The impact of diabetes on labour market outcomes in Mexico: a panel data and biomarker analysis

Abstract

Recent evidence for Mexico suggests important differences in health status between people with diagnosed and undiagnosed diabetes. However, there is at best scarce evidence on the economic consequences of diabetes, especially in contexts where the condition often remains undiagnosed, as is typically the case in low- and middle income countries. Using Mexican longitudinal ~~as well as~~ and biomarker data we estimated the ~~impact of diabetes and diabetesduration on~~ relationship between diabetes, as well as its duration, and employment probabilities, wages and working hours. We further explored how these ~~effects differed~~ relationships differ for those with diagnosed and undiagnosed diabetes. For the longitudinal analyses, nationally representative data from 11836 men and 13745 women 15 to 64 years old were taken from three waves (2002, 2005, 2009) of the Mexican Family Life Survey. We estimated a fixed effects model to account for unmeasured time-invariant confounders of diabetes. We found a reduction in the probability of being employed of ~~5.4 and 5.9~~ 7.7 and 6.3 percentage points for men and women, respectively, but no ~~effects on~~ significant relationship with hours worked or wages. Employment probabilities fell gradually with each year since diagnosis in men, but not women. Using cross-sectional biomarker data, ~~we observed~~ our results indicate that 68% of those exhibiting glycated hemoglobin (HbA1c) levels above the clinical diabetes threshold did not self-report a diagnosis, hence were undiagnosed. Nevertheless, regression analysis revealed that there was no association of diabetes with labour outcomes for undiagnosed women or men. This suggests that results based on self-reported diabetes cannot be extended to the ~~considerable~~ (rather large) population with undiagnosed diabetes, likely because of a selection of people in worse health and with a longer diabetes duration into the diagnosed population. ~~An earlier~~ Earlier diagnosis and improved treatment of diabetes therefore may prevent adverse health effects and related economic hardship in Mexico.

Keywords: Mexico; diabetes; biomarker; wages; fixed effects; employment, working hours

1. Introduction

Diabetes, a disease characterized by elevated blood glucose levels due to the body's inability to use insulin properly, has in the last two decades increasingly become a global problem, with over two-thirds of people with diabetes living in low- and middle-income countries (LMICs) (International Diabetes Federation 2015). In Mexico, diabetes prevalence has grown from 6.7% in 1994 to 14.4% in 2006 (Barquera, Campos-Nonato, et al. 2013) and 15.8% in 2015. Diabetes has become the number one contributor to mortality (International Diabetes Federation 2015), by increasing the risk for heart disease and stroke, blindness, kidney disease and neurologic problems, foot ulcers and amputations (Reynoso-Noverón et al. 2011). However, via effective self-management of the disease through regular monitoring, behaviour change and medication adherence, the occurrence of complications could be avoided or delayed in many cases (Gregg et al. 2012; Lim et al. 2011).

The observed increase in diabetes incidence has been attributed to a deterioration in diet and a reduction in physical activity (Barquera, Hernandez-Barrera, et al. 2008; Basu et al. 2013), while genetic predisposition among Mexicans with pre-Hispanic ancestry may also play a role (Williams et al. 2013). The onset of diabetes has been occurring at an ever earlier age in Mexico (Bello-Chavolla et al. 2017), increasing the risk of complications occurring during the productive lifespan. Only a minority of patients in Mexico achieves adequate blood glucose control (Barquera, Campos-Nonato, et al. 2013). Moreover, diabetes is related to diseases, including depression, hypertension and cardiovascular disease that impose a heavy burden onto the health system (World Health Organization 2016).

Despite the catastrophic impact of diabetes on health, its economic consequences, in particular in LMICs, have received little attention. This applies in particular to the evidence on the effects of diabetes on labour outcomes (Seuring, Archangelidi, et al. 2015). In high-income countries substantial economic losses have been observed (Brown, Pagán, et al. 2005; Brown, Perez, et al. 2011; Brown 2014; Latif 2009; Minor 2011, 2013; Mi-

nor and MacEwan 2016). A rare LMIC study exploited a natural experiment in China and found a significant reduction in income due to a recent diabetes diagnosis (Liu and Zhu 2014). A study for Mexico, using cross-sectional data from 2005, found a significant ($p<0.01$) reduction in employment probabilities for males by 10 percentage points (p.p.) and for females by 4.5 p.p. ($p<0.1$) (Seuring, Goryakin, et al. 2015). Most existing studies relied on instrumental variable (IV) estimation, using the genetic component of diabetes based on its family history, to address the potential endogeneity of diabetes. However, family history of diabetes may also proxy for other genetically transferred traits, including unobserved abilities, as well as intrahousehold or intergenerational dynamics that impact labour outcomes directly; the validity of this IV therefore remains debatable. Panel data methods provide the opportunity to account for time-invariant unobserved individual characteristics, which may play an important role, but—to the best of our knowledge—have not yet been used. Such unobservables, for instance hunger or nutrient deficiency experienced in early life, could adversely affect health as well as the propensity to develop type 2 diabetes later in life (Ewijk 2011; Li et al. 2010; Sotomayor 2013). Additionally, there may also be long-term effects on labour outcomes—either directly through reductions in contemporaneous productivity (Currie and Vogl 2013), or indirectly by limiting educational attainment and human capital accumulation (Ayyagari et al. 2011). These unobservables thereby present a major source of a potential bias that can be accounted for by panel data estimation.

In parallel to these identification challenges, heterogeneity in impact and measurement across the population also deserves further investigation. Recent evidence from Mexico points to a strong positive relationship of diabetes duration with mortality due to diabetes related complications (Herrington et al. 2018). A longer disease duration was found to be related with higher glycated hemoglobin (HbA1c) levels, and undiagnosed diabetes had the lowest diabetes related mortality risks. The latter points to potential selection issues when using self-reported diabetes data to investigate economic outcomes. Those who

self-report, and hence tend to be diagnosed, are in worse health than those undiagnosed. This can lead to an overestimation of the economic effects of diabetes, in particular in populations with a large undiagnosed population, such as in many LMICs (Beagley et al. 2014). So far, however, little evidence exists on the economic impact according to diabetes severity, duration or for those with undiagnosed diabetes.

The objective of this study was to provide new evidence on the impact of diabetes on labour outcomes, adding to previous work by paying close attention to the challenges of unobserved heterogeneity, to the chronic nature of diabetes and to undiagnosed diabetes. We used three waves of the Mexican Family Life Survey (MxFLS), covering the period 2002–2012. Applying a fixed effects model we accounted for time-invariant heterogeneity when assessing the impact of self-reported diabetes and time since diagnosis on labour outcomes. To assess the role of undiagnosed diabetes we used biomarker data from the last wave of the MxFLS.

2. Data

This paper used data from the Mexican Family Life Survey (MxFLS), a nationally representative longitudinal household survey, containing three waves conducted in 2002, 2005–2006 and 2009–2012. It is the only longitudinal household survey in Mexico that provides data on a wide range of social, demographic, economic and health characteristics (Rubalcava and Teruel 2013). Because the survey followed participants moving within Mexico as well as to the US, around 90% of the original sample have been reinterviewed in the third wave. Our samples were restricted to the working age population (15–64) and excluded pregnant women. Pregnant women have an increased diabetes risk and may not be able to work. Since their inclusion may have biased the estimates, we dropped all observations of women reporting to be pregnant at the time of the survey (N=764). We also dropped those reporting to be in school. The first part of the analysis used all three waves, ex-

ploiting the panel structure of the data. The second part used a biomarker subsample of the third wave (2009–2012). Because the biomarker sample included everybody above the age of 44, but only a random subsample of those aged 44 or below (Crimmins et al. 2015), its age structure was older and hence its self-reported diabetes prevalence higher. The analysis therefore compares with self-reported data for this specific subsample only.

