the effect of the signal on the use of DM medications. We look at both the extensive margin (whether any DM medication is prescribed) and intensive margin (total spending on DM medications), finding no significant effect (Figure A5 and A6).

Health Behavior. In Figure A7, we report results for health behavior. In this figure, we observe little effect of crossing the pre-diabetes threshold on health behavior. Local linear regressions reported in Panel B of Table 2 indicate that none of the coefficients are statistically significant. 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, there is no evidence that crossing the threshold affects health behaviors at this lower threshold.

In Figure A8, we check the robustness of the results by altering bandwidths and polynomials. Figure A9 further checks the results without covariates. The results are robust to these checks. Therefore, there is no evidence that crossing the pre-diabetes threshold affects health behavior, unlike the case of medical care utilization.

Health Outcomes. Figure 5 shows the impact on health outcomes of crossing the FBS=110 cutoff. In contrast to the impact on medical care utilization, we observe virtually no discontinuities in health outcomes at the threshold regardless of the health outcome we examine. Estimation results reported in Panel C of Table 2 confirm these observations. In particular, none of the estimated coefficients are significantly negative. As shown in Figures A10 and A11, these results stay the same even when we use different bandwidths, polynomials, and with and without covariates. All these results indicate that there is no evidence that the additional utilization at the pre-diabetes threshold leads to better health.

We note that the null effect is precise because of the large number of observations we have near this threshold. The 95% confidence intervals of our estimates rule out that crossing the pre-diabetes threshold decreases FBS and HbA1c values in the next year by more than 0.2 mg/dl and 0.018%, which correspond to a 0.2% and 0.3% decrease, respectively, from the means just before the threshold. This indicates that meaningful health improvements can be ruled out.

Alternatively, we can examine the precision of the null effect for those who visited physicians with a DM diagnosis, i.e., "treatment on the treated effect." For this, we include "Number of DM visits" (and "DM-related outpatient spending") on the right-hand side and instrument it with an indicator for crossing the 110 FBS threshold. As reported in Table 3, for example, the estimated coefficient is 0.763 for FBS with a standard error of 0.863 (95% CI [-0.963, 2.489]). The corresponding numbers for HbA1c are -0.020, standard error 0.034, and 95% CI [-0.088, 0.048]. Given that the means just before the threshold for FBS and HbA1c are 105 and 5.7, respectively, the largest improvement in health outcomes within the 95% confidence intervals would be -0.963/105*100=0.92% and -0.088/5.7*100=1.54%, respectively, for a DM visit. Thus, the null effect is still precise even in this case.

In addition to the results from individual regressions, in Table 2 we report statistics that take into account the multiple outcomes in each domain. First, family-wise adjusted p-values, reported in square brackets, indicate that the results stay the same with this adjustment; medical care utilization increases significantly, while we find no effect on health behaviors or health outcomes. Second, the standardized treatment effect that summarizes a global pattern within each domain also confirms these findings (see the last row of each panel); while medical care utilization significantly increases at the pre-diabetes cutoff, we do not find significant impacts on

health behaviors or health outcomes.[24] These results are again consistent with our previous results.

An important question is why health outcomes do not improve despite clear increases in utilization at this threshold. As noted in Section 2, physicians may treat a pre-diabetes patient by suggesting lifestyle changes and occasionally by prescribing DM medications. Unfortunately, we have no data on the verbal consultation given in physician offices. However, our analysis so far finds no evidence that crossing a pre-diabetes diagnostic threshold increases utilization of DM medications at either the extensive or intensive margin. This suggests that medication-related issues such as lack of drug adherence are less likely to explain why exceeding the pre-diabetes threshold has no impact. Alternatively, to the extent that preventive visits to physicians might also involve counseling to reduce lifestyle risk factors such as smoking and physical inactivity, this non-response along margins of health outcomes could be interpreted as evidence of the lack of effectiveness of non-drug care at this margin.

5.2. Effects of Crossing the Diabetic-type Threshold (FBS=126 mg/dl)

Medical Care Utilization. Figure 6 shows the effects on medical care utilization of crossing the FBS=126 mg/dl threshold, above which the person is considered a diabetic type. As shown in Figure 6, medical care utilization also appears to increase at this threshold, but the impacts are somewhat weaker than those found for the FBS=110 mg/dl threshold shown in Figure 4. For example, the probability of having at least one DM visit (Any DM visit) appears to increase about the same amount as we found for the FBS=110 threshold. The use of OGTT also appears to

[24] For health behaviors, we calculate the average standardized treatment effect by flipping the sign of the estimated coefficient for "Walk or Exercise." This is because we must have the same expected sign for all coefficients within a domain to calculate the average standardized treatment effect.

increase at the threshold, although we cannot be definitive without a regression analysis. The number of DM visits and DM-related outpatient medical expenditure also appear to increase only slightly at the threshold.

Local-linear regression results reported in Table 4 confirm these observations. The probability of having at least one 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, including on DM medications, do not significantly increase at this threshold. We also estimated the model using different bandwidths, polynomials, and with and without covariates. The results are robust as shown in Figure A12 (with covariates) and Figure A13 (without covariates). As before, we separately examine the effect of the signal on the use of DM medications, finding no robust evidence either at the extensive or intensive margins (Figures A14 and A15).

The somewhat weaker responses at FBS=126 mg/dl than at FBS=110 mg/dl may be because the number of people near this threshold is substantially fewer than that for the pre-diabetes threshold. Another possibility is that as revealed by Figure 2, few individuals receive the diabetic type signal because not many 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. 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 A16 indicates that crossing the threshold continues to have little effect on any measure of health behavior, although

the probability of "walk or exercise" and "drink every day" may have improved somewhat at the threshold. Panel B of Table 4 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 not affected very much at this margin. As before, we also performed robustness checks using different bandwidths and polynomials and with and without covariates. Figures A17 and A18 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, unlike the pre-diabetes threshold—where crossing the threshold had no effect at all on any health behavior—we have weak evidence that crossing the diabetic type threshold affects some types of health behavior, but the result is not robust across behaviors or empirical specifications.

Health Outcomes. Figure 7 shows the results for health outcomes. In Figure 7, intermediate health outcomes such as BMI and FBS are smooth around the FBS=126 mg/dl threshold and there is no clear evidence that crossing 126 improves health outcomes. We confirm these observations in Panel C of Table 4; results from local-linear regressions do not indicate that moving above the FBS=126 mg/dl threshold 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 A19 and A20 there is no evidence that exceeding the diabetic type threshold at a check-up screening affects subsequent health outcomes.

We also report statistics that take into account the fact that we have multiple outcomes in each domain. Family-wise adjusted p-values reported in Table 4 indicate that while the adjustment reduces statistical significance as expected, the overall pattern remains the same even

after the adjustment. Specifically, the coefficient for Any DM visit and OGTT continues to be statistically significant and weakly significant, respectively, even with the adjustment. Similarly, the results for health behaviors and health outcomes remain insignificant after the adjustment. The average standardized treatment effects further confirm these observations; the summary measures indicate that medical care utilization (reported in the last row of Panel A) significantly increases at the threshold, while health behaviors and health outcomes (see Panel B and Panel C, respectively) do not. Again, these results further support previous findings.

5.3. Longer-run Effects on Health Outcomes

In previous sections, we have looked at the short-run effects of screening results exceeding diagnostic thresholds and found no robust evidence that additional care triggered by such screening results improves health outcomes. However, changing lifestyles may take time and medical care can have cumulative effects; if either or both applied, we might observe stronger effects in the longer-run. To assess these possibilities, we examine the effects on health outcomes three and five years after a checkup. Focusing on the pre-diabetes threshold, Figure A21 shows that we find no effects on intermediate health outcomes three years after a checkup. The results for five years after a checkup are similar, as reported in Figure A22.

Regarding the results for the diabetic type threshold shown in Figure A23, we continue to find no statistically significant effects on health outcomes three years after a checkup. While SBP decreases in some models, this effect is not robust. In addition, no effect on health outcomes five years after a checkup is detected, as reported in Figure A24. Thus, even in the longer-run, there is no evidence that additional care for DM (around both the margins of pre-diabetes and diabetic type) improves health outcomes for the general population. These results suggest that any benefit

from incremental medical spending may dissipate rather than strengthen over time, and that it is not easy to make and sustain improvements in lifestyle.

