HEALTH INFORMATION AND LIFESTYLE BEHAVIOURS: THE IMPACT OF A DIABETES DIAGNOSIS

Alessio Gaggero

Joan Gil

Dolores Jiménez-Rubio

Eugenio Zucchelli

Title: Health information and lifestyle behaviours: the impact of a diabetes diagnosis

Abstract: We estimate short- and long-term causal impacts of a type-2 diabetes mellitus (T2DM) diagnosis on lifestyle behaviours. We employ a fuzzy regression discontinuity design exploiting the exogenous cut-off value in the diagnosis of T2DM provided by a biomarker (glycated haemoglobin, HbA1c). We make use of unique administrative longitudinal data from Spain and focus on the impact of a diagnosis on clinically measured BMI, smoking and alcohol consumption. We find that, following a T2DM diagnosis, individuals appear to reduce their weight in the short-term. These effects are particularly large among obese individuals and those diagnosed with depression. Patients who are younger, still in the labour market and healthier also present increased short-term probabilities of quitting smoking. In addition, we provide evidence of statistically significant long-term impacts of a T2DM diagnosis on BMI up to three years from the diagnosis. Our results are consistent across parametric and non-parametric estimations with varying bandwidths.

Authors:

Alessio Gaggero	Joan Gil	Dolores Jiménez-Rubio	Eugenio Zucchelli
University of Granada	Universitat de Barcelona and BEAT	University of Granada	MIAS and UAM, Lancaster University, IZA
Email:	**Email**:	**Email**:	**Email**:
alessiogaggero@ugr.es	joangil@ub.edu	dolores@ugr.es	eugenio.zucchelli@uam.es

Date: February 2021

1. Introduction

Acquiring timely information on an individual's health and related risks through screening tests and diagnoses could potentially affect patients' behaviour and reduce avoidable health care use and associated costs (WHO, 2016; Liang et al., 2019). Recent medical studies find that the diagnosis of different types of cancers may affect positively the lifestyle (as measured by smoking, drinking, dietary changes and exercise) of a minority of older male individuals (Jazieh et al., 2006) and that a significant proportion of patients diagnosed with lung or head/neck cancer continue smoking (Burris et al., 2015). Yet, such medical evidence appears to mainly focus on cancer diagnoses and is often based on limited cross-sectional samples as well as simple statistical correlations. Hence, whether a diagnosis might have a causal impact on lifestyle behaviours remains an open empirical question.

While the economics literature on the effectiveness of different types of broad health information campaigns is wide and increasingly exploits causal inference methods (e.g. Brown and Schrader, 1990; Chern at al., 1995; Kim and Chern, 1999; Roosen et al., 2009; Alleis et al., 2015; Fichera and von Hinke, 2020), only a handful of recent economic studies have so far explored the causal impact of a *diagnosis* on lifestyle behaviours (Zhao et al., 2013; Kim et al., 2019; Gaggero, 2020). Zhao et al. (2013) employ a regression discontinuity design (RDD) on data from the China Health and Nutrition Survey (CHNS) to investigate the impact of a hypertension diagnosis on several health outcomes. They find that individuals respond to a hypertension diagnosis by substantially decreasing their fat intake with the effect being especially relevant among wealthier individuals. Similarly, Kim et al. (2019) use a RDD on data drawn from the Korea's National Health Screening Program on type-2 diabetes mellitus (T2DM); hyperlipidaemia (high cholesterol); and obesity diagnoses. Their findings suggest overall limited effects of disease-related risk information if provided in isolation. However, they also find that individuals classified as high risk for diabetes who are prompted for a second visit exhibit reductions in waist circumference in the short-run.[1] Moreover, exploiting a RDD on a sample of individuals aged 50 years and above, Gaggero (2020) provides evidence of a short-term reduction in BMI and waist circumference following a T2DM diagnosis. Hence, results on the causal impact of a diagnosis on lifestyle behaviours appear to be mixed. Furthermore, all these studies appear to focus on short-term effects.

---

[1] Though not based on a formal diagnosis, Cook (2018) also exploits an RDD to estimate the role of personalised weight information on weight change and finds only a moderate impact concentrated among the "very overweight" with higher income.

In addition, the broader correlational literature on the effects of a T2DM diagnosis on health-behaviours also appear to be generally inconclusive. For instance, Chong et al. (2017) suggest a weak association between lifestyle changes and a T2DM diagnosis while Seuring et al. (2020) find some contradicting results among women (decreased BMI but also increased hypertension and reduced physical activity). Relevant to this paper, in one of the very few correlational studies exploring long-term effects, Slade (2012) suggests that the association between a T2DM diagnosis and changes in health-behaviours tend to disappear over time with evidence of recidivism (weight gain and decreased physical exercise) two years after a diagnosis.

The main objective of this paper is to identify short- and long-term causal impacts of a diagnosis of T2DM on key modifiable lifestyle behaviours, including obesity, smoking and alcohol consumption. We focus on T2DM as it is among the most widespread chronic conditions worldwide currently affecting 462 million individuals (6.28% of the world's population); it is the ninth leading cause of death globally; and its burden of disease is projected to raise at an increasingly faster pace in both developing and developed countries (Khan et al., 2020).

We employ a fuzzy RDD on comprehensive longitudinal administrative data from Spain and exploit the exogenous cut-off of an established biomarker used for T2DM diagnoses, namely glycated haemoglobin (HbA1c). We estimate the impact of a T2DM diagnosis on clinically measured BMI, quitting smoking and quitting drinking both short-term (within one year from the diagnosis) and long-term (up to three years after the diagnosis).

We find a statistically significant and substantial reduction in BMI following a T2DM diagnosis. This effect appears to be stronger among obese individuals and those diagnosed with depression. Differently from previous studies, we also provide evidence of statistically significant long-term impacts of a T2DM diagnosis on weight even after three years since the diagnosis. Moreover, younger individuals, those still in the labour market and who are healthier, present increased short-term probabilities of quitting smoking. Our results appear to be consistent throughout alternative parametric and non-parametric specifications with varying bandwidths and further robustness checks.

This paper contributes to the literature in several important ways. First, our findings are produced using uniquely suited panel data, including a reliable biomarker used as the running variable in our RDD approach, and a number of clinically measured anthropometric and health-related variables. Importantly, such data allows for the possibility of identifying long-term effects as well as heterogeneity in the effects of a diagnosis. Secondly, we focus our empirical

exploration on individuals affected by T2DM, one of the most prevalent chronic conditions worldwide. We do so by using a large sample of individuals drawn from the general population in Spain, a European country with one of the highest rates of metabolic diseases (Rojo-Martinez et al., 2020). Thirdly and differently from previous studies on the impact of a T2DM diagnosis employing a sharp RDD, our fuzzy RDD approach explicitly accounts for the possibility that a T2DM diagnosis may not be exclusively based on the value of a single biomarker but on a broader set of patients' characteristics. That is, our approach allows to more realistically take into account that medical doctors could inform their diagnoses using a range of information, including family history and the presence of further health conditions. Fourthly, whereas recent studies in the area (e.g. Kim et al., 2019; Gaggero, 2020) use an objective measure of blood sugar (i.e. Fasting Plasma Glucose, FPG), such measure is sensitive to short-term lifestyle changes and tend to underestimate diabetes prevalence. In this study, we employ a more reliable biomarker (i.e. glycated haemoglobin, HbA1c) which does not suffer from these limitations (American Diabetes Association, 2020).

Overall, this study provides novel causal evidence on the short- and long-term impacts of a T2DM diagnosis by employing a fuzzy RDD on rich longitudinal administrative data. Accordingly, this contributes directly to and extends the growing economics literature on the impact of health information on lifestyle behaviours among people with chronic conditions.

2. Data and institutional setting

2.1 Data

We make use of rich individual-level longitudinal data drawn from administrative records of patients followed over seven consecutive years (2004-2010) in six primary care centres and two hospitals in the municipality of Badalona (north-east of Barcelona, Spain), providing care to a population of around 104,000 individuals. Our sample includes patients aged 16+ who had at least one contact with the health care system between 1 January 2004 and 31 December 2010 and were assigned to one of such healthcare centres.[2] The dataset includes information about each patient's clinical measurements of height and weight; chronic conditions and other diagnosed diseases, according to the International Classification of Primary Care, second edition (ICPC-2); lifestyles (smoking and drinking); date of admission and discharge; type of

---

[2] We exclude from the analysis patients transferred or moved to other centres and those from other areas or regions. However, since movements across centres are rare events, this is unlikely to have a significant impact on our results.

health care professional(s) contacted and reasons of visit. Moreover, the dataset contains patients' age; gender; employment status (active versus non-active); as well as immigration status (defined by the place of birth, i.e. European Union, EU, versus non-EU) and current residence. While the initial sample includes a total of 123,453 patients, given the purpose of our analysis we restrict our estimating sample to 13,971 individuals with at least one biomarker measurement per year.