Our outcome variables of interest were employment status, weekly working hours, hourly wage, and occupation. Employment status was defined as having carried out an activity that helped with the household expenses the last week while working for at least four hours per week. We explicitly included informal employment and employment without monetary remuneration, for instance in family businesses. Hourly wage was constructed as reported monthly income from the first and second job, divided by average number of weeks per month and weekly working hours. Labour income was obtained from the response to questions on wages, income from piecework, tips, income from extra hours, meals, housing, transport, medical benefits and other earnings, or from the response to a question on aggregate labour income for the entire month. We adjusted calculated wages for inflation in the year of interview and considered the log of real wages. Due to a considerable number of missing or zero income reports, the sample used for the wage estimation was smaller than the sample for working hours. Working hours were combined from both the first and a potential second job. Descriptive statistics for the entire panel sample show that over 80% of men with diabetes and 87% of men without diabetes reported some form of employment, compared to 26% of women with diabetes and 37% of women without diabetes (see Table 1). Interestingly, men did not report considerably higher hourly wages than women but worked more hours per week. There were also little differences in working hours and wages between men and women with and without diabetes. Men worked more often in agricultural jobs while women were more likely to be self-employed or in non-agricultural wage employment. The educational attainment of women was lower than that for men on average. Similarly, those without

diabetes were better educated than those with diabetes. Further, the diabetes sample is about 15 years older on average than the non-diabetes sample, both for men and women.

The first part of the analysis focused on the relationship of labour outcomes with self-reported diabetes, which was based on the survey question: "Have you ever been diagnosed with diabetes?". Because the data did not distinguish between type 1 and type 2 diabetes, we assumed that the estimates represented the impact of type 2 diabetes, by far the most common type of diabetes in Mexico. As a robustness check, we re-estimated our main results categorizing diabetes into early-onset and late-onset cases, according to the age at which diabetes was first reported in the survey. This was a similar approach to Alegre-Díaz et al. (2016), who assumed that everybody diagnosed before age 35 and using insulin had type 1 diabetes. Accordingly, we assumed that those first reporting a diabetes diagnosis before the cut-off had type 1 diabetes while those above had type 2 diabetes. Nonetheless, because we cannot warranty that this is 100% accurate (as it ~~may be~~ is unlikely that both populations consisted exclusively of one type of diabetes) we preferred to think of the groups as of early- and late-onset groups. This separation also provides information about the effects for different age groups, as the late-onset group had an average age of onset of 50 compared to 28 for the early-onset group. In the pooled data, which combines all three waves, diabetes was self-reported by 5% of men and 6% of women. This is consistent with other reports from Mexico for the time, showing a prevalence of diagnosed diabetes of 7.5% in 2006 in a sample also including people over the age of 64 (Barquera, Campos-Nonato, et al. 2013). Apart from self-reported diabetes, which was available in all rounds, we also used information on the self-reported year of diagnosis as well as biometrically measured HbA1c levels for a subsample of respondents from the third wave.

Information on the self-reported year of diagnosis, reported in the third wave, allowed us to construct a measure of time since diagnosis. For those also present in previous waves, we inferred the time since diagnosis by the difference between the year of the interview and the year of diagnosis. This allowed us to use panel data methods for the duration

analysis as well, however limited to those reporting the year of diagnosis in the third wave.

The second part of the analysis assessed the role of undiagnosed diabetes. The ~~biometrically measured blood glucose value~~ HbA1c levels that allowed us to identify those with undiagnosed diabetes ~~, was~~ were available for over 6000 respondents in the third wave. We used the internationally recognized cut-off of an HbA1c $\geq 6.5\%$ to define diabetes as recommended by the World Health Organization (WHO) (World Health Organization 2011). As we show in Supplementary Table S5, 19% of self-reported diabetes cases had HbA1c levels below the diabetes threshold. We dropped those for our analysis as it was not clear if they had misreported their diabetes status or had achieved these low levels as a result of their successful disease management. Analysis including those cases led to qualitatively similar results (results available on request).

3. Estimation strategy

To investigate the relationship between self-reported diabetes and three labour outcomes—employment, weekly working hours and wages—we estimated a fixed effects model. The fixed effects model accounts for the potential bias introduced by time-invariant unobservables, providing an estimate of the ~~effect~~ association for cases that received a diagnosis throughout the survey.

$$Y_{it} = \beta_0 + \beta_1(D_{it} - \overline{D}_i) + \beta_2(X_{it} - \overline{X}_i) + \beta_{3i} + \beta_{4i} + (u_i + e_{it}), \quad (1)$$

The fixed effects model ~~uses~~ used only the within-person variation for identification, i.e. the difference between the diabetes indicator D_{it} and its cluster mean $\overline{D}_i$, so that β_1 represented the within-person variation of diabetes over time. The same ~~applies~~ applied to the other time-varying covariates X_{it}. Y_{it} was a binary variable taking a value of 1 if respondent i reported being in employment at time t and 0 otherwise. For ease of interpretation we chose to estimate a linear probability model for the ~~effects of diabetes~~

~~on~~ association of diabetes with employment.

To estimate the ~~effect on~~ association of diabetes with working hours and wages, our empirical models were estimated conditional on being in employment. Y_{it} represented the log hourly wage or the weekly working hours over the last year, for respondent i at time t.

~~We also included time variant confounders – as the fixed effects model accounts for time invariant confounding only. We controlled for changes in~~ In the main fixed effects (FE) models we only included calender year dummies as time-variant control variables. Other potential time-variant control variables to account for socioeconomic, demographic, geographic or health changes throughout the observation period could have been affected by the onset of diabetes, and were not controlled for as this would have prevented a causal interpretation of the relationship of diabetes with our labour market outcomes (Angrist and Pischke 2009). So is it thinkable that a diabetes diagnosis affected the place of residence, for example as people move back to their family to receive additional help. Diabetes may also have affected a person's chances to become married, as potential spouses could be deterred by a diabetes diagnosis and the potential health consequences it entails. Similarly, we did not account for changes in wealth, in particular because changes in employment outcomes due to diabetes would more than likely have affected the overall wealth of the person and its household. We also did not account for obesity. While part of the effect of diabetes may be due to potential adverse effects of obesity, its inclusion in the model would have lead to attenuated estimates if the diagnosis of diabetes also had an effect on body mass index (BMI), which has been shown to be the case in other studies (De Fine Olivarius et al. 2015; Seuring, Suhrcke, et al. 2018; Slade 2012). Similarly, we did not control for any diseases that were likely consequences of diabetes, such as heart disease or other micro- and macro-vascular complications (World Health Organization 2016). Nonetheless, we carried out a robustness analysis where we controlled for the level of urbanization, the level of education, the state of residence, marital status, the number of children below the age of six in the household ~~, a quadratic age term and calendar year dummies. We also~~

control for and household wealth approximated by a household asset index. We composed an indicator The household asset index was created using principal component analysis of household assets and housing following Filmer et al. (2001) Filmer and Pritchett (2001). The asset index reflected owning a vehicle, a second house, a washing machine, dryer, stove, refrigerator or furniture, any electric appliances, any domestic appliances, a bicycle, farm animals, and accounted for the physical condition of the house, proxied by the type of floor material and water access. In our main regression models we did not account for . While part of the effect of diabetes may be due to potential adverse effects of obesity , including an additional robustness check we also controlled for obesity by including an indicator that was one for a BMI as a control variable in the model would have led to attenuated estimates if the diagnosis of diabetes also had an effect on ≥ 30 and zero for a BMI , which has been shown to be the case in other studies . Similarly, we did not control for any diseases that were likely consequences of diabetes, such as heart disease or other micro- and macro-vascular complications , as this would prevent any causal interpretation of the relationship of diabetes with our labour market outcomes . Nonetheless, we carried out a robustness analysis controlling for obesity< 30. Stata 15 was used for all analyses (StataCorp 2017).

3.1. Labour outcomes and time since diagnosis

The chronic nature and irreversibility of diabetes provide good reason to explore the long term effects associations post diagnosis. To do this we replaced the binary diabetes indicator of Eq 1 with a continuous variable indicating years since the diagnosis was first reported. Simultaneous inclusion of year dummies and time since diagnosis (which varies by one unit in each time period) would typically not allow separate identification of the coefficient Further, to allow for non-linear relationships over time we also estimated a model where instead of the linear years since diagnosis variable we used a spline function $g(Dyears_{it})$. The simultaneous inclusion of variables that increase at the same rate between survey

years in a FE model is not possible due to perfect collinearity. In our case this caused the problem of identifying the effect of time since diagnosis ~~. In this case, identification relied on the presence of people without diabetes in the sample, for which diabetes duration did not increase~~separate from the effect of other linear time trends such as age or the year of the survey. To deal with this problem, we opted for the estimation of an interaction effect of the time since diagnosis at baseline with the survey year. This provided us with an estimate of the association of each additional year since diagnosis with the respective labour outcome independent of the linear time trend.