5.4. Effects of Screening on High-risk Individuals

So far, we have examined the impacts of health screening on the general population and found little evidence that exceeding a diagnostic threshold at screening improves subsequent health outcomes. However, instead of recommending office visits to everyone slightly over a single threshold, encouraging follow-up care only for high-risk individuals (i.e., those with multiple risk factors out of the normal range) may be more cost effective. The objective of this section is to investigate this scenario by examining whether FBS values above clinical thresholds have different effects on outcomes for high-risk individuals.

We define an individual as "high-risk" if the person's cholesterol and blood pressure levels are higher than the thresholds set by Japan's official metabolic syndrome screening guidelines. Then, restricting our sample to those who have high cholesterol and elevated blood pressure at an index checkup, we examine whether having an FBS value just above the diabetic-type threshold affects utilization, health behaviors, and health outcomes. Specifically, for cholesterol and other lipids, metabolic syndrome screening guidelines consider an individual to be high-risk if total triglycerides ("bad" lipids) are greater than or equal to 150 mg/dl, or if HDL cholesterol ("good" cholesterol) is less than 40 mg/dl. For blood pressure, an individual is classified as high-risk if systolic blood pressure (SBP) is greater than or equal to 130 mg/dl, or if diastolic blood pressure (DBP) is greater than or equal to 85 mg/dl. From our web search, we also confirm that these values correspond to the thresholds that are most frequently used in checkup reports.[25] Thus,

[25] Specifically, the following percentages of checkup reports adopt the same thresholds as mentioned above: triglyceride (95%), HDL cholesterol (98%), SBP (76%), and DBP (76%).

these high-risk individuals are likely to be told that they have multiple health measures "out of the normal range." We identify 5,666 and 1,136 high-risk individuals in the neighborhood of the pre-diabetes and diabetic type thresholds, respectively,[26] and examine whether the additional warning signal that may arise after screening detects elevated FBS affects utilization and health outcomes.

In Figures 8, Figure A25, Figure 9, and Table 5, we report the impact of crossing the diabetic type threshold on medical care utilization, health behaviors, and health outcomes for high-risk individuals.[27] For medical care utilization (Figure 8), we find that Any DM visit increases significantly and the point estimate (0.120) is larger than for the general population (0.04) reported in Table 4. For health behavior, there is weak evidence that having FBS above the diabetic type threshold reduces the probability of drinking every day (Figure A25 and Table 5). These results hold for a broad range of bandwidths as reported in Figures A32-A34.

The biggest departure from our previous results is that crossing the diabetic type threshold improves intermediate health outcomes, namely by decreasing FBS and HbA1c values. As reported in Figure 9 and Table 5, a high-risk individual who crosses the threshold of FBS=126 mg/dl in year *t* will have FBS and HbA1c values decline by 9.7 and 0.3, respectively, in year *t+1*. These represent about 8% and 5% reductions from the means just before the threshold, respectively. These results are robust across different bandwidths and specifications (Figure

[26] We exclude individuals from the analysis if waist circumference is more than or equal to 85 cm (and 90 cm for women) or BMI is more than or equal to 25 in addition to the person's cholesterol and blood pressure levels exceeding the thresholds. This exclusion is necessary because under the Japan's metabolic syndrome screening, such people may receive "specific health guidance," a guidance conducted under the metabolic syndrome screening program, which may also affect health outcomes. By focusing on those who do not receive the guidance, we can clearly identify the effects of check-up health signals on our outcome measures.

[27] We also performed the same analysis for the pre-diabetes signal and find that none of our health outcome measures are affected by the signal. We report these results in Appendix Figures A26-A31.

A34). The fact that both FBS and HbA1c decrease provides strong evidence that crossing the diabetic type threshold helps to improve DM outcomes for high-risk individuals.

As reported in square brackets in Table 5, many of these results continue to hold even after we adjust p-values for multiple hypothesis testing. Most importantly, exceeding the diabetic type cutoff significantly reduces FBS and HbA1c values at the five percent confidence level even after the adjustment (see p-values in Panel C). We also find that Any DM visit becomes only marginally significant. Turning to the results for the average standardized treatment effects, we find that medical care utilization continues to be statistically significant at the domain level. Moreover, the health outcomes domain shows a significant improvement among high-risk individuals at the higher threshold. This pattern is robust across different bandwidths and polynomials as reported in Figure A35. The same figure shows that the standardized effect for the health behaviors domain is also significant for about half of the bandwidths and polynomials. These results seem to indicate that the crossing the diabetic-type threshold may trigger some lifestyle improvements.

Figure A36 reports the longer-run effects on the four health outcomes. We find that the effect on HbA1c persists three years after an index checkup. In fact, the point estimate (-0.512) of our preferred specification is stronger than the effect after one year (-0.311) (both not reported). The effect on FBS is also negative in all specifications but not statistically significant. Thus, we have weak evidence that the effects on health outcomes last even in the longer-run.

An important remaining question is what may explain the improvements in health outcomes. One possibility is that individuals change their lifestyle after receiving a signal. In fact, our finding that crossing the threshold weakly reduces the standardized treatment effect for "bad health behavior" supports this explanation. Another possibility is that medical treatments such as

DM medications might help reduce FBS and HbA1c. However, as we report in Figure A37, we do not find evidence that crossing the diabetic-type cutoff increases utilization of DM medications either at the extensive margin (Any DM drug) or at the intensive margin (DM drug spending). Moreover, interestingly, we find that medical care utilization, including Any DM visits and number of DM visits, decreases (rather than increases) three years after the index checkup (see Figure A38). Since we observe improvements in health outcomes, while medical care utilization declines in the long-run, it appears that the diabetic-type signal, along with physician office visits, affects individual's health behavior and through that channel improves DM-related health outcomes.

5.5. Is Preventive Care Worth its Cost?

Our results so far indicate physician office visits increase among people whose blood sugar exceeds a pre-diabetes or diabetes threshold at their checkup, suggesting a nontrivial response to a "bad" health signal. An important question is whether the preventive care triggered by the check-up is efficient. For the general population, the answer appears, on average, to be negative around the two thresholds because while utilization clearly increases, we find no evidence that health outcomes improve at either of the thresholds in either the short- or longer-term.

For high-risk individuals, we find evidence that the crossing the diabetic-type threshold increases utilization and improves DM outcome measures. To calculate the cost-effectiveness of preventive care, we use a risk prediction model to estimate the impact of reduced blood sugar on mortality. Risk prediction models utilize an individual's current risk factors (or predictor variables), such as age, sex, and blood pressure, to predict the likelihood of a specific health outcome in the future, such as mortality within the next five years. Quan et al. (2019) developed

risk prediction scores for individuals residing in Hong Kong and Singapore by analyzing detailed clinical data for more than 1 million individuals over age 20 with type 2 diabetes. We use this Hong Kong University – Singapore (HKU-SG) risk score as our outcome measure, because risk scores based on East Asian populations are likely more accurate for Japanese that those based on western populations.

Specifically, we calculate the impact of the reduced blood sugar level identified in the previous section on mortality risk using the HKU-SG risk score in a hypothetical average "high-risk" person. We then calculate the value of reduced mortality assuming a range of values of a statistical life.[28] With respect to the additional spending induced by FBS exceeding a threshold, we consider only DM-related outpatient spending and not all healthcare spending. This is because our objective is to capture the short-run screening-induced spending for the purpose of preventing DM within a year after an index checkup, rather than the long-run impact on healthcare spending due to complications of DM. Because we focus on patients who are *not* diagnosed with DM at the time of an index checkup, non-DM healthcare spending, such as hospitalization for a traffic accident, is unlikely to be related to spending to prevent DM. For this reason, we restrict expenditures to DM-related outpatient spending in this analysis.[29] Appendix II provides greater detail on the construction and application of the HKU-SG risk prediction model.

We find that the estimated reduction in HbA1c of -0.311 (Table 5 Panel C) as evaluated with the HKU-SG risk score yields an average change in mortality risk over the next five years of -0.0870%, with a 95% CI of [-0.1493%, -0.0236%]. The average predicted all-cause mortality in

[28] We recognize that the cost effectiveness calculation ignores the patient's cost of time and quality of life improvements.