2.2. Key Variables

Our analysis focuses on the impact of a T2DM diagnosis on modifiable lifestyle behaviours (weight, alcohol and tobacco consumption), while accounting for a number of relevant health-related and sociodemographic observed variables. Our key treatment variable is a binary indicator based on a physician's diagnosis of type-2 diabetes reported using the corresponding ICPC-2 code. We couple this information with the one provided by a biomarker, glycated haemoglobin (HbA1c), which provides an accurate measure of glucose concentration over the previous 8-12 weeks (Goldstein et al. 2004; IEC, 2009; Lyons and Basu, 2012). HbA1c measurements are endorsed by the International Expert Committee (IEC) and the American Diabetes Association (ADA) as they appear to be more reliable if compared to other measures of blood sugar such as the ones based on Fasting Plasma Glucose (FPG) (IEC, 20; ADA, 2010, 2020; Ho-Pham et al., 2017). For instance, FPG tests have a limited time validity; are sensitive to short-term lifestyle changes and stress; and have been found to underestimate the prevalence of diabetes.[3]

Our outcomes of interest are three lifestyle behaviours whose improvement monitored as part of the standard treatment for T2DM prescribed by physicians: (1) body mass index (BMI); (2) smoking; and (3) alcohol consumption.[4] Importantly, our data includes clinically measured BMI, thus avoiding potential issues related to measurement error and reporting bias. Smoking and alcohol consumption are binary variables defining whether an individual is currently a smoker and drinks alcohol, respectively. These are also derived from patients' clinical records as reported by the physician (and thus they are not simply self-reported by patients). Since we can only observe smoking and drinking at the extensive margin, our interest is on the impact of a T2DM diagnosis on quitting smoking and quitting alcohol consumption.

---

[3] This is also a distinctive advantage in terms of the identification within our RDD as it is highly unlikely for patients to manipulate HbA1c scores, i.e. our running variable (see section 3 as well).

[4] BMI is defined as weight divided by the square of height.

2.3 Institutional setting

Our data are drawn from medical records of GP practices and hospitals based in Spain, an EU country with a universal health care system free at the point of delivery with the exception of pharmaceuticals, which may require co-payment (Bernal et al., 2018). In this setting, physicians follow standard national medical guidelines for managing patients with T2DM (Mata et al., 2013) where the threshold value of HbA1c ≥ 6.5 percent is used as the main criteria for a T2DM diagnosis.[5] However, it is worth noting that realistically medical doctors may not base their diagnosis exclusively on HbA1c values and could look at broader patients' characteristics, including family history around T2DM and whether they present other metabolic conditions, such as hypertension or dyslipidaemia (i.e. an abnormal amount of lipids in the blood, including triglycerides and cholesterol). Ultimately, this may imply that some patients with several metabolic conditions and a value of HbA1c just below 6.5 percent may be diagnosed with T2DM. We account for this in our empirical approach by both employing a fuzzy rather than a sharp RDD, allowing the discontinuity driven by a T2DM diagnosis not to depend only on the value of our running variable (HbA1c), and including general practitioners (GP) fixed effects.[6]

3. Econometric Methods

3.1 Regression Discontinuity Design

Our empirical strategy exploits the exogenous cut-off value in the diagnosis of T2DM provided by the biomarker glycated haemoglobin (HbA1c) via a Regression Discontinuity Design (RDD). This allows us identifying the causal impact of a T2DM diagnosis on our outcomes of interest, i.e. lifestyle behaviours.

Specifically, drawing from the potential outcomes framework outlined by Rubin (1974), let D_i be a binary indicator that equals unity for patients diagnosed with T2DM and zero otherwise. Let the pair $\{Y_i^1, Y_i^0\}$ denote the potential outcomes individual i would obtain in case he/she

---

[5] Note that blood sugar levels ranging between $5.7\% \leq$ HbA1c $\leq 6.4\%$ would define a prediabetes condition, i.e. blood glucose levels would be high, but not high enough to prompt a diabetes diagnosis.

[6] Note that upon being diagnosed with T2DM, patients are recommended a series of follow up visits where weight, height and HbA1c measurements are performed. This normally happens during an initial visit as well as further visits at 6 and 12 months since the diagnosis. Lifestyle changes are then monitored during follow up visits (Mata et al., 2013).

does or does not receive a T2DM diagnosis, respectively. The observed outcome, Y_i, can be written as follows:

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0)D_i, \qquad (1)$$

where $(Y_i^1 - Y_i^0)$, the difference between the two potential outcomes, is defined as the causal effect of a T2DM diagnosis for individual i. Since, at any moment, an individual can only be either diagnosed or not, we cannot observe the effect of T2DM diagnosis for each individual (Holland, 1986). Indeed, a naïve comparison between diagnosed versus undiagnosed individuals, would only provide a biased estimate of the true causal effect of the diagnosis. To illustrate this, we can show that the difference in outcomes between individuals with and without a T2DM diagnosis can be written as:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_i^1 - Y_i^0|D_i = 1] + E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0] \quad (2)$$

where $E[Y_i^1 - Y_i^0|D_i = 1]$ is the causal effect of interest (specifically, the average treatment effect on the treated, ATT), and the term $E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]$ represents selection bias, which in this case accounts for baseline differences between the two groups (i.e., diagnosed vs non-diagnosed individuals).

In this study, we exploit the fact that the probability of being diagnosed with T2DM changes discontinuously as a function of the value of the biomarker HbA1c being above or below a pre-determined cut-off and use a RDD to address selection bias. This type of design was first introduced by Thistlethwaite and Campbell (1960) and then formalised by Hahn et al. (2001), who derived the necessary conditions for the identification of causal effects. Let us define the pre-determined cut-off point, h_0, and let the running variable of interest, $h_{i,t}$, be the normalised HbA1c at time t, namely $h_{i,t} = (HbA1c_{it} - h_0)$. The main idea behind RDD is that individuals just above or below the pre-identified cut-off point are otherwise identical but in the probability of being diagnosed, and hence selection bias around the cut-off should disappear (e.g. Imbens and Lemieux, 2008; van der Klaauw, 2008; Lee and Lemieux, 2010; Moscoe et al., 2015).

In general, there are two types of RDD: the sharp and the fuzzy design (see Trochim, 1984). A sharp design assumes that treatment assignment is a deterministic function of the running variable whereas in a fuzzy design treatment assignment is assumed to be a stochastic function of the running variable. In our case, we would use the sharp design if the T2DM diagnosis would depend solely on the running variable, $h_{i,t}$, being above or below the pre-determined

threshold value. However, as mentioned in Section 2, because a T2DM diagnosis is ultimately at the discretion of the physician, and hence may depend also on other patients' characteristics, a fuzzy design is deemed to be more appropriate in this case.

Formally, let us define an indicator variable $Above_{i,t} = 1(h_{i,t} \geq 0)$, which takes the value of unity for patients with values of the HbA1c greater than or equal to the predetermined cut-off level. Then, we can write:

$$\Pr\{D_{i,t} = 1\} = \begin{cases} f_1(h_{i,t}), & if \ Above_{i,t} = 1 \\ f_0(h_{i,t}), & if \ Above_{i,t} = 0 \end{cases}; \tag{3}$$

where, owing to the discontinuity at the cut-off point, $f_1(h_{i,t}) \neq f_0(h_{i,t})$. In the spirit of Hahn et al. (2001), a regression framework for a fuzzy RDD is offered by the instrumental variable approach. Accordingly, we estimate a version of Equation (4) on a series of outcomes of interest as follows:

$$Y_{i,t+1} = g(h_{i,t}) + \beta D_{i,t} + \mathbf{X'_{i,t}}\gamma + \varepsilon_{i,t}, \tag{4}$$

where we use $Above_{i,t}$ as the instrument for the T2DM diagnosis, D_i:

$$D_{i,t} = f(h_{i,t}) + \rho Above_{i,t} + \mathbf{X'_{i,t}}\Omega + \epsilon_{i,t}. \tag{5}$$

where $Y_{i,t+1}$ denotes the outcome of interest and captures lifestyle behaviours at time $t+1$. Functions $g(.)$ and $f(.)$ are flexible polynomials of the running variable, which are allowed to vary around the cut-off.[7] β is the main parameter of interest which, under the additional assumption of monotonicity, represents the local average treatment effect (LATE) of the T2DM diagnosis among compliers around the cut-off. These are the patients whose T2DM diagnosis is actually affected by the HbA1c value being above or below the cut-off. $\mathbf{X'_{it}}$ is a set of pre-diagnosis controls whereas γ and Ω are their associated vectors of parameters. These include patients' demographic characteristics, such as age and gender; a set of dummy variables to account for employment status (active vs non-active); marital status (living vs not living alone); and immigration status (as there is evidence suggesting systematic differences in lifestyle

---

[7] In practice, this is done by interacting each polynomial with the assignment indicator.

behaviours among the immigrant populations, e.g. Carrasco-Garrido et al., 2007). We also account for pre-existing health conditions (hypertension and dyslipidaemia) and time elapsed from the diagnosis. We follow Lee and Lemieux (2010) and in order to reduce sampling error, we include covariates defining baseline (i.e. time t) lifestyle behaviours.