~~We also considered a spline function that allowed for non-linear effects over time, with~~ The spline function took the form $g(Dyears_{it}) = \sum_{n=1}^{N} \delta_n \cdot max\{Dyears_{it} - \eta_{n-1}\}I_{in}$ and $I_{in} = 1[\eta_{n-1} \leq Dyears_{it} < \eta_n]$, with η_n being the place of the n-th node for $n = 1, 2, \ldots, N$. The coefficient δ_n captured the effect of diabetes for the n-th interval. The effects are linear if $\delta_1 = \delta_2 =, \ldots, = \delta_n$.

Based on visual inspection (Fig. 1~~on page 1~~) we chose three nodes located at 3, 7 and 12 years after diagnosis. The first three years should capture any immediate ~~effects of~~ associations with the diagnosis, the years four to seven any ~~effects~~ associations during time of adaptation to the disease and the later terms the ~~long term effects~~associations after a longer time has passed. We also estimated a non-linear model using dummy variables for duration groups rather than splines, applying the same duration cut-offs. Because the year of diagnosis was only reported in the third wave, time since diagnosis for previous survey waves was not available for those who were not interviewed in the third round. Also, using the years of diabetes at baseline for the interaction effect excluded everybody that only received a diabetes diagnosis after entering the sample. A reported diagnosis in the year of the interview was counted as 'one year since diagnosis'.

3.2. Labour outcomes and biometrically measured diabetes

The biomarker analysis consisted of three steps. We first re-estimated Eq ~~2~~ 1 to assess the relationship between self-reported diabetes ~~with~~ and labour outcomes, but this time for the cross-sectional biomarker sample only, using the following specification:

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 X_i + c_i + v_i \tag{2}$$

where v_i were community fixed effects, which reflected local unobserved characteristics, such as access to healthcare, poverty and unemployment in the community. Communities (or *localidades* in Spanish) are the smallest administrative units (nested within municipalities) recognized by the Mexican Institute for Statistics and Geography (INEGI). We did not use household fixed effects since the average number of observations per household was close to one.

In a second step, we estimated the ~~relations between biomarker diabetes and~~ association of biomarker diabetes with labour outcomes, using the following equation:

$$Y_i = \beta_0 + \beta_1 Dbio^d + \beta_2 X_i + v_i + u_i, \tag{3}$$

where $Dbio^d$ was equal to 1 if HbA1c $\geq 6.5\%$.

To estimate the ~~effect~~ association of undiagnosed diabetes with our outcomes, we added self-reported diabetes back into the equation (Eq 4).

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i + v_i + u_i. \tag{4}$$

This changed the interpretation of $Dbio_i$, which now reflected the ~~effect on those with undiagnosed diabetes~~ association of undiagnosed diabetes with the outcomes, i.e. the respondents not self-reporting diabetes but with HbA1c levels equal to or above the threshold.

We further investigated ~~the effect of~~ how the severity of diabetes ~~on labour outcomes~~ ~~, replacing *Dbio* with $Dbio^c$,~~ was associated with labour outcomes for self-reported and undiagnosed diabetes, respectively (Eq 5). Therefore we created for both self-reported and undiagnosed cases a variable that was 0 for HbA1c $< 6.5\%$ and increased ~~by one for every percentage point increase in~~ continuously with HbA1c for those ~~with an $HbA1c \geq 6.5\%$ (Eq 5). This allowed us to investigate the effect of a one percentage point increase in levels for people with undiagnosed diabetes (β_2) as well as~~ $\geq 6.5\%$, by carrying out the following transformation: $\text{HbA1c} - 6.4$ for those with ~~self-reported diabetes (β_3)~~ $\text{HbA1c} \geq 6.5\%$.

$$Y_i = \beta_0 + \beta_1 Dsr DsrHbA1c_i + \beta_2 Dbio_i^c + \beta_3 Dsr_i * HbA1c DbioHbA1c_i + \beta_4 X_i + v_i + u_i. \quad (5)$$

4. Results

4.1. Labour outcomes and self-reported diabetes

The results of estimating Eq 1 in Table 2 indicated ~~statistically significant reductions in the probability~~ lower probabilities of employment for men and women with self-reported diabetes ~~. Employment probabilities were reduced by 5.4~~ of 7.7 p.p. for men and ~~5.9~~ 6.3 p.p. for women. There ~~was no significant relationship~~ were no statistically significant associations between diabetes and working hours or wages.

Dividing the diabetes population into early and late onset groups, men, and potentially also women, ~~saw their employment probabilities negatively affected by a diabetes onset later in life~~ with a later diabetes onset had lower employment probabilities (Supplementary Table S2). In particular ~~, womenwith~~ in women, also an early diabetes onset ~~experienced an adverse effect. For working hours , effects~~ was associated with lower employment probabilities. In particular for working hours but also for wages, the estimates were less precise ~~but may indicated increased working hours for men with an early diabetes~~

~~onset, while women reduced their work hours~~and quite large, indicating much remaining uncertainty. Finally, we found higher wages for women with an early diabetes onset ~~, but no effects~~ but no association for men.

To assess whether diabetes ~~affected~~ was associated with changes in the selection into different types of work, we investigated the role of diabetes for the probability of being in non-agricultural wage employment, agricultural employment or self-employment. We found a reduction in the probability to work in agriculture for women, but not for men (Table 3). Disaggregating the diabetes groups further according to their age showed that most statistically significant relationships were driven by the older-onset group (Supplementary Table S3). For male self-employment, diabetes increased the probabilities to be self-employed in the younger group, while it reduced the probabilities to be self-employed in the older-onset group.

We reestimated all regressions in this section including a binary control for obesity (BMI ≥ 30) (Supplementary Table S9 and Table S10) ~~. This led to a reduction in sample size due to the larger number of missing cases for BMI. Obesity itself did not appear to be an independent predictor of any labour market outcome. The estimates of the effect of diabetes remained similar for most outcomes. Only for male employment probabilities, the diabetes coefficient was no longer statistically significant at the 10 percent significance level. Estimating the original model without accounting for obesity, but using the sample with only non-missing BMI cases, showed similar changes in effects (results available on request)~~as well as other time-variant control variables that were excluded from the main analysis due to their unclear relationship with diabetes (Supplementary Table S6 and Supplementary Table S7). All estimates remained very similar.

4.2. Labour outcomes and time since diagnosis

Fig. 1 shows that the probability of employment for men steadily declined as time progressed, using a non-parametric kernel-weighted local polynomial regression. For women,

a first drop-off occurred right after diagnosis; though no consistent pattern emerged thereafter. The dynamics for working hours and wages were less clear, with a possibly long term negative trend for women but not for men.

Table 4 panel A shows the results of estimating ~~Eq ??, which indicated that male and~~ the association of an additional years since diagnosis with labour market outcomes. They indicate a reduction in male but not female employment probabilities ~~fell every year , with a larger effect shown by the within coefficient.~~

~~Wages were reduced for womenas diabetes progressed using the linear specification~~with every year since the diagnosis. We also found some indication that diabetes duration was associated with a reduction in wages for women. Using diabetes onset groups, there was ~~no~~ only evidence of an ~~effect~~ association of diabetes duration ~~for early onset groups~~ with employment for men in the late-onset group (see Supplementary Table S4). ~~Unfortunately the very limited number of diabetes cases in the~~ For monthly working hours the results indicate a negative association with early-onset ~~group prohibited the estimation of effects of early diabetesonset durationon wagesand working hours~~diabetes, but not with late-onset diabetes duration. Further, we found a large positive association of early-onset diabetes with female wages, but also a negative association of female wages with late-onset diabetes.

The non-linear results for the spline function and dummy variable approach are presented in panels B and C, respectively. ~~They suggest that~~ The spline function results for the association of the time since diagnosis with employment probabilities were not very precisely measured for both men and women, only providing suggestive evidence that male employment chances were lower during the first three years and eight to twelve years after diagnosis. Further, results for wages indicate reductions in men with the longest time since diagnosis and in women immediately after diagnosis. For working hours we found higher working hours in women with the longest time since diagnosis. The dummy variable models suggest lower employment probabilities for men immediately after diagnosis and 13 years after diagnosis. For women, a negative association was found after eight to twelve years

after diagnosis. Contrary to the splines model, the ~~main adverse effects appeared after more than seven years since diagnosis. The same was true for female wages . Re-estimating the specifications with a random effects modelshowed that, at least for the models with dummies, an immediate reduction in employment probabilities that became stronger the longer a person had diabetes (see Supplementary Table ??)~~dummy model did not suggest higher female working hours 13 years after diagnosis. However, similar to the splines model we found lower wages for men and women with diabetes, in particular in the group with the longest time since diagnosis. Note that we did not estimate models splitting diabetes ~~in early and late onset~~ into early and late-onset groups, as this implied strong reductions in statistical power.