[29] We note, however, that we construct DM-related outpatient spending by selecting monthly claims that include DM as one of the individual's medical conditions. For example, if a patient sees a doctor for DM, hypertension, and common cold in month t, we consider healthcare spending due to all these diseases as DM-related spending in month t. Because healthcare costs due to other diseases are also included, our spending measure may actually overestimate the cost related to DM-prevention and thus yield a conservative estimate of cost-effectiveness.

the pre-period is 3.64%, which decreases to 3.55% in the post-period.[30] We estimate the value of this reduction in mortality risk using three different values for one year of life: USD $50,000, $100,000, and $200,000. This range spans commonly used values, including those proposed by the WHO for interventions to be "cost-effective" (cost per life-year less than three times mean per capita GDP) and "very cost-effective" (less than mean per capita GDP) (Hutubessy et al. 2003). For these values of a life-year and an exchange rate of 110 JPY to the dollar, we find that the reduction in mortality risk represents a value of 4,784 JPY (95% CI [1,297, 8,210]), 9,568 JPY (95% CI [2,595, 16,420]), and 19,136 JPY (95% CI [5,191, 32,841]) in the final period, respectively, if we simply multiply the risk reduction by the three aforementioned values of a life-year. However, given that the Quan model calculates risks for a five-year period, we also consider an alternative method which takes into account that the reduction in risk is actually a reduction in five-year mortality rate (See Appendix II below). Doing so, we find that the reduction in mortality rate represents a value of 12,889 JPY (95% CI [3,496, 22,119]), 25,779 JPY (95% CI [6,993, 44,239]), and 51,557 JPY (95% CI [13,986, 88,477]) respectively given values of one year of life of $50,000, $100,000 and $200,000, respectively.

Thus, for all but the lowest value of a life-year, the benefit of reduced mortality risk is comparable to the cost of the medical care utilization induced by crossing the threshold (7,141 JPY as reported in Table 5 Panel A) and average cost of screening (8,449 JPY)[31], implying that screening targeted to high-risks is cost-effective.

5.6 Implications for Optimal Thresholds

[30] This represents a 2.4 percent reduction in risk of death, a non-trivial improvement for otherwise healthy middle-aged individuals.

[31] Based on a survey conducted among 162 health checkup providers. (Source: http://www.zeneiren.or.jp/cgi-bin/pdfdata/20160530094947.pdf accessed on August 30, 2018)

We have examined the cost effectiveness of preventive DM care at two diagnostic thresholds. Can we say anything about the optimality of currently defined thresholds based on these results? First, it should be noted that at the optimal threshold, the return to preventive care should be zero for a marginal patient, and it should be positive for those with a higher FBS as long as the benefit of care increases with FBS values. In our study, we find that for the general population, both the pre- and diabetic thresholds induce additional DM care without improving health outcomes. These results imply that returns to preventive care are likely to be negative at these thresholds. In other words, the optimal thresholds for the general population are likely to be higher than the current thresholds.

For high-risk individuals, we find that crossing a threshold also leads to additional DM care at both pre- and diabetic thresholds. Moreover, health outcomes improve at the diabetic threshold (but not at the pre-diabetes threshold) and we find that the cost of preventive DM care is roughly equal to the benefit of care. This result indicates that for the high-risk individuals we examine, the diabetic-type threshold is close to the optimal value. More generally, because the definition of "high risk" may differ by clinical context, the optimal threshold may be lower than the diabetic threshold if patients have more serious medical conditions than what we consider.

5.7 Additional Analysis

5.7.1. Placebo Tests

As another check on inference, we re-estimate the models using placebo thresholds that range between 115 and 121, a range in which we do not expect any discontinuous increase or decrease in our outcome measure. As reported in Appendix IV, none of the placebo thresholds exhibit any robust effects on utilization, which provides additional confidence on our previous

results.

5.7.2 Effects on Those Who Did Not Receive the Signal in the Previous Year

One might expect that some people 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. Appendix V explores such a possibility by excluding individuals whose FBS value exceeded the pre-diabetes threshold in the previous year. We find that although the estimated impacts are somewhat stronger, the results are not substantially different from what we reported in the main text.

6. Conclusion

Individuals at the cusp of laboratory threshold values ("marginal individuals") often receive warning signals when their health measures exceed a threshold. Are such health signals valuable for the marginal individuals? Using unique individual-level panel data, we investigated whether people respond to checkup results indicating one or more health measures fall outside the normal range, and if so, whether medical care so triggered is worth its cost. We did so in the context of mandatory health checkups in Japan, focusing on marginal patients just above and below clinical thresholds for diabetes and any associated preventive medical care for DM.[32]

We find that, first, individuals whose blood sugar exceeded the threshold for pre-diabetes increase their probability of visiting a physician. For example, we estimate that medical spending

[32] Such "marginal" individuals may constitute a majority of working-age adults. In our sample of over 600,000, 48% have FBS within 10mg/dl of the pre-diabetes threshold (110mg/dl), and another 7% have FBS within 10mg/dl of the diabetic-type threshold (126mg/dl).

increases by approximately 40,000 JPY (or $360) per year for those whose FBS crossed the pre-diabetes threshold. This result confirms that health signals can encourage people to visit physicians' offices for follow-on care, potentially leading to prevention of, or better control of, chronic illnesses like diabetes. However, this result also implies that non-optimally set thresholds may exacerbate wasteful over-use of some types of medical care while not effectively promoting the use of medical resources that are under-utilized relative to their value (Baicker, Mullainathan, and Schwartzstein, 2015).

Second, for the general population, we do not find evidence that additional medical care triggered by screening results above a threshold is effective for either the pre-diabetes or diabetic type thresholds: while DM-related medical care utilization increases at both thresholds, health outcomes do not improve. These results suggest that the current thresholds may need to be reexamined.

Third, focusing on high-risk individuals with elevated blood pressure and high cholesterol, we find that crossing the diabetic type threshold lowers DM-related biomarkers, i.e., both FBS and HbA1c values. Additional analyses utilizing a risk prediction model suggest that the value of predicted life saved due to reduced blood sugar is comparable to additional medical spending triggered by the health signal. This suggests that focusing on those who have multiple risk factors rather than the generic public would be a more fruitful approach. The marginal value of check-up signals could also be increased by stratifying check-up items and frequency in subsequent years by an individual's identified risk, such as exempting low-risk individuals from annual screening of fasting serum glucose levels. Clinicians have pointed out the potential for overtreatment when a large fraction of the population is classified as pre-diabetes (Yudkin and Montori 2014). Our results are consistent with such a concern.

Our results also suggest that clearly communicating health risks to individuals might significantly enhance the value of existing investments in health check-ups. Predicted risks—such as those for mortality that we estimate with a 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 health signals for those at high risk. Such risk prediction models are currently used in the Korean national screening program – which reports with a "traffic light" system – 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 many diagnosis thresholds for multiple conditions that could trigger additional preventive care. 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.

As chronic conditions related to lifestyle loom larger in the burden of mortality and morbidity, understanding appropriate information flows to individuals and providers regarding risk mitigation is becoming ever more important for a larger share of the population. Our methods could be applied to enhance the marginal value of health signals (e.g. from wearable devices) as such signals proliferate.

Acknowledgements

This paper is a substantially revised version of the previously-circulated paper entitled "Is Preventive Care Worth the Cost? Evidence from Mandatory Checkups in Japan." We are grateful

to Philip Clarke, Martin Gaynor, Chie Hanaoka, Daniel Herrera, Hyuncheol Kim, Ming-Jen Lin, Nada Wasi, and seminar and conference participants at Stanford University, University of Southern California, Tokyo University of Science, AEA meetings, AWEHE, EuHEA Conference, iHEA Congress, IRDES-Dauphine Health Economics Workshop, JHEA Conference, and the Korea-Taiwan-Japan Health Economics Associations Joint Conference for helpful comments and suggestions. The Kikawada Foundation, Abe Fellowship Program, and JSPS Bilateral Open Partnership Joint Research Projects are gratefully acknowledged for their financial support. All remaining errors are our own.