Additionally, our econometric specifications included time-, health area- and GP fixed-effects (FE). In particular, the inclusion of GP FE allows us controlling for systematic differences across physicians that might affect both a T2DM diagnosis and health outcomes. For example, some GPs may be stricter and diagnose patients with T2DM as soon as the HbA1c approaches the pre-determined cut-off, while some others may be more lenient and only officially diagnose T2DM when the HbA1c value reaches higher levels. Similarly, other physicians may be more cautious about the use of drugs and more prone to recommend lifestyle changes, while others may rely mainly on pharmaceutical treatments for the management of T2DM. Our empirical strategy enables accounting for such differences. Finally, $\varepsilon_{i,t}$ and $\epsilon_{i,t}$ are random error terms. We cluster standard errors on the running variable based on the recommendation of Lee and Card (2008).

A RDD can be implemented parametrically by using all data points, both "close to" and "far away" from the cut-off, to estimate Equation (4). Alternatively, one could also use a non-parametric approach (e.g. local linear regression), where Equation (4) would be estimated on a sample of individuals in an arbitrarily small neighbourhood around the cut-off, h_0, and without the polynomial term, $g(.)$. We choose the former approach to make use of the full size of sample, but we also explore the robustness of our results via a non-parametric estimation. Following the recommendations of Gelman and Imbens (2018), our preferred specification (baseline scenario) includes a linear polynomial of the running variable (i.e., $g(.)$ is assumed to be a linear function). However, we further test the robustness of our findings with polynomials of orders between 1 and 4 (see Section 4).

3.2 Validity

RDD methods such as the one employed in this study rely on the standard local continuity assumption typical of all RDD approaches. That is, we are assuming that patients with a HbA1c value just above and below the cut-off should be identical, in terms of both pre-determined observables and unobservables. One indirect way to test for this assumption is to show that the distribution of the observed baseline covariates does not change discontinuously around the cut-off point. Accordingly, Figure 1 displays a set of local polynomial smoothing (LPS)

regressions for pre-diagnosis covariates: *(a)* age; *(b)* gender; *(c)* employment status; *(d)* immigration status; and other potential medical conditions, such as *(e)* hypertension, and *(f)* dyslipidaemia.[8] Figure 1 does not appear to show any significant discontinuity at the cut-off for any of these variables, suggesting that the continuity assumption should hold in this case.

[FIGURE 1 AROUND HERE]

An additional concern for the validity of our identification strategy relates to the potential manipulation of the running variable. McCrary (2008) shows that if individuals can systematically manipulate the value of the running variable in order to (or not to) receive the treatment, then the continuity assumptions will not hold. However, here it seems implausible to assume that patients would be able to manipulate their HbA1c test values and influence a T2DM diagnosis. In particular, since HbA1c values capture the average glucose concentration over the previous 8-12 weeks, in order to significantly change the value of the biomarker, individuals should have started to substantially change their lifestyle 8-12 weeks prior to the test. Although such behaviour seems unlikely, in Figure 2 we investigate this issue by examining whether the density function of our running variable is smoothly distributed around the cut-off. As expected, Figure 2 reveals no evidence of a discontinuity at the (normalised) cut-off.[9]

[FIGURE 2 AROUND HERE]

Finally, Table 1 reports estimates of a placebo test. This test should show a zero effect for the outcome of interest at values other than the pre-determined cut-off value. Indeed, results from our placebo test confirm that when using an alternative cut-off value of HbA1c, namely 4.75%, estimated effects are not significantly different from zero.

[TABLE 1 AROUND HERE]

---

[8] Similar graphs can be obtained for the full set of covariates and are available upon request.

[9] A formal McCrary test for the manipulation of the assignment variable confirms this finding, with an estimated log difference in height of $\hat{\theta} = -0.071$ (s.e. 0.0157), failing to reject the null hypothesis of no discontinuity. Similarly, following Cattaneo et al. (2018) we also run a robust bias-corrected manipulation test based on an unrestricted inference with a local-polynomial of order 2, triangular Kernel, a Jackknife standard error estimator and a data-driven bandwidth selection. A manipulation test score of $T = -0.4614$ (p-value of 0.6445) was obtained, indicating no evidence of systematic manipulation of the running variable.

4. Results

4.1. Sample descriptives

Table 2 displays summary statistics of the variables used in our empirical analysis.[10] In terms of our main outcomes of interest, the table shows that individuals in the estimating sample present an average BMI of around 30 (widely considered as the standard threshold for obesity); that 18 percent of them are current smokers, while only 3 percent report drinking alcohol.[11] The table also reports that 67 percent of individuals in the sample have been diagnosed with T2DM together with an average of 3.1 years since the onset of the disease. The mean HbA1c value for the patients in our sample is of 6.6 percent, just above the threshold used to diagnose T2DM. With respect to sociodemographic characteristics, Table 2 shows an average age of 65; that 52 percent of the sample consists of women; and that 87, 27, and 2 percent of the individuals in the sample are not living alone; still active in the labour market; and are immigrants (born outside the European Union), respectively. Additionally, the Table reports that 59, 53 and 18 percent of the patients are also diagnosed with hypertension, dyslipidaemia and depression, correspondingly.

[TABLE 2 AROUND HERE]

4.2 Main results

Figure 3 presents evidence of the causal impact of a T2DM diagnosis by means of a series of RDD plots of the outcomes of interest as a function of the (normalised) HbA1c, our running variable. Interestingly, the Figure exhibits a statistically significant discontinuity at the normalised threshold for the BMI outcome. Specifically, the plots appear to imply that patients with HbA1c just above the cut-off exhibit a lower BMI value than their counterparts. However, the Figure does not show any significant jump with respect to smoking or drinking cessation. We next test the significance of these findings in a regression framework while controlling for a number of potential confounding factors, as specified in Equations (4) and (5).

[FIGURE 3 AROUND HERE]

---

[10] Summary statistics by year can be found in Table A1 of the Appendix.

[11] The low prevalence of smoking and drinking in our sample might not be entirely surprising as these are all individuals who have been seeking medical attention at some point during the period considered in this study. In addition, a large proportion of them were diagnosed with medical conditions. Hence, it is likely that these patients have been previously advised to change their lifestyle behaviours.

Table 3 reports first stage RDD estimates of a T2DM diagnosis as outlined in Equation (5) using different specifications. Specifically, column (1) includes estimates produced by a model with no covariates; column (2) adds a set of observed covariates; column (3) considers baseline lifestyle behaviours; column (4) includes time and area FE and, finally, column (5) adds GP FE (preferred specification). Estimates appear to be consistent throughout all specifications and show that the parameter of interest (ρ) is positive and statistically significant. This implies that individuals with an HbA1c above the cut-off are significantly more likely to be diagnosed with T2DM.

[TABLE 3 AROUND HERE]

The main estimates of this paper are reported in Table 4. This Table reports the RDD impacts on next-year lifestyle behaviours of interest (Equation 4) based on a parametric approach. Our findings confirm results displayed in Figure 3 implying a statistically significant RDD estimate of a T2DM diagnosis on BMI. More specifically, the estimated coefficient implies that individuals diagnosed with T2DM exhibit a decrease of 1.013 points in their next-year BMI level. Interestingly, in our data a T2DM diagnosis does not appear to affect either smoking or drinking behaviour.[12]

[TABLE 4 AROUND HERE]

4.3 Heterogeneity Analysis

In Tables 5-7, we investigate heterogeneous effects.[13] Table 5 reports RDD estimates when splitting the sample by gender, age, and labour market status. The Table highlights that, with respect to BMI, the impact of a T2DM diagnosis is quantitatively stronger among women, individuals younger than 60 years of age and those active in the labour market (either employed or actively looking for work). Additionally, the Table shows that for patients below the age of

---

[12] We also allowed for the possibility that individuals with BMI measures may not be randomly drawn from our sample of patients and explored the presence of a potential sample selection issue following Wooldridge (2010). Accordingly, we estimated probit models for each time period (wave), where the dependent variable was a dummy taking value 1 for not having a BMI measure and including the full set of exogenous regressors and produced corresponding inverse Mills ratios. These were then included in our RDD models and results produced using such models showed no evidence of selection bias.