Controlling for other time-variant variables or additionally for obesity, results remained ~~very similar and obesity itself was not found to be affecting any labour market outcome~~ similar (Supplementary Table S8 and Supplementary Table S11).

4.3. Cross-sectional biomarker analysis

As reported in Supplementary Table S5, 18% of the observations in the biomarker sample were undiagnosed, which ~~accounts~~ accounted for 68% of all cases above the diabetes threshold. Comparing the health status and diabetes risk factors of the diagnosed and undiagnosed diabetes populations ~~suggested~~ suggests that those with self-reported diabetes were older and in worse health, both objectively and subjectively, compared to those undiagnosed.

Table 6 presents the results from ~~estimating~~ investigating relationships of self-reported diabetes, diabetes as defined by HbA1c, undiagnosed diabetes and diabetes severity with labour market outcomes (see Eq. 2 – 5). Overall we did not find evidence for any associations of diabetes with working hours or wages, so that in the following we only focus on employment probabilities. Panel A confirms the earlier longitudinal results using self-reported diabetes for the cross-sectional biomarker sample. The results in panel B in-

dicate that the relationship with employment became weaker when ~~using diabetes defined by the biomarker~~ diabetes was defined by HbA1c levels instead of self-reported diabetes, in particular for men. Results in Panel C ~~were obtained from estimating Eq. 4 and~~ indicate an absence of a (statistically significant) negative relationship between undiagnosed diabetes and labour outcomes.

~~To explore whether the adverse effects increased with higher levels, we estimated Eq. 5.~~ The results in panel D ~~only show a borderline statistically significant~~ show an adverse association of ~~0.9 percentage points per percentage point increase in~~ employment probabilities with increasing HbA1c ~~for female employment probabilities~~ levels of 1.2 p.p. lwith employment chances of women with diagnosed diabetes. However, a lot of uncertainty remained around this estimate as indicated by the relatively large standard errors. For undiagnosed diabetes, we did not find any association.

5. Discussion

Diabetes is now one of the most common chronic diseases in LMICs, as well as high-income countries (HICs), with ~~potential~~ severe impacts on the health and economic well-being of those affected. Yet rigorous evidence on the economic consequences for LMICs remains scarce.

To address key methodological challenges, this paper used rich longitudinal panel data from Mexico that also contained diabetes biomarkers. The biomarker data showed alarming levels of clinically tested diabetes (27% prevalence) and indicate that a large proportion of the Mexican population (18%) has undiagnosed diabetes.

The paper provided evidence for ~~adverse effects~~ an adverse association of self-reported diabetes ~~on~~ with employment, working hours and wages. While earlier work showed evidence for Mexico for employment (Seuring, Goryakin, et al. 2015), this paper presented, by our knowledge, the first evidence on the relationship of diabetes, working hours and

wages. Furthermore, we added to the study of Seuring, Goryakin, et al. (2015) by using longitudinal instead of cross-sectional data~~to identify a causal relationship~~. We provided first evidence of the ~~long term impact of diabetes~~ association of diabetes with labour market outcomes over the longer term in Mexico and explored the ~~extent and effects~~ role of undiagnosed diabetes. We confirmed ~~the findings of~~ earlier findings for Mexico by Seuring, Goryakin, et al. (2015) insofar that we found an adverse ~~effect of diabetes on~~ association of diabetes with male employment. We further showed more conclusive evidence that also ~~women experience a reduction in their employment probabilitiesdue to diabetes~~for women diabetes is associated with lower employment probabilities. Taking into account the general differences in employment between men and women, the found ~~effects translate into relative reductions of their~~ associations translate into relatively lower employment probabilities of ~~6~~almost 9% for men and ~~14~~17% for women with diabetes. We also found that the ~~effects~~ associations were mainly driven by those with a diabetes onset at a relatively later state, consisting of older people with most likely type 2 diabetes. This ~~was also found by~~ is similar to the findings of Seuring, Goryakin, et al. (2015) in their stratified analysis of an older and younger age group. Analyses of the long term impact indicated that the employment ~~probability~~ probabilities fell gradually in the years following diagnosis, albeit only for men. The results using ~~a~~ non-linear ~~model~~ models were less clear, potentially due to reductions in statistical power~~, but~~. They suggested that adverse ~~effects became stronger with time since diagnosis . The linear effect~~ associations with employment probabilities and also wages appeared in particular immediately after the diabetes diagnosis and then again after a considerable time of living with the disease. The significant linear association found in this analysis contrasts with estimates for the USA, where such an effect was absent; however allowing for non-linearity, revealed falling employment probabilities after 11 to 15 years for females and after 2-5 years for males (Minor 2013).

Overall, a relationship of diabetes and working hours or wages was mostly absent.

Although any explanation at this point is speculative, it may be that higher paid and more educated individuals were able to remain employed without experiencing wage reductions, for instance due to their particular set of skills. They may also have had access to better health care leading to better diabetes related health outcomes. Low paid workers, on the other hand, may have lacked access to quality diabetes care, making it more likely that they developed severe complications earlier (Flores-Hernández et al. 2015). They may also have been more likely to be in informal employment and low skilled jobs with less job security, and thus more prone to being laid off and replaced with healthier workers.

We found that self-reported diabetes cases were not representative of the entire diabetes population in Mexico. A large share of people with diabetes were undiagnosed and significantly healthier and younger, suggesting a selection into the diagnosed group based on the severity and true duration of diabetes. Consequently, diabetes as defined by the HbA1c threshold, was less related to reduced employment probabilities compared to self-reported diabetes. Further analysis showed that this was due to the absence of an association between undiagnosed diabetes and employment. These results are similar to those found for the USA, where ~~no~~ a statistically significant relationship was ~~observed between undiagnosed diabetes on employment, while a significant effect of diagnosed diabetes was observed~~ only observed between diagnosed diabetes and employment, but not between undiagnosed diabetes and employment (Minor and MacEwan 2016). Our results further indicated that the ~~difference in employment effects~~ association of diagnosed and undiagnosed diabetes ~~was not mediated~~ with employment status was not necessarily related with disease severity as proxied by current HbA1c levels. This is similar to findings for Mexican-Americans in the USA, where employment outcomes were unrelated to higher HbA1c levels (Brown, Perez, et al. 2011). A possible explanation may be that HbA1c levels are primarily informative for the last three months, and are not the only ~~or~~ nor best indicator for the severity of diabetes. Overall, it seems that a longer diabetes duration with its related health consequences, and selection into the diagnosed population based on

emerging diabetes related health problems, could have been driving the ~~adverse economic effects among those with~~ found adverse associations between a self-reported diagnosis and employment probabilities.

Our study had several limitations. While our model accounted for any time-invariant confounding, the estimates may have been affected by unobserved time-variant confounders. Reverse causality, where employment status affects the propensity to develop or be diagnosed with diabetes, may also have played a role. Existing studies that looked at this particular direction of causality, however, have not found strong evidence for an effect of employment status on diabetes (Bergemann et al. 2011; Schaller and Stevens 2015), though they were carried out in HICs. We did not control for the effects of obesity, hypertension, self-reported health or other diseases in our models due to the high probability that they were affected by diabetes themselves, which would have made a causal interpretation of the estimates more difficult. Robustness checks including obesity and other time-variant control variables indicated that our main findings remained mostly unchanged~~and that any occurring changes were likely the result of a smaller sample size rather than of~~, indicating that the main results are robust to the inclusion of ~~obesity. For~~ additional time-variant variables. A limitation of the duration analysis~~an additional limitation~~, imposed by the data, was that the year of diagnosis was only reported in the third wave. While this still allowed us to construct an estimate of the time since diagnosis for the previous waves, it restricted the analysis to those that were present in the last wave. Further, the strategy to interact the time since diagnosis at baseline with a linear time trend led to the exclusion of diabetes cases that had not yet received a diagnosis when they joined the survey. The results of the duration analysis are therefore not directly comparable to those using a binary diabetes indicator. Finally, we used a WHO recommended HbA1c cut-off to diagnose diabetes, due to the lack of a Mexico specific cut-off. There is some evidence that HbA1c may be affected by ethnicity (Sacks 2011). Hence, if Mexican ethnicity would lead to different HbA1c levels, the use of our cut-off could have led to misclassification based

on the used biomarkers. Finally, the analysis using early- and late diabetes onset groups may suffer from low statistical power in the early onset group, due to a low prevalence of diabetes in this group, making these estimates less reliable.