References

Alalouf, M., S. Miller, and L.R. Wherry (2019). "What Difference Does a Diagnosis Make? Evidence from Marginal Patients," NBER Working Paper No. 26363, October 2019.

Almond, D., J. Doyle, A. Kowalski and H. Williams (2010). "Estimating marginal returns to medical care: Evidence from at-risk newborns," *Quarterly Journal of Economics*, 591-634.

Baicker, K., Cutler, D., and Song, Z. "Workplace wellness programs can generate savings." *Health Affairs* 29.2 (2010): 304-311.

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

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.

Barry, E., Roberts, S., Oke, J., Vijayaraghavan, S., Normansell, R., and Greenhalgh, T. "Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: Systematic review and meta-analysis of screening tests and interventions." *BMJ* 2017, 356, i6538.

Bloomgarden, Z. T., "Glycemic Control in Diabetes: A Tale of Three Studies" *Diabetes Care* 2008 Sep; 31(9): 1913-1919.

Canay, I. A., and Kamat, V. (2018). "Approximate permutation tests and induced order statistics in the regression discontinuity design." *The Review of Economic Studies*, 85(3), 1577-1608.

Chatterjee, S., Khunti, K. and Davies, M. J. “Type 2 diabetes.” *Lancet* 389, 2239–2251 (2017).

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*, 358: 661-663.

Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes et al. “Report of the Committee on the Classification and Diagnostic Criteria of Diabetes Mellitus” *Journal of Diabetes Investigation*, 2010;1:212–228

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). “Is more information better? The effects of ‘report cards’ on health care providers.” *Journal of Political Economy*, 111(3), 555-588.

Dranove, D., & Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, *48*(4), 935-63.

Finkelstein, A., et al. 2012. “The Oregon Health Insurance Experiment: Evidence from the First Year.” *Quarterly Journal of Economics* 127 (3): 1057–106

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.

Gabir M. M., Hanson R. L., Dabelea D, Imperatore G, Roumain J, Bennett P. H., Knowler W. C. The 1997 American Diabetes Association and 1999 World Health Organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. Diabetes Care 2000;23: 1108 – 1112

Goroll, A. H., “Toward Trusting Therapeutic Relationships — In Favor of the Annual Physical” *N Engl J Med*. 2015; 373:1487-1489.

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.

Handel, B., & Kolstad, J. (2017). Wearable technologies and health behaviors: new data and new methods to understand population health. *American Economic Review*, *107*(5), 481-85.

Haneda, M., et al. "Japanese clinical practice guideline for diabetes 2016." *Journal of diabetes investigation* 9.3 (2018): 657-697.

Herman, W. H., et al. (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.

Hutubessy R, Chisholm D, Edejer TT. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. Cost Eff Resour Alloc. 2003 Dec 19;1(1):8. doi: http://dx.doi.org/10.1186/1478-7547-1-8 PMID: 14687420

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

Iizuka, T., Nishiyama, K., Chen, B., & Eggleston, K. (2017). Is Preventive Care Worth the Cost? Evidence from Mandatory Checkups in Japan. NBER w23413.

Ito, C., Maeda, R., Ishida, S., Sasaki, H., Harada, H. "Correlation among fasting plasma glucose, two-hour plasma glucose levels in OGTT and HbA1c", *Diabetes Res. Clin. Pract.* 50 (2000) 225–230.

Jin, G. Z., & Leslie, P. (2003). "The effect of information on product quality: Evidence from restaurant hygiene grade cards." *The Quarterly Journal of Economics*, 118(2), 409-451.

Jones, D., Molitor, D., & Reif, J. (2019). What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. *The Quarterly Journal of Economics*, *134*(4), 1747-1791.

Kim, Hyuncheol Bryant, Suejin A. Lee, and Wilfredo Lim. "Knowing is Not Half the Battle: Impacts of Information from the National Health Screening Program in Korea." *Journal of Health Economics* (2019). 65 (2019) 1–14.

Kling, J., and Liebman, J. ''Experimental Analysis of Neighborhood Effects on Youth,'' *Mimeo*, 2004; available online at http://www.nber.org/mtopublic/483.pdf.

Kolesár, M., and Rothe, C. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review*, 108 (8): 2277-2304.

Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, *103*(7), 2875-2910.

Lipska, J. K., and Krumholz, M. H. "Is Hemoglobin A1c the right outcome for studies of diabetes?" *JAMA.* 2017 Mar 14; 317(10): 1017–1018.

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

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.

Mehrotra, A., and Prochazka, A. "Improving Value in Health Care — Against the Annual Physical" *N Engl J Med.* 2015 Oct 15;373(16):1485-7.

Myerson, R., Lakdawalla, D., Colantonio, D. L., Safford, M., and Meltzer, D. "Effects of Expanding Health Screening on Treatment - What Should We Expect? What Can We Learn?" NBER Working Paper No. 24347, 2018.

Oster, E. 2018. "Diabetes and Diet: Purchasing Behavior Change in Response to Health Information." *American Economic Journal: Applied Economics*, 10.4 (2018): 308-48.

Oster, E., Shoulson, I., and E. Dorsey (2013). "Optimal expectations and limited medical testing: evidence from Huntington disease." *American Economic Review* 103.2: 804-30.

Patel, S. M., Asch, A. D., and Volpp, G. K. "Wearable devices as facilitators, not drivers, of health behavior change." *JAMA* 313.5 (2015): 459-460.

Quan, J., D. Pang, T.K. Li, C.H. Choi, S.C. Siu, S.Y. Tang, …, G.M. Leung, 2019. "Risk prediction scores for mortality, cerebrovascular, and heart disease amongst Chinese people with type 2 diabetes," forthcoming, *The Journal of Clinical Endocrinology & Metabolism*. https://academic.oup.com/jcem/advance-article/doi/10.1210/jc.2019-00731/5528893.

Reif, J. 2018. "WYOUNG: Stata module to perform multiple testing corrections," Statistical Software Components S458440, Boston College Department of Economics.

Rubin, R. "Debating Whether Checkups Are Time Wasted or Time Misused" *JAMA.* 2019;322(2):101-102

Sacks, D. "A1C Versus Glucose Testing: A Comparison." *Diabetes Care*, (2011) 34(2), 518–523.

Seino, Y., K. Nanjo, N. Tajima, T. et al. “Report of the Committee on the classification and diagnostic criteria of diabetes mellitus.” *Diabetol Int* (2010) 1:2–20.

Song, Z. and K. Baicker. 2019. “Effect of a Workplace Wellness Program on Employee Health and Economic Outcomes: A Randomized Clinical Trial.” *JAMA.* 321(15):1491–1501.

Tanaka, S., Tanaka, S., Iimuro, S., Yamashita, H., Katayama, S., Akanuma, Y., Yamada, N., Araki, A., Ito, H., Sone, H., Ohashi, Y., 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. 1998. "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: 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.

Westfall, P. H. and S. S. Young (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment, Volume 279. John Wiley & Sons.

Whelton P. K., Carey R. M., Aronow W. S., 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 Nov. 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.

Yudkin S. J.“Prediabetes”: Are There Problems With This Label? Yes, the Label Creates Further Problems!. *Diabetes Care* 2016;39:1468–1471

Zhao, M., Konishi, Y., and Glewwe, P. “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	Normal Range
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.