[13] Due to the reduced size of the resulting estimating sample, we chose not to rely on, and thus not to report, heterogenous RDD estimates for quitting alcohol consumption. Results are available upon request.

60 and who remain active, a T2DM diagnosis increases significantly the probability of quitting smoking in the short-term.

[TABLE 5 AROUND HERE]

In Table 6, we investigate the differential short-term effects of a T2DM by different weight status, using standard BMI categories (healthy weight, 18.5≤BMI≤24.9; overweight, 25≤BMI≤29.9; obese; BMI ≥ 30; severely obese; BMI≥35). We find the more substantial impact among obese patients, corresponding to a 1.2 points reduction in BMI one year after a T2DM diagnosis. We do not find a statistically significant short-term effect of a diagnosis on quitting smoking across any weight categories.

[TABLE 6 AROUND HERE]

Finally, in Table 7 we explore whether and to what extent the short-term impact of a T2DM differ by pre-existing health conditions by looking at patients with no pre-existing conditions versus those with hypertension, dyslipidaemia, and depression. With respect to BMI, while our estimates do not show a statistically significant effect of the diagnosis among individuals without pre-existing conditions, they do display strong and statistically significant effects for those with previous conditions. Interestingly, the effects on BMI are larger for patients previously diagnosed with depression, leading to a reduction of 1.85 points on the BMI scale.

[TABLE AROUND 7 HERE]

Overall, our findings suggest that, following a T2DM diagnosis, the more substantial BMI reductions occur among individuals with comorbidities, either obese or with pre-existing medical conditions. On the contrary, smoking cessation after a T2DM diagnosis is more likely among diagnosed patients who are younger and with no pre-existing conditions.

4.4 Robustness checks

Our results appear to be robust to a battery of further checks. Table 8 reports RDD estimates obtained using a non-parametric approach focusing on a small neighbourhood around the cut-off. For comparison purposes, Panel A of Table 8 presents the benchmark RDD estimates using the parametric approach. From Panels B to E, we report a series of non-parametric RDD estimates using only observations within intervals defined at varying bandwidths around the

HbA1c threshold of 6.5%. Although the estimates are somewhat smaller in magnitude, the findings appear to confirm a statistically significant decrease in BMI following a T2DM diagnosis, even in a substantially smaller sample. No causal short-term impact is found for either smoking or alcohol cessation.

[TABLE 8 AROUND HERE]

Furthermore, Table 9 shows results obtained by estimating the fuzzy RDD parametrically using different polynomial orders, ranging from a polynomial of order 1 (Panel A or baseline estimate) to a polynomial of order 4 (Panel D). Once again, estimates allowing for different flexibility of the running variable are qualitatively the same.[14]

[TABLE 9 AROUND HERE]

Additionally, we use information about time elapsed since a T2DM diagnosis and estimate Equation (4) by restricting the sample to the more recently diagnosed patients, including only those reported to have a time since diagnosis of less than a year. Despite significantly reducing the sample size, we believe this analysis may provide valuable information as it explores whether the short-term impact of a T2DM diagnosis on lifestyle behaviours could differ among newly diagnosed patients (and ultimately whether the time elapsed from a diagnosis plays a role). Table 10 reports estimates of this exercise and shows that the estimated RDD coefficients are still statistically significant and similar to those reported in Table 4.

[TABLE 10 AROUND HERE]

4.5 Long-term Impacts of a T2DM Diagnosis

Another advantage brought by the use of our longitudinal administrative dataset is that it allows examining both short- and long-term impacts of a T2DM diagnosis. Hence, we also investigate whether the observed reductions in BMI following a T2DM diagnosis are sustained

---

[14] Although issues around non-random heaping in the running variable might not be a concern in our study (because of the accuracy in the measurement of the biomarker), we also run a robustness check around this following Barreca et al. (2011; 2016). As expected, an exploration of the distribution of the running variable reveals no evidence of heaping in the data (corresponding graphs are available upon request).

over time, namely 1, 2, and 3 years after the diagnosis. To do this, we estimate the following version of Equation (4):

$$Y_{i,t+j} = g(h_{i,t}) + \beta D_{i,t} + \boldsymbol{X}'_{i,t}\gamma + \varepsilon_{i,t}, \quad \text{where } j = 1, 2, 3. \tag{6}$$

Accordingly, the estimation of model (6) implies the use of at least four consecutive BMI measurements.[15] Results reported in Table 11 suggest that the estimated BMI reductions after a T2DM diagnosis are observed beyond year 1 and sustained over time. More specifically, we find statistically significant decreases of 1.089, 1.361, and 2.045 on the BMI scale for 1, 2, and 3 years after the T2DM diagnosis, respectively. This appears to imply that the health information included in a T2DM diagnosis might not just have a statistically significant casual impact in the short-term but that it can also have a longer-term impact on BMI reduction. Such sustained causal impact also appears to increase over time with the largest reduction in BMI occurring three years after the diagnosis. This may be the result of a cumulated effect driven by consistent weight losses throughout the years for diagnosed patients.

While this finding might be interesting, we may need to consider that it could be also partly caused by a selection process among patients in our sample. This is because estimates from model (6) are produced using a sample of patients presenting at least four consecutive BMI measurements, thus potentially including more conscientious individuals who might be more aware of the health consequences of their condition. Conversely, those who are excluded from the sample may include individuals who are less conscientious or less prone to follow up on their health status. However, such findings are among the first to present causal evidence on the long-term impact of a T2DM diagnosis on BMI.

[TABLE 11 AROUND HERE]

5. Conclusions

This study provides new evidence on short- and long-term causal impacts of a diagnosis of type-2 diabetes mellitus (T2DM) on key modifiable lifestyle behaviours. In order to identify

---

[15] Note that we estimate Equation (6) on the sample of recently diagnosed individuals. We chose to do so as the identification of the long-term impacts on lifestyle behaviours using the more general sample of patients (including individuals with longer time elapsed from a T2DM diagnosis) might be confounded by other factors. For example, patients with earlier T2DM diagnoses are more likely to have developed further comorbidities or to have experienced additional treatments. Accordingly, this might have an impact on their weight loss over and above the one of a T2DM diagnosis. Indeed, estimates of long-term impacts produced using the general sample appear to confirm this as they present quantitatively larger effects (these are available upon request).

causal impacts, we exploit the exogenous cut-off of blood sugar in the diagnosis of T2DM provided by a reliable and accurate biomarker (HbA1c) via a fuzzy RDD and measure its impact on objectively measured BMI, quitting smoking and quitting drinking. This builds on and directly extends the emerging economics literature on the causal effects of health information on lifestyle behaviours by focusing on T2DM, a major public health concern whose prevalence is projected to increase worldwide. Importantly, we make use of comprehensive longitudinal medical records from Spain, including a number of diagnosed health conditions and relevant sociodemographic variables. This allows identifying long-term impacts as well as exploring heterogeneity in the impact of a diagnosis.

We find a substantial reduction in body weight as a response to a T2DM diagnosis with larger short-term effects among patients who are obese and were diagnosed with depression. Short-term weight reductions appear to be more substantial also among women and middle-aged individuals. Differently from most previous correlational studies, we provide evidence of statistically significant long-term impacts of a T2DM diagnosis on BMI up to three years from the diagnosis. Our results are robust to a series of robustness checks, including parametric estimations with different polynomials orders, as well as non-parametric estimations with several bandwidths.

Interestingly, we also find that patients who are younger, still in the labour market and healthier also present increased short-term probabilities of quitting smoking. However, we should keep in mind that our data are not capable of capturing the intensive margin of smoking (or drinking), and so we cannot observe whether individuals diagnosed with T2DM would eventually reduce the daily number of cigarettes smoked or units of alcohol. It is also worth noting that our paper employs administrative data including individuals seeking medical advice at some point during the period examined in our analysis. As such, these patients might have been already advised to either losing weight or quitting alcohol and smoking before a potential T2DM diagnosis. Thus, our estimates might represent a lower band of the true impact of a T2DM diagnosis on lifestyle behaviours. Yet, this might not be necessarily the case for the long-term effects identified for BMI, as the sample used for the corresponding RDD models might include more conscientious patients who are more prone to follow up their condition.

Our study has some potential limitations. Firstly, although our data includes a measure of employment status, as in most administrative datasets, specific information on education and income is not directly available, therefore limiting the scope of our heterogeneity analysis. Secondly, the lack of information on the specific medications prescribed to patients, prevents us from controlling for their potential effects. Despite such limitations, our paper offers novel

causal evidence on the impact of the health information provided by a T2DM diagnosis on key lifestyle behaviours both in the short-term and over time.

References

American Diabetes Association (ADA) January 01 2020; volume 43 issue Supplement 1.