Despite these limitations, our findings bear important implications. First, the ~~impact~~ association of self-reported diabetes on labor outcomes in Mexico seemed mostly limited to its ~~effect on~~ relationship with lower employment probabilities. Second, its effect on employment was much stronger for females, though the underlying reasons for this remain unclear. Potential explanations are that lower working hours or wages for women make a dropout less costly. Other evidence suggests that women with diabetes are in worse metabolic health compared to men when they cross the diabetes threshold (Peters et al. 2015), making it more likely for them to drop out. Third, caution is needed when estimates based on self-reported diabetes are interpreted in terms of the entire population, i.e. extending to those with undiagnosed diabetes. Ideally, studies would include a biomarker analysis, acknowledge the differences between diagnosed and undiagnosed subpopulations, and carry out a separate analysis whenever feasible. If this is not possible, study conclusions about the effects of self-reported diabetes should be limited to this specific part of the population. This is of particular importance in LMICs where the share of undiagnosed diabetes is often high.

The large proportion of previously undiagnosed cases found in this paper indicates that diagnosis—at least in Mexico—happens late or not at all. This may reduce the possibilities to prevent complications via treatment and self-management, increasing the risk of severe complications appearing premature. Earlier diagnosis and ensuing effective treatment may lighten the health and economic burden. Therefore, more research is needed to investigate the economic impact of diabetes over time. Longitudinal biomarker information could be used to observe the true duration and severity of diabetes as well as the time that passes till a medical diagnosis. This would allow for a better understanding of when adverse economic effects start to appear. Further, future research should investigate how time of

diagnosis and treatment of diabetes affect the occurrence of adverse labour market effects of diabetes. The results of such research could allow costing studies to include more detailed information on the indirect costs of diabetes; or inform cost-effectiveness analyses that aim to include a measure of the potential benefit of the intervention to employers or society at large.

References

J. Alegre-Díaz, W. Herrington, M. López-Cervantes, L. Gnatiuc, R. Ramirez, M. Hill, C. Baigent, M. I. McCarthy, S. Lewington, R. Collins, G. Whitlock, R. Tapia-Conyer, R. Peto, P. Kuri-Morales, and J. R. Emberson (2016). "Diabetes and Cause-Specific Mortality in Mexico City". In: *New England Journal of Medicine* 375.20, pp. 1961–1971. DOI: 10.1056/NEJMoa1605368.

J. Angrist and J. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

P. Ayyagari, D. Grossman, and F. Sloan (2011). "Education and health: evidence on adults with diabetes". In: *International Journal of Health Care Finance and Economics* 11.1, pp. 35–54. DOI: 10.1007/s10754-010-9087-x.

S. Barquera, I. Campos-Nonato, C. Aguilar-Salinas, R. Lopez-Ridaura, A. Arredondo, and J. Rivera-Dommarco (2013). "Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy". In: *Globalization and Health* 9.1, p. 3. DOI: 10.1186/1744-8603-9-3.

S. Barquera, L. Hernandez-Barrera, M. L. Tolentino, J. Espinosa, S. W. Ng, J. A. Rivera, and B. M. Popkin (2008). "Energy Intake from Beverages Is Increasing among Mexican Adolescents and Adults". In: *Journal of Nutrition* 138.12, pp. 2454–2461. DOI: 10.3945/jn.108.092163.

S. Basu, P. Yoffe, N. Hills, and R. H. Lustig (2013). "The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data". In: *PLoS ONE* 8.2, e57873. DOI: 10.1371/journal.pone.0057873.

J. Beagley, L. Guariguata, C. Weil, and A. a. Motala (2014). "Global estimates of undiagnosed diabetes in adults". In: *Diabetes Research and Clinical Practice* 103.2, pp. 150–160. DOI: 10.1016/j.diabres.2013.11.001.

O. Y. Bello-Chavolla, R. Rojas-Martinez, C. A. Aguilar-Salinas, and M. Hernández-Avila (2017). "Epidemiology of diabetes mellitus in Mexico". In: *Nutrition Reviews* 75.suppl 1, pp. 4–12. DOI: 10.1093/nutrit/nuw030.

A. Bergemann, E. Grönqvist, and S. Gudbjörnsdottir (2011). "The effects of job displacement on the onset and progression". In: *Netspar Discussion Paper* 25, pp. 1–37.

H. S. Brown, J. A. Pagán, and E. Bastida (2005). "The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit". In: *Health Economics* 14.5, pp. 537–544. DOI: 10.1002/hec.942.

H. S. Brown, A. Perez, L. M. Yarnell, J. a. Pagan, C. L. Hanis, S. P. Fischer-Hoch, and J. B. McCormick (2011). "Diabetes and employment productivity: does diabetes management matter?" In: *American Journal of Managed Care* 17.8, pp. 569–576.

T. T. Brown (2014). "How effective are public health departments at preventing mortality?" In: *Economics & Human Biology* 13, pp. 34–45. DOI: 10.1016/j.ehb.2013.10.001.

E. Crimmins, T. McDade, L. Rubalcava, T. Seeman, G. Teruel, and D. Thomas (2015). *Health of the Mexican population: Results from the Mexican Family Life Survey (MxFLS)*. Tech. rep.

J. Currie and T. Vogl (2013). "Early-Life Health and Adult Circumstance in Developing Countries". In: *Annual Review of Economics* 5.1, pp. 1–36. DOI: 10.1146/annurev-economics-081412-103704.

N. De Fine Olivarius, V. D. Siersma, R. Køster-Rasmussen, B. L. Heitmann, and F. B. Waldorff (2015). "Weight changes following the diagnosis of type 2 diabetes: The impact of recent and past weight history before diagnosis. Results from the Danish Diabetes Care in General Practice (DCGP) Study". In: *PLoS ONE* 10.4, pp. 1–14. DOI: 10.1371/journal.pone.0122219.

R. van Ewijk (2011). "Long-Term Health Effects on the Next Generation of Ramadan Fasting during Pregnancy". In: *Journal of Health Economics* 30.6, pp. 1246–1260.

D. Filmer and L. Pritchett (2001). "Estimating wealth effects without expenditure data-Or tears: An application to educational enrollments in states of India". In: *Demography* 38.1, pp. 115–132.

S. Flores-Hernández, P. J. Saturno-Hernández, H. Reyes-Morales, T. Barrientos-Gutiérrez, S. Villalpando, and M. Hernández-Ávila (2015). "Quality of Diabetes Care: The Challenges of an Increasing Epidemic in Mexico. Results from Two National Health Surveys (2006 and 2012)". In: *Plos One* 10.7, e0133958. DOI: 10.1371/journal.pone.0133958.

E. W. Gregg, H. Chen, L. E. Wagenknecht, J. M. Clark, L. M. Delahanty, J. Bantle, H. J. Pownall, K. C. Johnson, M. M. Safford, A. E. Kitabchi, F. X. Pi-Sunyer, R. R. Wing, and A. G. Bertoni (2012). "Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes". In: *Journal of the American Medical Association* 308.23, p. 2489. DOI: 10.1001/jama.2012.67929.

W. G. Herrington, J. Alegre-Díaz, R. Wade, L. Gnatiuc, R. Ramirez-Reyes, M. Hill, M. Solano-Sánchez, C. Baigent, S. Lewington, R. Collins, R. Tapia-Conyer, R. Peto, P. Kuri-Morales, and J. R. Emberson (2018). "Effect of diabetes duration and glycaemic control on 14-year cause-specific mortality in Mexican adults: a blood-based prospective cohort study". In: *The Lancet Diabetes & Endocrinology* 8587.18, pp. 1–9. DOI: 10.1016/S2213-8587(18)30050-0.

International Diabetes Federation (2015). *Diabetes Atlas*. Tech. rep. International Diabetes Federation.

E. Latif (2009). "The impact of diabetes on employment in Canada". In: *Health Economics* 18.5, pp. 577–589. DOI: 10.1002/hec.1390.