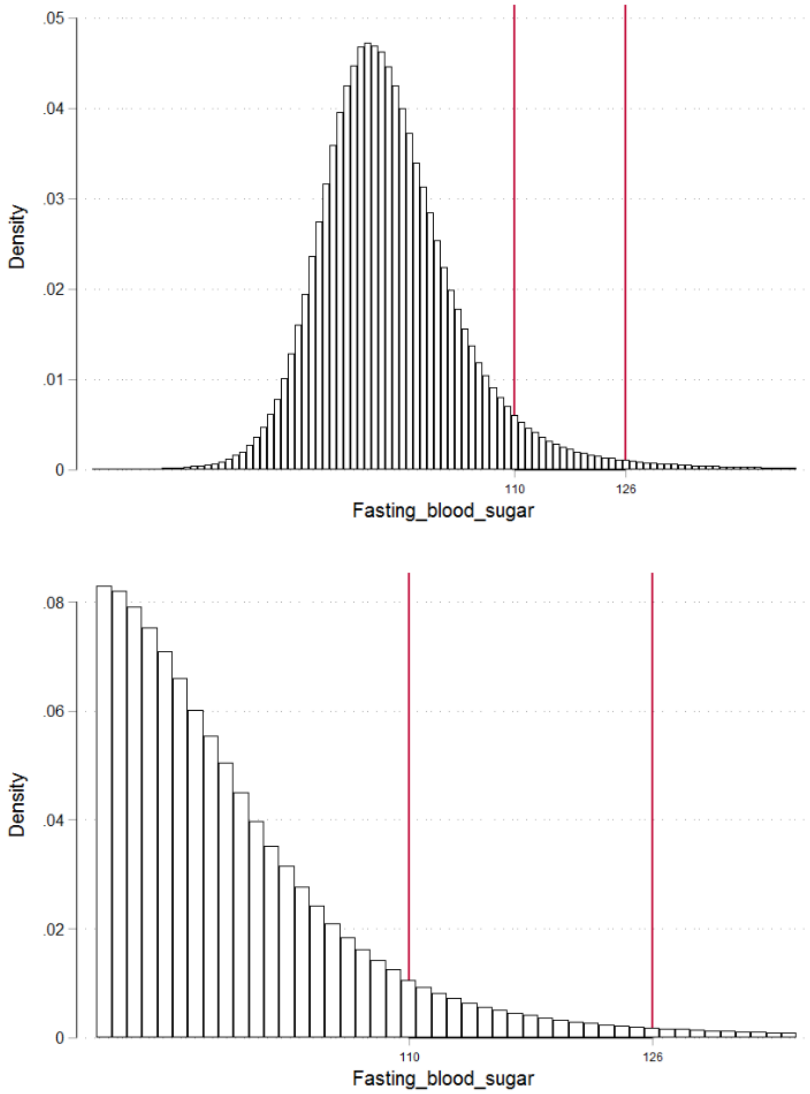
Physician's Signature	**********	Office/Institutions	*** clinic

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

Figure 2. Distribution of the upper bound of the normal range for FBS (N=326)

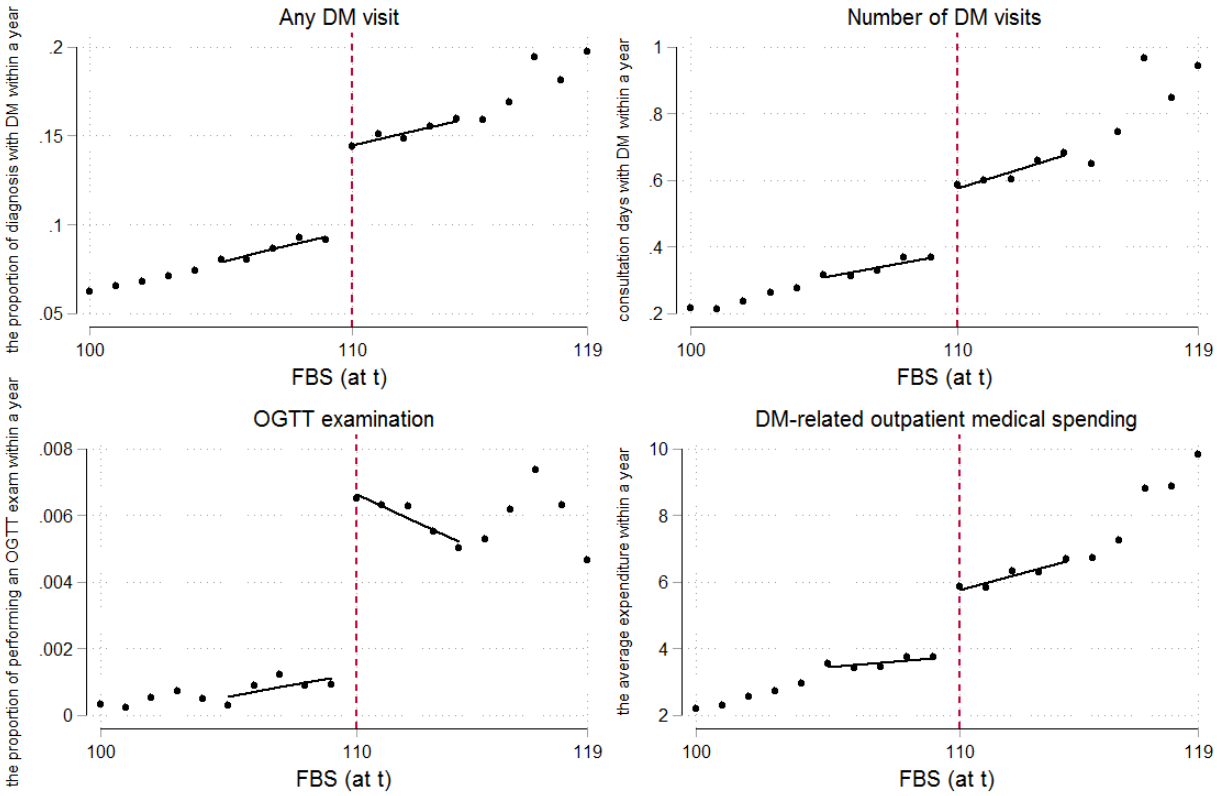
Notes: This figure shows the distribution of the upper limits for the normal range used in medical facilities that conduct health checkups. Each number counts how many facilities out of a total of 326 use that upper limit value for normal FBS. The data is taken from the survey conducted by the National Federation of Industrial Health Organizations (*Zeneiren*) in 2014.

Figure 3. Distribution of FBS values



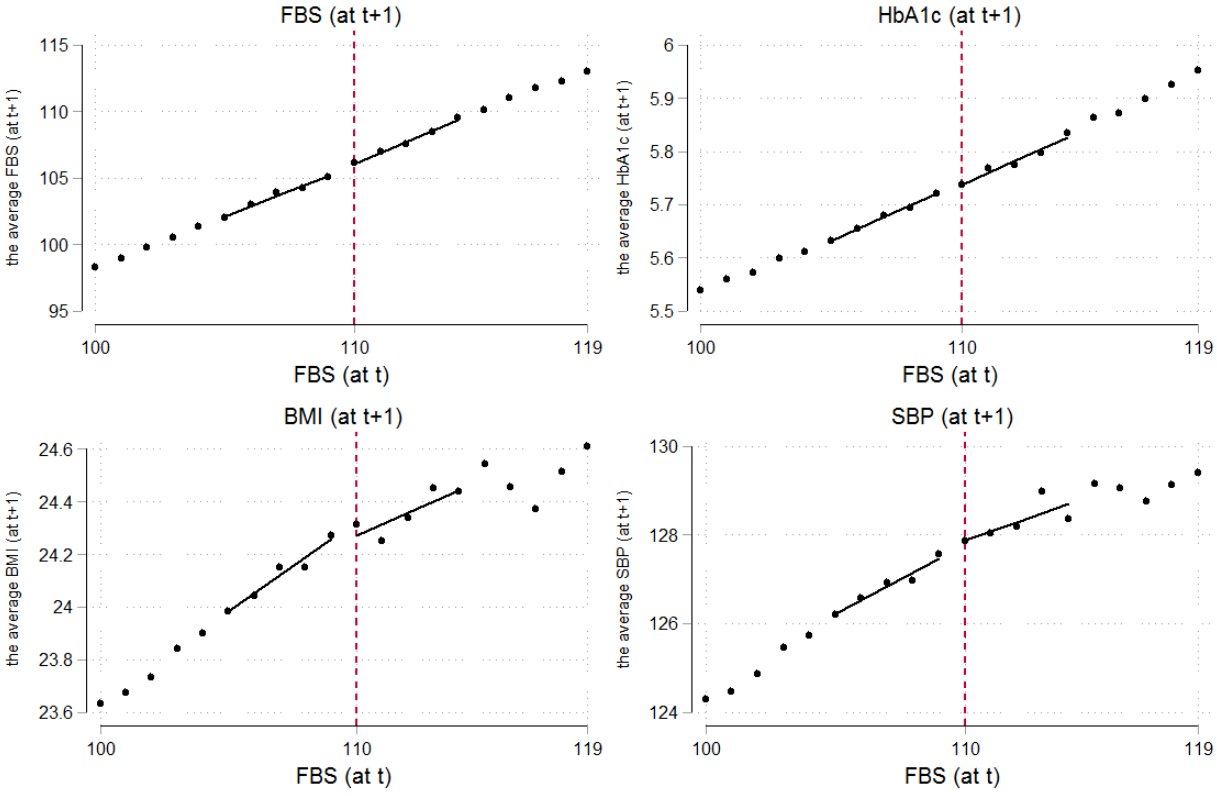
Notes: The histogram shows the density of FBS values within 1 point bins of the FBS value around the "pre-diabetes (FBS=110 mg/dl)" and "diabetic type (FBS=126 mg/dl)" thresholds. The bottom histogram is an enlarged version of the top one, zooming into the area around the thresholds.