Allais, O., Etilé, F. and Lecocq, S., 2015. Mandatory labels, taxes and market forces: An empirical evaluation of fat policies. *Journal of Health Economics*, *43*, pp.27-44.

Barreca, A.I., Guldi, M., Lindo, J.M., Waddell, G.R., 2011. Saving Babies? Revisiting the Effects of Very Low Birth Weight Classification. *Quarterly Journal of Economics*, 126(4), pp. 2217-2123.

Barreca, A.I., Lindo, J.M., Waddell, G.R., 2016. Heaping-Induced Bias in Regression-Discontinuity Designs. *Economic Inquiry*, 54(1), pp. 268-293.

Bernal-Delgado, E., García-Armesto, S., Oliva, J., Martínez, F., Repullo, J., Peña-Longobardo, L., ... & Hernández-Quevedo, C. (2018). Spain-health system review.

Brown, D.J. and Schrader, L.F., 1990. Cholesterol information and shell egg consumption. *American Journal of Agricultural Economics*, *72*(3), pp.548-555.

Burris, J.L., Studts, J.L., DeRosa, A.P. and Ostroff, J.S., 2015. Systematic review of tobacco use after lung or head/neck cancer diagnosis: results and recommendations for future research. *Cancer Epidemiology and Prevention Biomarkers*, *24*(10), pp.1450-1461.

Carrasco-Garrido, P., Gil de Miguel, A., Hernández-Barrera, V., Jiménez-García, R., 2007. Health Profiles, Lifestyles and Use of Health Resources by the Immigrant Population Resident in Spain. *European Journal of Public Health*, 17(5), pp. 503-507.

Cattaneo M.D., Jansson M., Ma X., 2018. Manipulation Testing Based on Density Discontinuity. *The Stata Journal*, 18(1): 234-261

Chern, W.S., Loehman, E.T. and Yen, S.T., 1995. Information, health risk beliefs, and the demand for fats and oils. *The Review of Economics and Statistics*, pp.555-564.

Corbin, L.J., Richmond, R.C., Wade, K.H., Burgess, S., Bowden, J., Smith, G.D. and Timpson, N.J., 2016. BMI as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using Mendelian randomization. *Diabetes*, *65*(10), pp.3002-3007.

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

Chong, S., Ding, D., Byun, R., Comino, E., Bauman, A., & Jalaludin, B. (2017). Lifestyle changes after a diagnosis of type 2 diabetes. *Diabetes Spectrum*,*30*(1), 43-50.

Fichera, E. and von Hinke, S., 2020. The response to nutritional labels: Evidence from a quasi-experiment. *Journal of Health Economics*, p.102326.

Gaggero, A., 2020. The effect of type 2 diabetes diagnosis in the elderly. *Economics & Human Biology*, 37, pp. 1-23.

Gelman, A., Imbens, G., 2018. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics* 37(3), pp. 447-456.

Goldstein, D.E., Little, R.R., Lorenz, R.A., Malone, J.I., Nathan, D., Peterson, C.M., Sacks, D. B., 2004. Tests of Glycaemia in Diabetes. *Diabetes Care*, 27(7), 1761-1773.

Hahn, J., Todd, P., van der Klaauw, W., 2001. Identification and Estimation of Treatment Effects with Regression-Discontinuity Design. *Econometrica*, 69(1), pp. 201-209.

Holland, P.W., 1986. Statistics and Causal Inference (with discussion). *Journal of the American Statistical Association*, 81(396), pp. 945-970.

Ho-Pham, L.T., Nguyen, U.D., Tran, T.X. and Nguyen, T.V., 2017. Discordance in the diagnosis of diabetes: Comparison between HbA1c and fasting plasma glucose. *PLoS One*, *12*(8), p.e0182192.

International Expert Committee, 2009. International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. *Diabetes Care*, 32(7), 1327-1334.

Imbens, G.W., Lemieux, T., 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2), pp. 615-635.

Jazieh, A.R., Foraida, M., Ghouse, M., Khalil, M.M., Kopp, M. and Savidge, M., 2006. The impact of cancer diagnosis on the lifestyle and habits of patients served at a Veterans Administration Hospital. *Journal of Cancer Education*, *21*(3), pp.147-150.

Kim, H.B., Lee, S.A. and Lim, W., 2019. Knowing is not half the battle: impacts of information from the National health screening program in Korea. *Journal of Health Economics*, *65*, pp.1-14.

Kim, S.R. and Chern, W.S., 1999. Alternative Measures of Health Information and Demand for Fats and Oils in Japan. Journal of Consumer Affairs, 33: 92-109.

Khan, M.A.B., Hashim, M.J., King, J.K., Govender, R.D., Mustafa, H. and Al Kaabi, J., 2020. Epidemiology of Type 2 Diabetes–Global Burden of Disease and Forecasted Trends. J*ournal of Epidemiology and Global Health*, *10*(1), p.107.

Lee, D.S., Lemieux, T., 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), pp. 281-355.

Lee, D.S., Card, D., 2008. Regression discontinuity inference with specification error. *Journal of Econometrics*,142(2), pp. 655-674.

Liang, Y. W., Chang, H. F., & Lin, Y. H. (2019). Effects of health-information-based diabetes shared care program participation on preventable hospitalizations in Taiwan. *BMC health services research*, *19*(1), 890.

Lyons, T-J., Basu, A., 2012. Biomarkers in Diabetes: Hemoglobin A1c, Vascular and Tissue Markers. *Translational Research*, 159(4), 303-312.

Mata, M., Cos, F.X., Morros, R., Diego, L., Barrot, J., Berengué, M., et al., 2013. Abordatge de la Diabetes Mellitus Tipus 2. 2a. Edició. Institut Català de la Salut 2013 (accessible at http://www.gencat.cat/ics/professionals/guies/diabetis/diabetis.htm).

McCrary, J., 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2), pp. 698-714.

Moscoe, E., Bor, J., Bärnighausen, T., 2015. Regression Discontinuity Designs Are Underutilised in Medicine, Epidemiology and Public Health: A Review of Current and Best Practice. *Journal of Clinical Epidemiology*, 68(2), pp. 132-143.

Narayan, K.V., Boyle, J.P., Thompson, T.J., Gregg, E.W. and Williamson, D.F., 2007. Effect of BMI on lifetime risk for diabetes in the US. *Diabetes Care*, *30*(6), pp.1562-1566.

OECD Statistics. (Organisation for Economic Cooperation and Development). Retrieved October 2, 2020, from https://stats.oecd.org/Index.aspx?ThemeTreeId=9

Public Health Agency of Canada. Diabetes in Canada: national statistics and opportunities for improved surveillance, prevention and control. Available from http://www.phac-aspc.gc.ca/cd-mc/ diabetes-diabete/dic-dac-99/index-eng.php.

Roosen, J., Marette, S., Blanchemanche, S. and Verger, P., 2009. Does health information matter for modifying consumption? A field experiment measuring the impact of risk information on fish consumption. *Review of Agricultural Economics*, *31*(1), pp. 2-20.

Rubin, D., 1974. Estimating Causal Effects of Treatments in Randomised and Non-Randomised Studies. *Journal of Educational Psychology*, 66, pp. 688-701.

Rojo-Martínez, G., Valdés, S., Soriguer, F., Vendrell, J., Urrutia, I., Pérez, V., Ortega, E., Ocón, P., Montanya, E., Menéndez, E. and Lago-Sampedro, A., 2020. Incidence of diabetes mellitus in Spain as results of the nation-wide cohort di@bet.es study. *Scientific reports*, *10*(1), pp.1-9.

Seuring, T., Serneels, P., Suhrcke, M., Bachmann, M., 2020. Diabetes, employment and behavioural risk factors in China: Marginal structural models versus fixed effects models. *Economics & Human Biology*, 39.

Slade, A. N. (2012). Health investment decisions in response to diabetes information in older Americans. *Journal of Health Economics*, *31*(3), 502-520.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. *Journal of Education Phycology*, 51(6), pp. 309-317.

Trochim, W.M.K., 1984. Research Design for Program Evaluation: The Regression Discontinuity Approach. Beverly Hills: Sage Publications.

Van der Klaauw, W. 2008. Regression Discontinuity Analysis: A Survey of Recent Developments in Economics. *Labour*, 22(2), pp. 219-245.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. Second Edition. The MIT Press.

World Health Organization. (2016). Assessing health services delivery performance with hospitalizations for ambulatory care sensitive conditions. *WHO Regional Office for Europe*.

Zhao, M., Konishi, Y. and Glewwe, P., 2013. Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption. *Journal of Health Economics*, *32*(2), pp.367-385.