Y. Li, Y. He, L. Qi, V. W. Jaddoe, E. J. M. Feskens, X. Yang, G. Ma, and F. B. Hu (2010). "Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood". In: *Diabetes* 59.10, pp. 2400–2406. DOI: 10.2337/db10-0385.

E. L. Lim, K. G. Hollingsworth, B. S. Aribisala, M. J. Chen, J. C. Mathers, and R. Taylor (2011). "Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol". In: *Diabetologia* 54.10, pp. 2506–2514. DOI: 10.1007/s00125-011-2204-7.

X. Liu and C. Zhu (2014). "Will knowing diabetes affect labor income? Evidence from a natural experiment". In: *Economics Letters* 124.1, pp. 74–78. DOI: 10.1016/j.econlet.2014.04.019.

T. Minor (2011). "The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey". In: *Health Economics* 20.12, pp. 1468–1486. DOI: 10.1002/hec.1685.

— (2013). "An investigation into the effect of type I and type II diabetes duration on employment and wages". In: *Economics & Human Biology* 11.4, pp. 534–544. DOI: 10.1016/j.ehb.2013.04.004.

T. Minor and J. P. MacEwan (2016). "A comparison of diagnosed and undiagnosed diabetes patients and labor supply". In: *Economics & Human Biology* 20, pp. 14–25. DOI: 10.1016/j.ehb.2015.10.003.

S. A. E. Peters, R. R. Huxley, N. Sattar, and M. Woodward (2015). "Sex Differences in the Excess Risk of Cardiovascular Diseases Associated with Type 2 Diabetes: Potential Explanations and Clinical Implications". In: *Current Cardiovascular Risk Reports* 9.7, pp. 1–7. DOI: 10.1007/s12170-015-0462-5.

N. Reynoso-Noverón, R. Mehta, P. Almeda-Valdes, R. Rojas-Martinez, S. Villalpando, M. Hernández-Ávila, and C. a. Aguilar-Salinas (2011). "Estimated incidence of cardiovas-

cular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey". In: *Cardiovascular Diabetology* 10.1, p. 1. DOI: 10.1186/1475-2840-10-1.

L. Rubalcava and G. Teruel (2013). *User's Guide for the Mexican Family Life Survey Third Round*. Tech. rep.

D. B. Sacks (2011). "A1C Versus Glucose Testing: A Comparison". In: *Diabetes Care* 34.2, pp. 518–523. DOI: 10.2337/dc10-1546.

J. Schaller and A. H. Stevens (2015). "Short-run effects of job loss on health conditions, health insurance, and health care utilization". In: *Journal of Health Economics* 43, pp. 190–203. DOI: 10.1016/j.jhealeco.2015.07.003.

T. Seuring, O. Archangelidi, and M. Suhrcke (2015). "The Economic Costs of Type 2 Diabetes: A Global Systematic Review". In: *PharmacoEconomics* 33.8, pp. 811–831. DOI: 10.1007/s40273-015-0268-9.

T. Seuring, Y. Goryakin, and M. Suhrcke (2015). "The impact of diabetes on employment in Mexico". In: *Economics & Human Biology* 18, pp. 85–100. DOI: 10.1016/j.ehb.2015.04.002.

T. Seuring, M. Suhrcke, P. Serneels, and M. Bachmann (2018). "Diabetes, Employment and Behavioural Risk Factors in China : Marginal Structural Models versus Fixed Effects Models Diabetes , Employment and Behavioural Risk Factors in China : Marginal Structural Models versus Fixed Effects Models". In: *IZA Discussion Papers* 11817.

A. N. Slade (2012). "Health Investment Decisions in Response to Diabetes Information in Older Americans". In: *Journal of Health Economics* 31.3, pp. 502–520.

O. Sotomayor (2013). "Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes". In: *Economics & Human Biology* 11.3, pp. 281–293. DOI: 10.1016/j.ehb.2012.02.009.

StataCorp (2017). *Stata Statistical Software: Release 15*. College Station, TX.

A. L. Williams et al. (2013). "Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico". In: *Nature* 506.7486, pp. 97–101. DOI: 10.1038/nature12828.

World Health Organization (2011). *Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation*. Geneva: World Health Organization.

— (2016). *Global Report on Diabetes*. Geneva: World Health Organization.

X. Yuan, T. Liu, L. Wu, Z.-Y. Zou, and C. Li (2015). "Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study". In: *British Medical Journal Open* 5.4, e006633–e006633. DOI: 10.1136/bmjopen-2014-006633.

A. Zajacova, J. Dowd, R. F. Schoeni, and R. B. Wallace (2010). "Consistency and precision of cancer reporting in a multiwave national panel survey". In: *Population Health Metrics* 8.1, p. 20. DOI: 10.1186/1478-7954-8-20.

Supplementary material

Strategies to deal with inconsistent self-reporting over time

Reporting error can pose a considerable challenge in the use of self-reported data. Fortunately, the MxFLS data provide several possibilities to assess the amount of misreporting and apply corrections before estimating the labour market effects of diabetes. In what follows we describe how we have dealt with inconsistencies in self-reported diabetes over time.

Throughout the surveys, self-reported diabetes was measured by the question 'Have you ever been diagnosed by diabetes'. One of the key advantages of panel data is the repeated measurement which results in more than one data point, allowing to uncover inconsistencies for cases with multiple observations. Very little is known about inconsistencies in self-reported diabetes over time. Zajacova et al. (2010) assess the consistency of a self-reported cancer diagnosis over time in the USA. The study found that 30% of those who had reported a cancer diagnosis at an earlier point, failed to report the diagnosis at a later point in time. A more recent diagnosis was found to be reported with greater consistency, possibly due to increasing recall problems as time since diagnosis advanced.

When assessing the MxFLS, we also found inconsistencies in the diabetes self-reports across the three waves, with between 10–20% of those reporting diabetes in one wave not doing so in one of the subsequent waves. To improve the validity of diabetes self-reports, we were interested in reducing the amount of reporting inconsistencies.

For diabetes, the main concern with mismeasurement is related to a lack of a diagnosis. Wrong self-reports indicating a diagnosis of diabetes we deemed less of a problem since incentives to falsely report a diabetes diagnosis seem to be very limited—although we cannot exclude this. A study from China found that the vast majority (98%) of those who self-reported diabetes were tested positive for diabetes, while only a minority of those who were tested positive for diabetes (40%) actually self-reported the disease (Yuan et al.

2015). Our data showed a similar pattern, with a low proportion (2%) of the respondents being tested negative while self-reporting diabetes, while the majority of those who were tested positive (68%) did not self-report diabetes.

We used the above information to infer the "true" diabetes status for those with inconsistent reports. For respondents present in all three waves, we corrected inconsistencies as reported in Supplementary Table S1. We assumed that if diabetes was reported only once in the first two waves (either in 2002 or 2005) and then not reported again in the ensuing waves, this diabetes report was likely to be false (see lines 3 and 4 in Supplementary Table S1) and that the person never had received a diagnosis. If a diabetes diagnosis was reported in two of the three waves (in 2002 and 2009 but not 2005, or in 2002 and 2005 but not in 2009), we assumed that the respondent had diabetes in all three waves (see lines 1 and 2 in Supplementary Table S1). For cases where we only had information from two waves, we assumed that if a diabetes diagnosis had been reported in a prior wave they also had diabetes in the ensuing wave, even if it was not reported in the latter (see lines 5 and 6 in Supplementary Table S1), given that most diabetes self-reports tend to be correct.

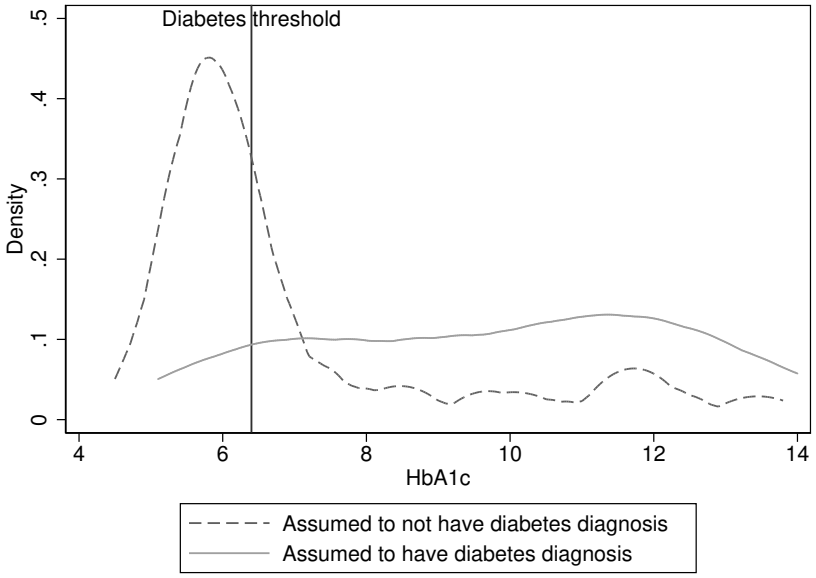
Table S1. Inconsistencies in diabetes self-report in MxFLS.