Figure 4. Effects on medical care utilization of exceeding the pre-diabetes threshold (FBS=110)



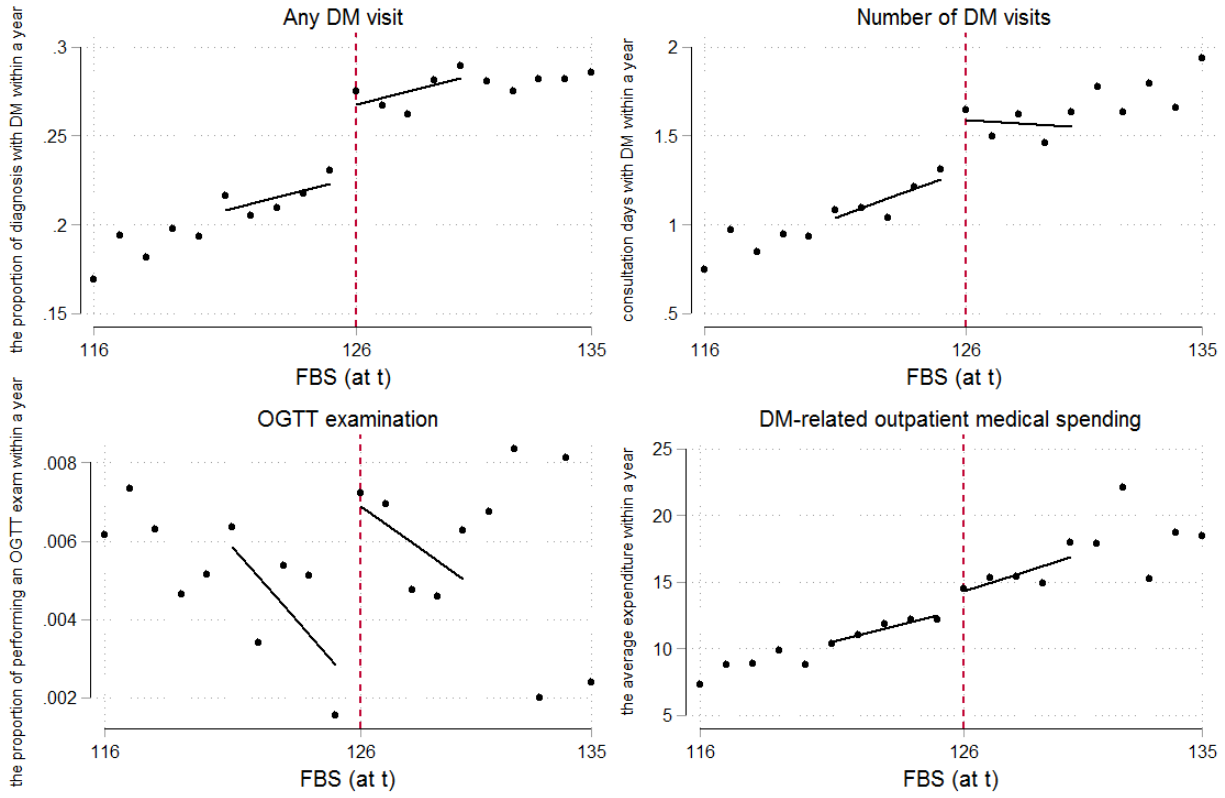
Notes: The scatter plot shows the mean of medical care utilization variable within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 5. Effects on intermediate health outcomes of exceeding the pre-diabetes threshold (FBS=110)



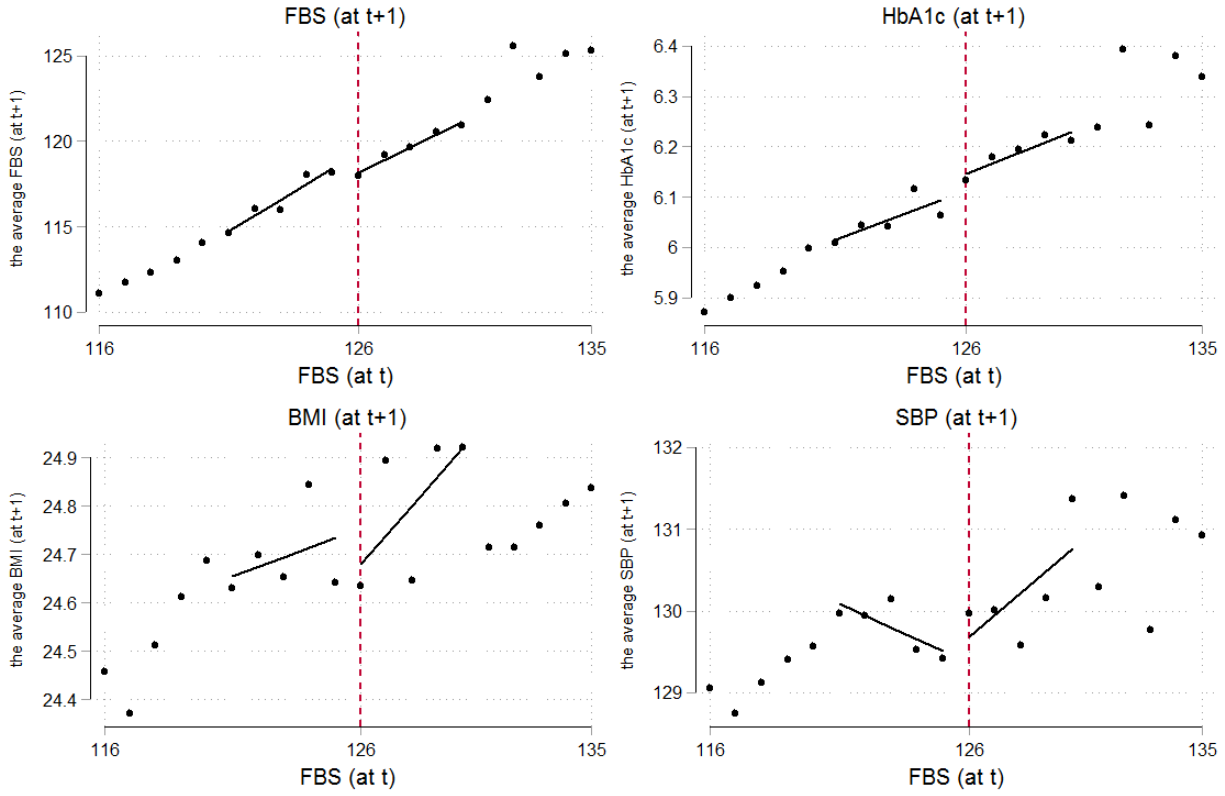
Notes: The scatter plot shows the mean of intermediate health outcome variable within 1 point bins of the FBS value. The vertical line indicates the FBS=110 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 6. Effects on medical care utilization of exceeding the diabetic-type threshold (FBS=126)