Tables and Figures

Table 1: RDD Estimates of the Impact of a T2DM Diagnosis on Lifestyle Behaviours - Placebo Cut-off

	(1) Body Mass Index	(2) Quitting Smoking	(3) Quitting Drinking
Cut-off: HbA1C= 4.75			
T2DM Diagnosis [0,1]	-0.124	0.086	-0.325
	(0.550)	(0.060)	(0.364)
Observations	13852	4220	687

Note: The Table reports RDD estimates on the outcomes of interest, when using a placebo cut-off. The alternative cut-off value is 4.75 (%), for which no effects should be observed. Although not shown in the Table, estimates are conditional on a set of covariates, time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

Table 2: SUMMARY STATISTICS

	Mean	S.D.	Min	Max	Obs
Lifestyle Behaviours:					
Body Mass Index	30.29	5.05	14	68	27920
Smoking [0,1]	0.18	0.38	0	1	39688
Drinking [0,1]	0.03	0.17	0	1	39688
T2DM Variables:					
T2DM Diagnosis [0,1]	0.67	0.47	0	1	39688
Onset of T2DM	3.13	3.72	0	39	34741
HbA1C (%)	6.60	1.43	0	20	39994
Demographics:					
Years of Age	65.10	12.62	16	106	39688
Female [0,1]	0.52	0.50	0	1	39688
Not Living Alone [0,1]	0.87	0.33	0	1	39688
Active [0,1]	0.27	0.44	0	1	39594
Immigrant [0,1l	0.02	0.13	0	1	39688
Other Conditions:					
Hypertension [0,1]	0.59	0.49	0	1	39688
Dyslipedimia [0,1]	0.53	0.50	0	1	39688
Depression [0,1]	0.18	0.38	0	1	39688
Years:					
2004	0.11	0.31	0	1	39994
2005	0.12	0.32	0	1	39994
2006	0.13	0.34	0	1	39994
2007	0.14	0.35	0	1	39994
2008	0.15	0.36	0	1	39994
2009	0.17	0.37	0	1	39994
2010	0.18	0.39	0	1	39994
Observations	39994				

Note: The Table reports summary statistics of the main variables of interest.

Figure 1: CONTINUITY TEST
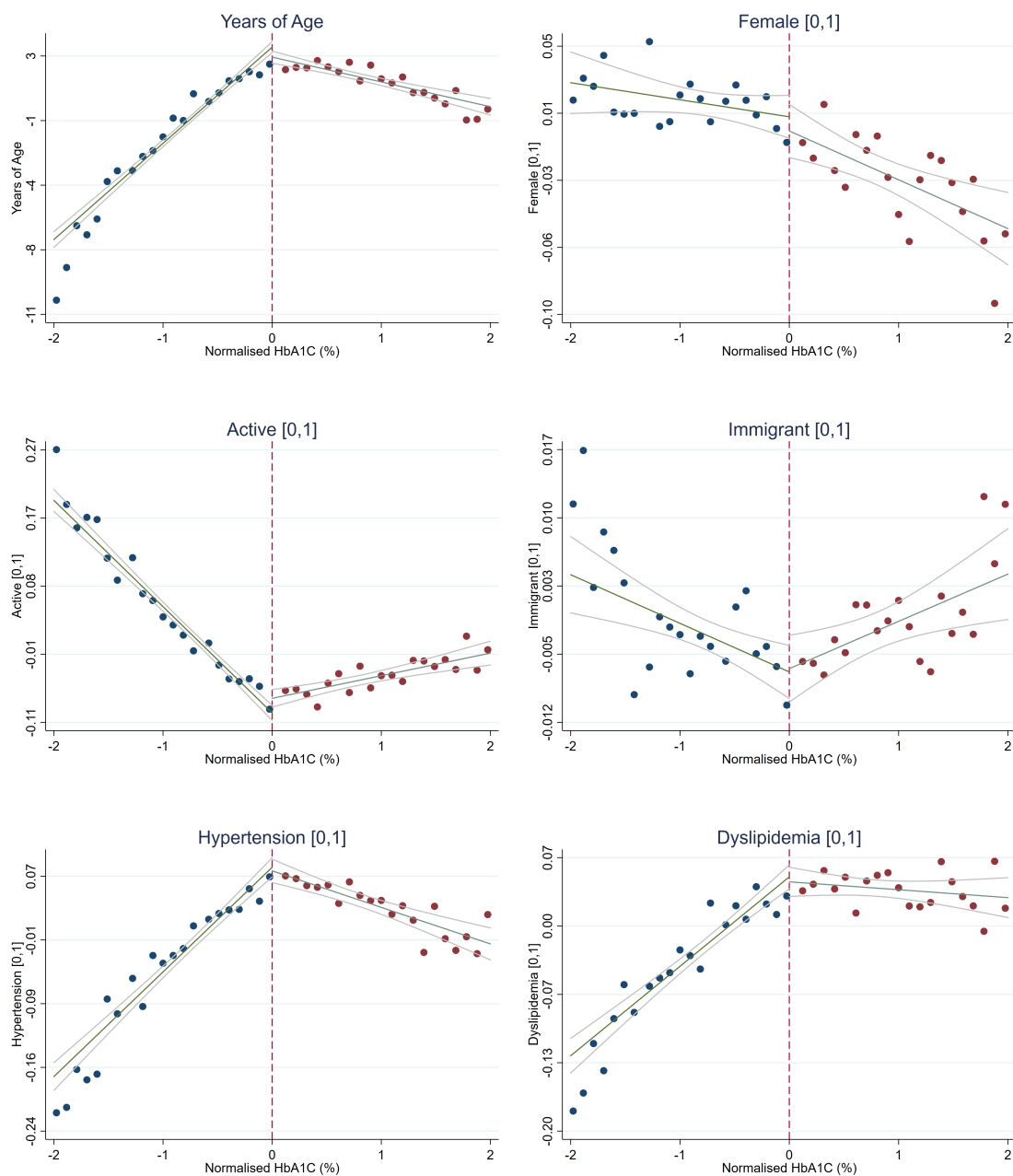
Note: The Figure shows local polynomial estimates of a number of covariates as a function of the running variable.

Figure 2: DENSITY OF THE RUNNING VARIABLE
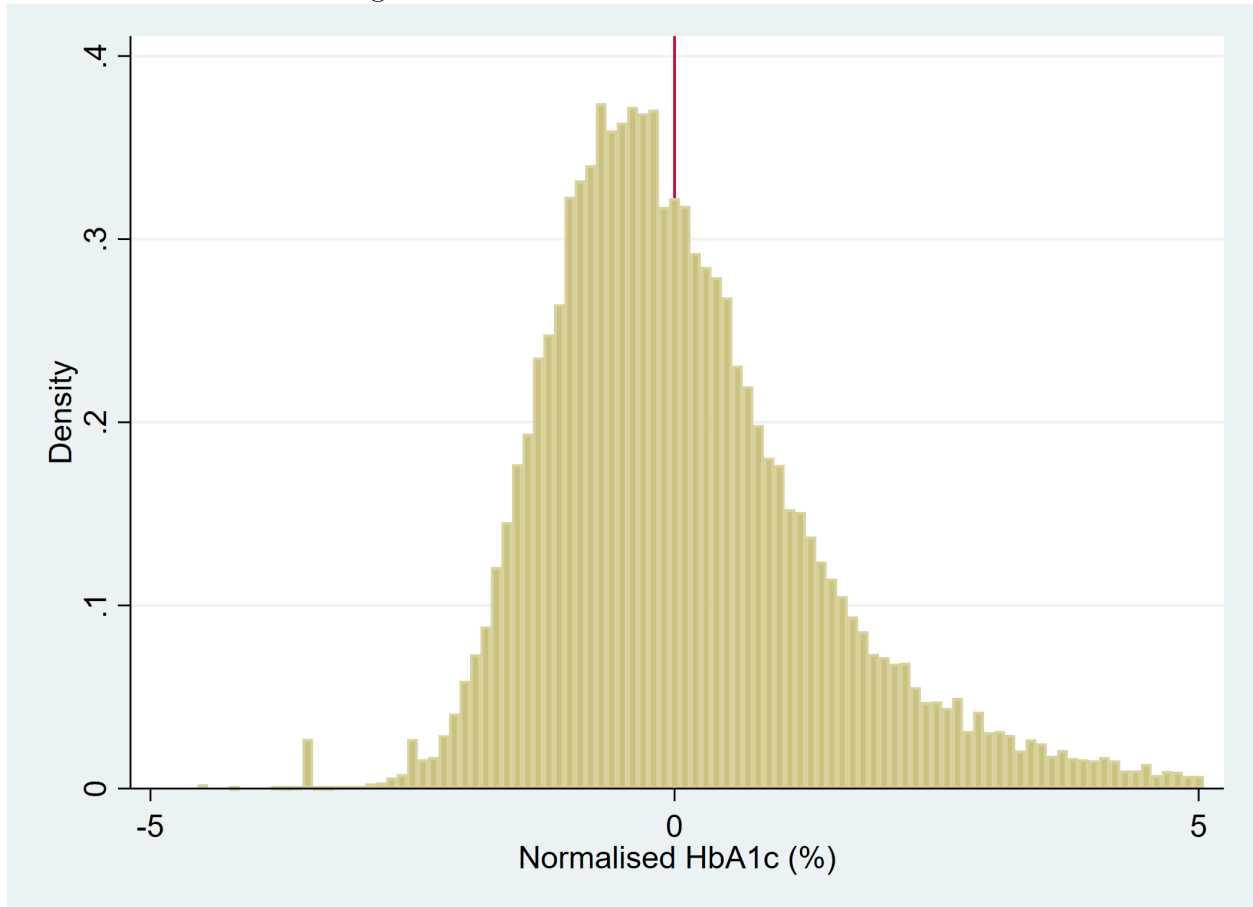
Note: The Figure shows evidence of no manipulation of the running variable. Bin size = 0.1. The bin size has been selected by means of the McCrary test Stata routine, i.e. `DCdensity`.