	Inconsistency	Assumption	Number of observations replaced
1	Diabetes self-report only in 2002, but not in 2005 and 2009	Has no diabetes in 2002 either	66
2	Diabetes self-report only in 2005, but not in 2002 and 2009	Has no diabetes in 2005 either	52
3	Diabetes self-report in 2002, 2005 but not in 2009	Has diabetes in 2009 as well	19
4	Diabetes self-report in 2002, 2009 but not in 2005	Has diabetes in 2005 as well	63
5	Diabetes self-report in 2002, but not in 2005. Not in survey in 2009	Has diabetes in 2005 as well	44
6	Diabetes self-report in 2005, but not in 2009. Not in survey in 2002	Has diabetes in 2009 as well	23

We then tested if the respondents we categorized as not having a diabetes diagnosis based on above rules, were actually more likely to not have biometrically measured diabetes, using the biomarker data from wave 3. Of those with inconsistencies in their diabetes self-reports, 95 were present in the biomarker sample (46 with two self-reports (from lines 3 and 4 in Table S1) and 49 with one self-report of diabetes (from lines 1 and 2 in Supplementary Table S1)). Supplementary Figure S1 illustrates the difference between

both groups and suggests that indeed those with two self-reports of diabetes were much more likely to have HbA1c values above the diabetes threshold. A t-test comparing the mean HbA1c for the two groups indicated that those with two self-reports also had significantly ($p<0.001$) higher HbA1c levels than those with only one self-report of diabetes (9.7% vs. 7.1%). Further, of those with one self-report, only 30% had an HbA1c$\geq$ 6.5% compared to 87% of those with two self-reports. Based on these results it appears that we did minimize misclassification of people into diabetes or no diabetes.

Figure S1: Kernel density of HbA1c values for those with one inconsistent and two inconsistent reports.



~~Early~~ Early- versus ~~late onset~~ late-onset of diabetes

Table S2. **Labour outcomes and self-reported diabetes by diabetes onset.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
~~Early onset~~Early-onset	~~0.133~~ 0.134	~~-0.206~~-0.195**	~~14.856~~14.395*	~~-18.250~~-18.665*	~~-0.490~~ -0.513*	~~0.375~~0.362***
	(0.176)	(0.086)	(~~8.329~~8.377)	(~~9.515~~9.650)	(~~0.337~~0.311)	(~~0.057~~0.039)
~~Late onset~~Late-onset	~~-0.059~~-0.082***	~~-0.047~~-0.053**	~~-0.936~~ -1.360	~~-1.346~~ -1.267	~~0.071~~ 0.016	~~0.074~~ 0.059
	(~~0.026~~0.025)	(0.025)	(~~1.510~~1.500)	(~~2.553~~2.565)	(~~0.068~~0.067)	(~~0.161~~0.165)
N	21388	27339	~~13828~~ 17618	~~7068~~ 9115	~~17616~~ 13830	~~9112~~ 7070

Notes Robust standard errors in parentheses. All models include ~~variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year~~ calendar year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S3. **Selection into types of work and self-reported diabetes by diabetes onset.**

	Non-agric.		Agriculture		Self-employed	
	Males	Females	Males	Females	Males	Females
~~Early-onset~~ Early-onset	~~0.030~~ 0.036	−0.105	~~-0.226~~ -0.233*	~~-0.068~~ -0.066	0.328**	~~-0.027~~ -0.020
	(~~0.216~~0.215)	(0.074)	(0.139)	(0.047)	(0.161)	(~~0.048~~0.049)
~~Late-onset~~ Late-onset	~~-0.007~~ -0.024	~~0.007~~ 0.006	~~-0.001~~ -0.008	~~-0.018~~ -0.019**	~~-0.054~~ -0.056**	~~-0.029~~ -0.033*
	(0.029)	(0.019)	(0.022)	(0.009)	(0.026)	(0.019)
N	~~20719~~ 20537	~~26575~~ 26478	~~20719~~ 20537	~~26575~~ 26478	~~20719~~ 20537	~~26575~~ 26478

Notes Robust standard errors in parentheses. All models include ~~variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year~~ calendar year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S4. **Relationship between self-reported years since diagnosis and employment probabilities using continuous duration by diabetes onset.**

	Employment		Monthly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
~~Early onset~~Survey year	~~-0.005~~ 0.003***	~~0.007~~ 0.004***	-0.019	0.181**	0.014***	0.019***
	(~~0.017~~0.001)	(~~0.013~~0.001)	(0.050)	(0.080)	(0.003)	(0.004)
~~Late onset~~Years since diagnosis at baseline (early-onset) × Survey year	~~-0.012~~ 0.003	0.001	0.197	-2.924***	−0.008	0.263***
	(0.003)	(0.001)	(0.234)	(0.040)	(0.019)	(0.002)
Years since diagnosis at baseline (late-onset) × Survey year	-0.003***	-0.001	0.007	0.042	-0.004	-0.010***
	(0.001)	(0.001)	(0.062)	(0.104)	(0.003)	(0.002)
N	20760	26313	17137	8863	13481	6885

Notes Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S5. **Number of observations with diabetes (HbA1c $\geq$ 6.5%) and self-reported diabetes.**

	$HbA1c < 6.5\%$	HbA1c $\geq$ 6.5%	Total
No self-reported diabetes (N)	4544	1181	5725
Row %	79%	21%	100%
Cell %	**71%**	**18%**	-
Self-reported diabetes (N)	129	554	683
Row %	19%	81%	100%
Cell %	**2%**	**9%**	-
Total (N)	4673	1735	6408

718 Robustness checks

719 Additional time-variant controls

Table S6. **Labour outcomes and self-reported diabetes including additional time-variant controls.**

	Employment		Monthly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Diabetes	-0.074***	-0.070***	-0.906	-2.170	0.021	0.071
	(~~0.005~~0.025)	(~~0.004~~0.024)	(~~0.275~~1.496)	(~~0.505~~2.515)	(~~0.013~~0.066)	(0.157)
N	21388	27339	17618	9115	13830	7070

Notes Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, $p < 0.01$.

Table S7. **Selection into types of work and self-reported diabetes including additional time-variant controls.**

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Diabetes	-0.018	-0.012	-0.048*	-0.007	-0.023***	-0.033*
	(0.029)	(0.022)	(0.026)	(0.018)	(0.009)	(0.018)
N	~~16308~~ 20537	~~22450~~ 20537	~~13592~~ 20537	~~7394~~ 26478	~~10778~~ 26478	~~5748~~ 26478

Notes ~~The effects of early onset diabetes on wages and working hours could not be estimated due to no within-variation for diabetes.~~ Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, ~~age squared~~ and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $p < 0.01$.