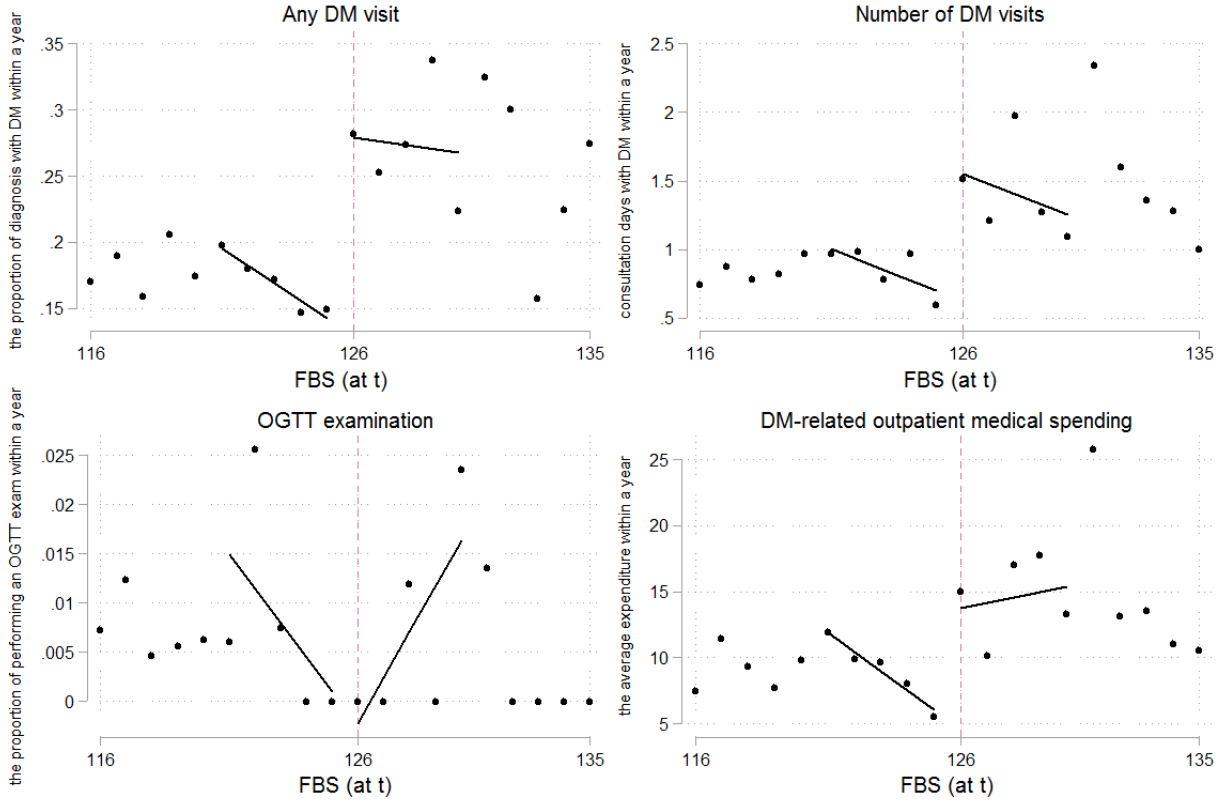
Notes: The scatter plot shows the mean of medical care utilization variable within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 7. Effects on intermediate health outcomes of exceeding the diabetic-type threshold (FBS=126)



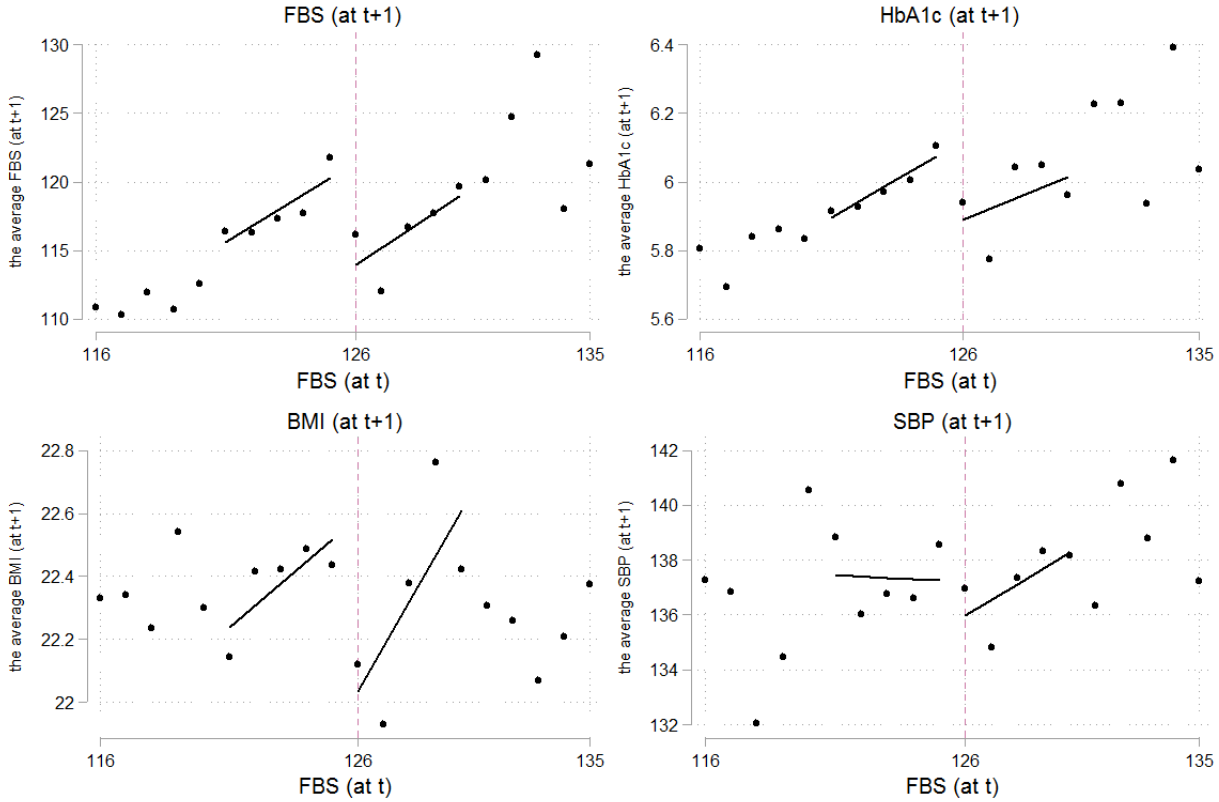
Notes: The scatter plot shows the mean of intermediate health outcome variable within 1 point bins of the FBS value. The vertical line indicates the FBS=126 mg/dl threshold. We fit the values using a linear function within 5 mg/dl of FBS values around the threshold.

Figure 8. Effects of exceeding the diabetic-type threshold (FBS=126) on medical care utilization among high-risk individuals



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 9. Effects of exceeding the diabetic type threshold (FBS=126) on intermediate health outcomes among high-risk individuals



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.

Table 1. Summary statistics

	Full Sample				Sample within 100≦FBS≦119				Sample within 116≦FBS≦135			
	N	Unique N	mean	sd	N	Unique N	mean	sd	N	Unique N	mean	sd
【running variable (at t)】												
FBS (mg/dl)	1,741,209	624,605	92.9	14.2	297,220	171,737	105.6	4.9	43,960	32,228	122.5	5.4
【covariates】												
Age	1,741,209	624,605	45.0	8.3	297,220	171,737	48.2	8.2	43,960	32,228	49.3	8.4
Female	1,741,209	624,605	0.36	0.48	297,220	171,737	0.19	0.40	43,960	32,228	0.14	0.35
【medical care utilization (from t to t+1)】												
Any DM visit	1,741,209	624,605	0.06	0.24	297,220	171,737	0.09	0.29	43,960	32,228	0.22	0.41
Number of DM Visits	1,741,209	624,605	0.24	1.80	297,220	171,737	0.36	2.04	43,960	32,228	1.16	4.27
OGTT examination	1,741,209	624,605	0.001	0.024	297,220	171,737	0.002	0.041	43,960	32,228	0.006	0.074
DM-related outpatient medical spending (in 1,000 JPY)	1,741,209	624,605	2.6	25.6	297,220	171,737	3.7	26.6	43,960	32,228	11.6	51.0
Number of total visits	1,741,209	624,605	7.54	12.05	297,220	171,737	7.97	12.77	43,960	32,228	8.52	13.41
Total medical spending	1,741,209	624,605	83.6	314.3	297,220	171,737	100.2	361.7	43,960	32,228	118.2	379.8
【health behavior (at t+1)】												
Walk or Exercise	967,757	402,150	0.44	0.50	165,535	103,817	0.44	0.50	23,868	18,386	0.44	0.50
Smoke	1,146,643	457,510	0.30	0.46	198,740	122,747	0.33	0.47	30,201	23,022	0.39	0.49
Drink everyday	1,073,649	437,235	0.27	0.44	185,217	115,692	0.36	0.48	27,720	21,328	0.38	0.48
Eat after dinner	940,030	386,449	0.18	0.39	160,551	100,022	0.17	0.38	23,084	17,723	0.16	0.37
【intermediate health outcomes (at t+1)】												
FBS	1,141,578	458,496	93.3	14.1	197,455	122,206	102.6	11.9	29,498	22,466	116.0	21.8
HbA1c	1,015,687	443,461	5.47	0.49	176,564	113,628	5.66	0.47	29,646	22,525	6.05	0.81
SBP	1,209,467	478,494	120.5	15.5	209,922	129,088	126.2	15.4	31,962	24,288	129.7	16.3
BMI	1,209,582	478,505	22.8	3.4	209,949	129,092	24.0	3.5	31,969	24,289	24.6	3.9

Note: This table presents summary statistics for the full sample and for the sample that falls within 10mg/dl of the pre-diabetes threshold (110mg/dl) and diabetic-type threshold (126mg/dl).