Figure 3: GRAPHICAL EVIDENCE
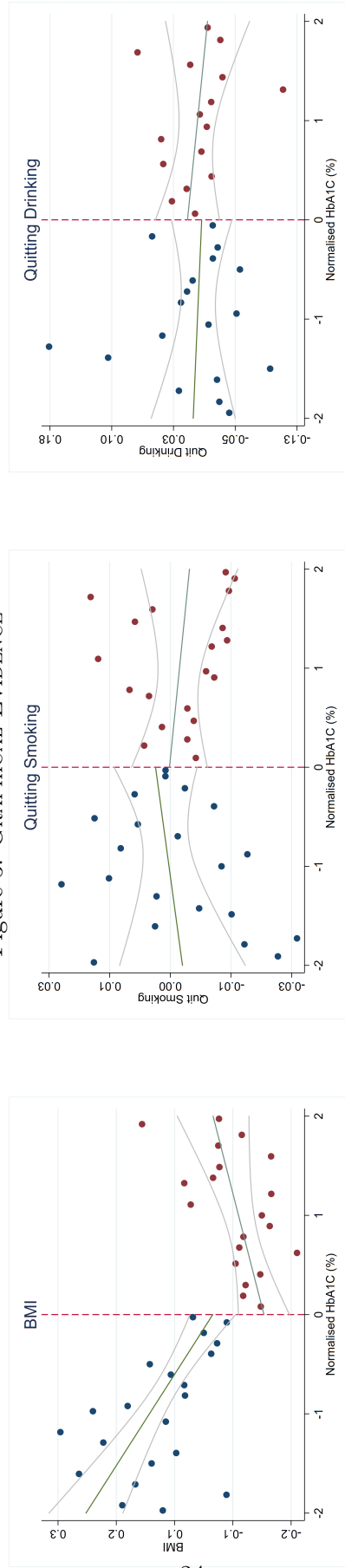
Note: The figure shows local polynomial estimates of the outcomes of interest at time $t+1$, as a function of the running variable at time t.

Table 3: RDD Estimates of a T2DM Diagnosis

	(1) T2DM Diagnosis	(2) T2DM Diagnosis	(3) T2DM Diagnosis	(4) T2DM Diagnosis	(5) T2DM Diagnosis
Above [0,1]	0.196***	0.110***	0.076***	0.070***	0.065***
	(0.032)	(0.025)	(0.022)	(0.023)	(0.022)
Above * HbA1c	-0.236***	-0.162***	-0.184***	-0.203***	-0.199***
	(0.025)	(0.017)	(0.017)	(0.020)	(0.019)
Running Variable:					
HbA1c (%)	0.247***	0.163***	0.184***	0.200***	0.196***
	(0.025)	(0.017)	(0.017)	(0.020)	(0.018)
Attributes:					
Years of Age		-0.004***	-0.004***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)
Female [0,1]		-0.040***	-0.037***	-0.039***	-0.039***
		(0.005)	(0.007)	(0.006)	(0.006)
Not Living Alone [0,1]		-0.021***	-0.027***	-0.021***	-0.023***
		(0.005)	(0.006)	(0.006)	(0.006)
Active [0,1]		-0.076***	-0.074***	-0.013**	-0.012*
		(0.007)	(0.007)	(0.006)	(0.006)
Immigrant [0,1l		-0.069***	-0.070***	-0.027	-0.037*
		(0.018)	(0.024)	(0.021)	(0.021)
Onset of T2DM		0.066***	0.063***	0.073***	0.073***
		(0.006)	(0.006)	(0.005)	(0.005)
Pre-existing Conditions:					
Hypertension [0,1]		0.084***	0.064***	0.042***	0.040***
		(0.005)	(0.005)	(0.005)	(0.005)
Dyslipedimia [0,1]		0.055***	0.051***	0.034***	0.036***
		(0.003)	(0.005)	(0.004)	(0.004)
Depression [0,1]		-0.017***	-0.018***	-0.015***	-0.012**
		(0.005)	(0.006)	(0.005)	(0.005)
Baseline Lifestyle:					
BMI			-0.000	0.001	0.001
			(0.001)	(0.001)	(0.001)
Smoking			-0.000	0.008	0.009
			(0.006)	(0.006)	(0.006)
Drinking			-0.027**	-0.015	-0.010
			(0.012)	(0.011)	(0.011)
Time FE				✓	✓
Area FE				✓	✓
GP FE					✓
Observations	39688	34359	24211	24209	24182

Note: The Table reports RDD estimates of a T2DM diagnosis. Although not shown in the Table, the presented estimates are conditional on time, area, and GP fixed effects. Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Fuzzy RDD estimates of the Impact of a T2DM Diagnosis on Lifestyle Behaviours

	(1) Body Mass Index	(2) Quitting Smoking	(3) Quitting Drinking
T2DM Diagnosis [0,1]	-1.013***	0.014	-0.095
	(0.166)	(0.015)	(0.075)
Running Variable:			
HbA1c (%)	0.019	-0.000	0.001
	(0.017)	(0.001)	(0.006)
Attributes:			
Years of Age	-0.009***	0.000	-0.001
	(0.002)	(0.000)	(0.001)
Female [0,1]	-0.042*	-0.005	-0.020
	(0.023)	(0.006)	(0.035)
Not Living Alone [0,1]	-0.043	0.004	0.015
	(0.037)	(0.006)	(0.036)
Active [0,1]	-0.072*	-0.002	0.003
	(0.042)	(0.005)	(0.036)
Immigrant [0,1l	0.303*	-0.003	-0.299
	(0.178)	(0.027)	(0.225)
Onset of T2DM	0.072***	-0.002	0.011
	(0.013)	(0.001)	(0.008)
Pre-existing Conditions:			
Hypertension [0,1]	0.062**	-0.007	0.029
	(0.027)	(0.005)	(0.021)
Dyslipedimia [0,1]	0.072***	-0.003	-0.007
	(0.021)	(0.006)	(0.022)
Depression [0,1]	0.032	-0.001	0.046
	(0.035)	(0.005)	(0.035)
Baseline Lifestyle:			
BMI	0.968***	-0.000	0.000
	(0.003)	(0.000)	(0.003)
Smoking	0.046	-	-0.041**
	(0.037)	(-)	(0.019)
Drinking	0.163**	-0.003	-
	(0.073)	(0.007)	(-)
Time FE	✓	✓	✓
Area FE	✓	✓	✓
GP FE	✓	✓	✓
Observations	13852	4220	687

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours. Although not shown in the Table, estimates are conditional on time, areas and GP fixed effects. Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: HETEROGENEITY BY SOCIODEMOGRAPHIC CHARACTERISTICS

	(1) Body Mass Index	(2) Quitting Smoking
Men:		
T2DM Diagnosis [0,1]	-0.910***	0.003
	(0.223)	(0.017)
Observations	6353	3309
Women:		
T2DM Diagnosis [0,1]	-1.082***	0.030
	(0.234)	(0.034)
Observations	7499	911
Age<60		
T2DM Diagnosis [0,1]	-1.060***	0.043**
	(0.303)	(0.018)
Observations	3570	2144
Age ≥ 60		
T2DM Diagnosis [0,1]	-0.990***	-0.024
	(0.186)	(0.031)
Observations	10282	2076
Active:		
T2DM Diagnosis [0,1]	-1.140***	0.058**
	(0.346)	(0.023)
Observations	2730	1842
Out of the Labour Force:		
T2DM Diagnosis [0,1]	-0.955***	-0.031
	(0.197)	(0.026)
Observations	11122	2378

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours. Although not shown in the Table, estimates are conditional on a set of covariates, as described in Section 4, as well as time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: HETEROGENEITY BY WEIGHT CATEGORY

	(1) Body Mass Index	(2) Quitting Smoking
Healthy weight:		
T2DM Diagnosis [0,1]	-0.883**	0.005
	(0.377)	(0.040)
Observations	1298	604
Overweight:		
T2DM Diagnosis [0,1]	-0.861***	0.023
	(0.192)	(0.027)
Observations	5805	1882
Obese:		
T2DM Diagnosis [0,1]	-1.205***	-0.015
	(0.268)	(0.025)
Observations	6738	1727
Severely Obese:		
T2DM Diagnosis [0,1]	-0.973*	-0.051
	(0.569)	(0.060)
Observations	2196	501

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours. Following the WHO classification, an individual with a BMI between 18.50 and 24.99 is considered to have a healthy weight; an individual with a BMI between 25 and 29.99 is considered overweight; an individual with a BMI equals to or greater than 30 is considered obese; and an individual with a BMI equals to or greater than 35 is considered severely obese. Although not shown in the Table, estimates are conditional on a set of covariates, as described in Section 4, as well as on time, area and GP fixed effects (FE). Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: HETEROGENEITY BY HEALTH STATUS

	(1) Body Mass Index	(2) Quitting Smoking
No Pre-existing Conditions:		
T2DM Diagnosis [0,1]	-0.694	0.088**
	(0.435)	(0.044)
Observations	2019	855
Hypertension:		
T2DM Diagnosis [0,1]	-0.965***	0.005
	(0.196)	(0.023)
Observations	9371	2312
Dyslipidemia:		
T2DM Diagnosis [0,1]	-1.117***	-0.018
	(0.219)	(0.026)
Observations	8062	2461
Depression:		
T2DM Diagnosis [0,1]	-1.850***	0.009
	(0.585)	(0.049)
Observations	2305	609

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours. Although not shown in the Table, estimates are conditional on a set of covariates as described in Section 4, as well as time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: FUZZY RDD ESTIMATES - NON-PARAMETRIC APPROACH

	(1) Body Mass Index	(2) Quitting Smoking	(3) Quitting Drinking
Panel A: Benchmark			
T2DM Diagnosis [0,1]	-1.013***	0.014	-0.095
	(0.166)	(0.015)	(0.075)
Observations	13852	4220	687
Panel B: Bandwidth = 2			
T2DM Diagnosis [0,1]	-0.716***	0.026	-0.013
	(0.166)	(0.020)	(0.082)
Observations	12621	3687	604
Panel C: Bandwidth = 1.75			
T2DM Diagnosis [0,1]	-0.688***	0.012	-0.115
	(0.172)	(0.024)	(0.130)
Observations	12115	3475	562
Panel D: Bandwidth = 1.5			
T2DM Diagnosis [0,1]	-0.709***	0.020	-0.091
	(0.211)	(0.026)	(0.154)
Observations	11546	3302	525
Panel E: Bandwidth = 1.25			
T2DM Diagnosis [0,1]	-0.742***	-0.011	-0.060
	(0.271)	(0.037)	(0.115)
Observations	10298	2924	465
Panel E: Bandwidth = 1			
T2DM Diagnosis [0,1]	-0.645**	0.049	0.081
	(0.280)	(0.034)	(0.107)
Observations	9258	2584	414

Note: The Table reports non-parametric RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours. Each coefficient in the table report the effect of being diagnosed with T2DM on lifestyle behaviours. Although not shown in the Table, estimates are conditional on a set of covariates, as described in Section 4, as well as time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: FUZZY RDD ESTIMATES - DIFFERENT POLYNOMIALS

	(1) Body Mass Index	(2) Quitting Smoking	(3) Quitting Drinking
Panel A: Linear			
T2DM Diagnosis [0,1]	-1.013***	0.014	-0.095
	(0.166)	(0.015)	(0.075)
Observations	13852	4220	687
Panel B: Quadratic			
T2DM Diagnosis [0,1]	-1.087***	-0.006	-0.119
	(0.246)	(0.017)	(0.106)
Observations	13852	4220	687
Panel C: Cubic			
T2DM Diagnosis [0,1]	-0.737***	-0.012	-0.050
	(0.247)	(0.018)	(0.108)
Observations	13852	4220	687
Panel D: Quartic			
T2DM Diagnosis [0,1]	-0.710***	-0.018	0.037
	(0.242)	(0.021)	(0.124)
Observations	13852	4220	687

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours, when considering different polynomials of the running variable. Although not shown in the Table, estimates are conditional on a set of covariates, as described in Section 4, as well as time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: FUZZY RDD ESTIMATES - RESTRICTED SAMPLE

	(1) Body Mass Index	(2) Quitting Smoking	(3) Quitting Drinking
T2DM Diagnosis	-0.817**	0.107	-0.088
	(0.365)	(0.080)	(0.108)
Running Variable:			
HbA1c (%)	0.025	-0.007	0.004
	(0.051)	(0.008)	(0.012)
Attributes:			
Years of Age	-0.011***	0.001	-0.002
	(0.004)	(0.000)	(0.002)
Female [0,1]	-0.014	0.014*	0.044
	(0.050)	(0.009)	(0.057)
Not Living Alone [0,1]	0.023	0.000	-0.014
	(0.093)	(0.007)	(0.035)
Active [0,1]	-0.109	-0.006	0.005
	(0.086)	(0.007)	(0.040)
Immigrant [0,1l	0.254	0.035	-0.434
	(0.314)	(0.043)	(0.355)
Onset of T2DM	0.000	0.000	0.000
	(.)	(.)	(.)
Pre-existing Conditions:			
Hypertension [0,1]	0.082	-0.012	0.069**
	(0.062)	(0.008)	(0.027)
Dyslipedimia [0,1]	-0.019	-0.002	0.017
	(0.048)	(0.007)	(0.035)
Depression [0,1]	0.001	-0.009	0.063
	(0.064)	(0.010)	(0.055)
Baseline Lifestyle:			
BMI	0.956***	-0.000	-0.003
	(0.006)	(0.001)	(0.004)
Smoking	0.177**	0.000	-0.038
	(0.090)	(.)	(0.031)
Drinking	0.271*	-0.008	0.000
	(0.149)	(0.009)	(.)
Observations	3266	1427	248

Note: The Table reports RDD estimates of the impact of a T2DM diagnosis on lifestyle behaviours, for the sample of individuals not diagnosed at time t Although not shown in the Table, estimates are conditional on a set of covariates, time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: FUZZY RDD ESTIMATES OF BODY MASS INDEX - LONG-TERM EFFECTS

	(1) 1 Year After Diagnosis	(2) 2 Years After Diagnosis	(3) 3 Years After Diagnosis
T2DM Diagnosis [0,1]	-1.089**	-1.361**	-2.045***
	(0.492)	(0.556)	(0.722)
Observations	1535	1210	899

Note: The Table reports RDD estimates of the long-term effect of a T2DM diagnosis on body mass index, for the sample of individuals not diagnosed at time t. Although not shown in the Table, estimates are conditional on a set of covariates, time, areas and GP fixed effects (FE). Robust standard errors are clustered on the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Table A.1: Descriptive Statistics by Year

	(1) 2004	(2) 2005	(3) 2006	(4) 2007	(5) 2008	(6) 2009	(7) 2010
Lifestyle Behaviours:							
Body Mass Index	30.40	30.38	30.34	30.32	30.20	30.18	30.27
Smoking [0,1]	0.18	0.18	0.17	0.18	0.18	0.19	0.18
Drinking [0,1]	0.03	0.03	0.03	0.03	0.03	0.03	0.03
T2DM Variables:							
T2DM Diagnosis [0,1]	0.75	0.75	0.71	0.71	0.69	0.62	0.56
Onset of T2DM	1.84	2.34	2.74	3.18	3.52	3.71	3.73
HbA1C (%)	6.72	6.46	6.57	6.39	6.37	6.72	6.85
Demographics:							
Years of Age	63.38	64.02	64.80	64.98	65.77	65.80	65.95
Female [0,1]	0.54	0.53	0.54	0.52	0.51	0.51	0.51
Not Living Alone [0,1]	0.88	0.87	0.87	0.87	0.87	0.87	0.88
Active [0,1]	0.22	0.23	0.24	0.26	0.27	0.30	0.33
Immigrant [0,1l	0.01	0.01	0.01	0.01	0.02	0.02	0.02
Other Conditions:							
Hypertension [0,1]	0.63	0.62	0.61	0.60	0.60	0.56	0.53
Dyslipedimia [0,1]	0.55	0.56	0.55	0.54	0.53	0.52	0.50
Depression [0,1]	0.18	0.18	0.18	0.18	0.18	0.17	0.17
Observations	4405	4769	5303	5551	5968	6638	7360

Note: The Table reports means and standard deviations of the main variables used in the paper.