Table 2. Effects of exceeding the pre-diabetes threshold (FBS=110)

	Coefficient	mean at 109	Obs.	p-values	
Panel A: medical care utilization					
Any DM visit	0.047***	0.091	120,735	{0.000}	[0.000]
	(0.004)				
Number of DM visits	0.188***	0.368	120,735	{0.000}	[0.000]
	(0.027)				
OGTT examination	0.005***	0.001	120,735	{0.000}	[0.000]
	(0.001)				
DM outpatient spending	1.933***	3.753	120,735	{0.000}	[0.000]
	(0.333)				
Number of total visits	0.272*	8.060	120,735	{0.083}	[0.144]
	(0.157)				
Total medical spending	5.083	100.768	120,735	{0.208}	[0.209]
	(4.037)				
Standardized treatment effect	0.095***			{0.000}	
	(0.012)				
Panel B: health behavior					
Exercise or walk	0.010	0.434	66,870	{0.194}	[0.559]
	(0.008)				
Smoke	-0.006	0.341	80,640	{0.401}	[0.648]
	(0.007)				
Drink every day	-0.010	0.376	75,048	{0.183}	[0.559]
	(0.007)				
Eat after dinner	0.001	0.163	64,924	{0.832}	[0.831]
	(0.006)				
Standardized treatment effect	-0.012			{0.115}	
	(0.008)				
Panel C: health outcomes					
FBS (next year)	0.163	105.1	80,074	{0.375}	[0.829]
	(0.183)				
HbA1c (next year)	-0.004	5.72	73,907	{0.549}	[0.858]
	(0.007)				
BMI (next year)	-0.036	24.3	85,306	{0.487}	[0.858]
	(0.051)				
SBP (next year)	0.094	127.6	85,288	{0.661}	[0.858]
	(0.215)				
Standardized treatment effect	0.000			{0.989}	
	(0.009)				

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. To estimate standardized treatment effects, we use individuals with FBS=109 as the control group. Standard errors, corrected for clustering at the person level, are in parentheses. Per comparison p-values are in curly braces. Family-wise error rate adjusted p-values based on Westfall and Young (1993) are in square brackets. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.

Table 3. Effects of medical care utilization on health outcomes (2SLS fuzzy regression discontinuity estimates)

	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)	
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)	

Note: This table shows the results from the two-stage linear regressions using a fuzzy regression discontinuity framework with a 5 mg/dl bandwidth and covariates. We implemented the estimation using exceeding the threshold as an instrumental variable for two endogenous variables: (A) Number of DM visits, and (B) DM-related outpatient medical spending. Only the coefficients for the endogenous explanatory variables are reported. Standard errors, corrected for clustering at the person level, are in parentheses. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.

Table 4. Effects of exceeding the diabetic-type threshold (FBS=126)

	Coefficient	mean at 125	Obs.	p-values	
Panel A: medical care utilization					
Any DM visit	0.040***	0.230	19,241	{0.002}	[0.010]
	(0.013)				
Number of DM visits	0.264*	1.312	19,241	{0.075}	[0.227]
	(0.148)				
OGTT examination	0.005**	0.002	19,241	{0.018}	[0.079]
	(0.002)				
DM outpatient spending	1.225	12.169	19,241	{0.318}	[0.532]
	(1.228)				
Number of total visits	0.454	8.244	19,241	{0.248}	[0.532]
	(0.393)				
Total medical spending	-11.079	130.008	19,241	{0.351}	[0.532]
	(11.869)				
Standardized treatment effect	0.053**			{0.023}	
	(0.023)				
Panel B: health behavior					
Exercise or walk	0.041**	0.418	10,461	{0.041}	[0.153]
	(0.020)				
Smoke	0.013	0.390	13,322	{0.449}	[0.696]
	(0.017)				
Drink every day	-0.019	0.376	12,201	{0.277}	[0.621]
	(0.017)				
Eat after dinner	-0.003	0.158	10,100	{0.816}	[0.818]
	(0.015)				
Standardized treatment effect	-0.026			{0.179}	
	(0.019)				
Panel C: health outcomes					
FBS (next year)	-1.018	118.1	12,920	{0.197}	[0.545]
	(0.789)				
HbA1c (next year)	0.034	6.06	13,045	{0.247}	[0.562]
	(0.029)				
BMI (next year)	-0.077	24.6	14,085	{0.567}	[0.722]
	(0.134)				
SBP (next year)	0.396	129.4	14,081	{0.475}	[0.722]
	(0.555)				
Standardized treatment effect	0.001			{0.975}	
	(0.023)				

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. To estimate standardized treatment effects, we use individuals with FBS=125 as the control group. Standard errors, corrected for clustering at the person level, are in parentheses. Per comparison p-values are in curly braces. Family-wise error rate adjusted p-values based on Westfall and Young (1993) are in square brackets. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.

Table 5. Effects of exceeding the diabetic type threshold (FBS=126) among high-risk individuals

	Coefficient	mean at 125	Obs.	p-values	
Panel A: medical care utilization					
Any DM visit	0.120**	0.149	1,136	{0.014}	[0.063]
	(0.049)				
Number of DM visits	0.794*	0.595	1,136	{0.055}	[0.205]
	(0.414)				
OGTT examination	0.001	0.000	1,136	{0.831}	[0.836]
	(0.006)				
DM outpatient spending	7.141*	5.514	1,136	{0.065}	[0.213]
	(3.869)				
Number of total visits	2.180	7.438	1,136	{0.156}	[0.378]
	(1.537)				
Total medical spending	68.167	98.152	1,136	{0.247}	[0.447]
	(58.899)				
Standardized treatment effect	0.214**			{0.039}	
	(0.103)				
Panel B: health behavior					
Exercise or walk	0.059	0.357	639	{0.473}	[0.718]
	(0.082)				
Smoke	0.008	0.512	827	{0.915}	[0.913]
	(0.071)				
Drink every day	-0.143*	0.545	751	{0.056}	[0.206]
	(0.075)				
Eat after dinner	-0.052	0.159	624	{0.351}	[0.718]
	(0.056)				
Standardized treatment effect	-0.135			{0.102}	
	(0.083)				
Panel C: health outcomes					
FBS (next year)	-9.722***	121.8	791	{0.008}	[0.028]
	(3.670)				
HbA1c (next year)	-0.311***	6.10	802	{0.007}	[0.028]
	(0.116)				
BMI (next year)	-0.429*	22.43	876	{0.096}	[0.181]
	(0.257)				
SBP (next year)	-1.973	138.6	876	{0.369}	[0.366]
	(2.195)				
Standardized treatment effect	-0.296***			{0.000}	
	(0.082)				

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. To estimate standardized treatment effects, we use individuals with FBS between 123 and 125 as the control group. Standard errors, corrected for clustering at the person level, are in parentheses. Per comparison p-values are in curly braces. Family-wise error rate adjusted p-values based on Westfall and Young (1993) are in square brackets. ***: 1 % confidence level, **: 5 % confidence level, *: 10 % confidence level.