~~Random effects model~~

Additionally controlling for time-variant controls and obesity

Table S8. **Relationship between self-reported years since diagnosis and employment probabilities using continuous duration and duration splines including additional time-variant controls.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Survey year	0.002***	0.004***	-0.021	0.198**	0.014***	0.023***
	(0.001)	(0.001)	(0.051)	(0.089)	(0.003)	(0.005)
Years since diagnosis at baseline × Survey year	-0.007 -0.003***	-0.004 -0.001*	0.039 0.010	-0.130 0.030	0.010 0.004	−0.009***
	(0.002 0.001)	(0.001)	(0.102 0.062)	(0.127 0.103)	(0.005 0.003)	(0.003 0.003)
Panel B: splines						
Years since SR diagnosis Survey year	0.002***	0.004***	-0.017	0.198**	0.014***	0.024***
	(0.001)	(0.001)	(0.051)	(0.090)	(0.003)	(0.005)
Interaction: Years since diagnosis at baseline with survey year						
0–3	-0.008 -0.006*	-0.015 -0.000	-0.035 -0.048	0.507 0.249	0.038 -0.003	0.031 -0.033**
	(0.006 0.004)	(0.006 0.003)	(0.346 0.206)	(0.611 0.237)	(0.017 0.013)	(0.029 0.014)
4–7	0.001 -0.006	0.004 -0.003	0.242 -0.301	-0.570 -0.274	-0.032 -0.022	-0.018 -0.018
	(0.011 0.008)	(0.011 0.004)	(0.665 0.468)	(1.062 0.467)	(0.032 0.023)	(0.052 0.024)
8–12	-0.008 -0.015*	0.002 -0.003	-0.116 0.406	-0.080 -0.973	-0.003 0.047*	-0.074 0.021
	(0.015 0.009)	(0.011 0.005)	(0.855 0.518)	(1.098 0.724)	(0.041 0.027)	(0.050 0.040)
13+	-0.012 -0.003	-0.004 -0.001	0.035 0.185	-0.339 -2.080***	0.029 -0.026**	0.011 -0.043
	(0.008 0.002)	(0.003 0.001)	(0.119 0.212)	(0.241 0.713)	(0.018 0.012)	(0.017 0.037)
Panel C: dummies						
0–3 Survey year	-0.036 0.002***	-0.041 0.004***	-0.015	0.198**	-0.821 -0.014***	1.091 -0.024***
	0.134 (0.001)	(0.001)	(0.051)	(0.090)	(0.003)	(0.005)
Interaction: Years since diagnosis at baseline with survey year						
0–3	-0.024**	0.021 -0.009	-0.256	0.218	-0.033	-0.078*
	(0.021 0.011)	(0.021 0.009)	(1.154 0.667)	(1.826 0.618)	(0.054 0.040)	(0.083 0.044)
4–7	-0.014 -0.017	-0.056 0.001	0.877 -1.190	1.200 0.617	0.093 -0.034	-0.003 -0.072
	(0.022 0.016)	(0.023 0.009)	(1.375 0.769)	(2.530 1.177)	(0.059 0.047)	(0.118 0.047)
8–12	-0.069 -0.030	-0.043 -0.047***	0.427 -0.462	0.302 -3.305*	-0.070 -0.052	-0.148 -0.011
	(0.037 0.020)	(0.039 0.012)	(2.288 1.174)	(2.995 1.840)	(0.101 0.071)	(0.117 0.115)
13+	-0.121 -0.042**	-0.043 -0.010	-0.568 -1.757	-2.104 -1.305	0.242 -0.156**	-0.279 -0.131***
	(0.045 0.021)	(0.031 0.014)	(2.280 1.431)	(3.088 2.071)	(0.126 0.067)	(0.153 0.020)
N	16308 20760	22450 26313	13592 17137	7391 8863	10778 13481	5748 6885

Notes Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications. Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth and health insurance status, age squared and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $p < 0.01$.

Table S9. **Labour outcomes and self-reported diabetes controlling for obesity.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Obese (BMI ≥ 30)	0.007 0.009	-0.005 -0.002	-0.127 -0.025	-1.144 -1.079	0.018 0.027	0.082 0.081
	(0.012)	(0.013)	(0.773 0.771)	(1.188 1.184)	(0.038)	(0.061)
Diabetes	-0.046 -0.064 -0.065**	-0.689 -0.078***	-0.169 -1.097	0.036 -0.291	0.033 0.001	0.037
	(0.028)	(0.027)	(1.772 1.768)	(2.904 2.909)	(0.078 0.076)	(0.183 0.181)
N	17992	24145	14866 14867	7929 7931	11711 11712	6166 6167

Notes Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S10. **Selection into types of work and self-reported diabetes controlling for obesity.**

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Obese (BMI ≥ 30)	0.005 0.007	-0.032 -0.031**	0.036 0.035**	-0.021 -0.019*	0.003 0.002	0.010 0.011
	(0.017)	(0.013)	(0.014)	(0.011)	(0.004)	(0.009)
Diabetes	0.010 0.001	0.002 -0.001	-0.060 -0.067**	-0.011 -0.018	-0.020 -0.023**	-0.025 -0.030
	(0.033 0.034)	(0.023)	(0.028)	(0.020)	(0.010)	(0.021)
N	17414 17261	17414 17261	17414 17261	23458 23377	23458 23377	23458 23377

Notes Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S11. **Relationship between self-reported years since diagnosis and employment probabilities using continuous duration and duration splines and controlling for obesity.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Obese (BMI ≥ 30)	0.003 0.011	-0.010 -0.003	0.059 -0.249	-0.412 -0.914	0.026 0.024	0.035 0.058
	(0.012)	(0.014)	(0.831 0.791)	(1.247 1.222)	(0.040)	(0.064 0.063)
Survey year	0.002*	0.004***	-0.039	0.182*	0.014***	0.020***
	(0.001)	(0.001)	(0.060)	(0.103)	(0.003)	(0.006)
Years since diagnosis at baseline × Survey year	-0.019 0.004***	-0.008 -0.001*	0.259 0.039	-0.008 -0.023	-0.016 -0.004	-0.073 0.011**
	(0.006 0.001)	(0.006 0.001)	(0.375 0.064)	(0.731 0.092)	(0.019 0.003)	(0.031 0.005)
Panel B: splines						
Years since SR diagnosis						
Obese (BMI ≥ 30)	0.003 0.011	-0.009 -0.002	0.073 -0.221	-0.371 -0.898	0.027 0.025	0.036 0.055
	(0.013 0.012)	(0.014)	(0.832 0.785)	(1.247 1.225)	(0.040)	(0.064 0.063)
Survey year	0.002**	0.004***	-0.037	0.178*	0.014***	0.021***
	(0.001)	(0.001)	(0.060)	(0.103)	(0.003)	(0.006)
0–3	-0.014 -0.008*	-0.022 0.003	0.806 0.094	3.762 0.053	-0.070 -0.009	0.015 -0.033**
	(0.015 0.004)	(0.017 0.003)	(1.651 0.242)	(3.169 0.212)	(0.057 0.015)	(0.139 0.014)
4–7	-0.003 0.000	0.009 -0.008	-0.293 -0.488	-3.921 0.288	0.035 -0.002	-0.121 -0.015
	(0.018 0.009)	(0.015 0.005)	(0.914 0.527)	(1.811 0.434)	(0.044 0.024)	(0.108 0.026)
8–12	-0.023 -0.007	0.001 -0.003	-0.098 0.599	3.082 1.090	-0.062 0.017	-0.085 0.028
	(0.022 0.010)	(0.016 0.007)	(1.356 0.541)	(1.736 0.727)	(0.068 0.021)	(0.074 0.045)
13+	-0.038 0.003	-0.024 0.000	0.855 0.135	-1.128 -1.010	0.005 -0.017	-0.065 -0.066
	(0.017 0.005)	(0.012 0.003)	(1.029 0.179)	(1.421 1.271)	(0.053 0.011)	(0.063 0.051)
Panel C: dummies						
Obese (BMI ≥ 30) ≥ 30)	0.005 0.011	-0.009 -0.003	0.044 -0.236	-0.378 -0.892	0.026 0.024	0.031 0.055
	(0.012)	(0.014)	(0.831 0.787)	(1.245 1.223)	(0.040)	(0.064 0.063)
0–3 Survey year	0.028 0.002**	-0.032 0.004***	1.184 -0.035	22.434 0.181*	0.047 0.014***	-0.658 0.021***
	(0.059 0.001)	(0.065 0.001)	(3.825 0.060)	(11.579 0.104)	(0.212 0.003)	(0.622 0.006)
4–7 0–3	0.001 -0.029**	-0.054 -0.001	2.399 -0.148	12.909 -0.326	-0.047	-0.083
	(0.013)	-0.793 (0.011)	(0.817)	(0.527)	(0.048)	(0.052)
4–7	-0.034*	0.004	-0.467	0.928	-0.046	-0.080*
	(0.018 0.018)	(0.055 0.011)	(3.181 0.731)	(11.063 1.195)	(0.154 0.044)	(0.616 0.049)
8–12	-0.064 -0.048**	0.010 -0.062***	0.296 -0.831	15.604 -2.434	-0.293 0.043	-1.125 -0.000
	(0.069 0.020)	(0.066 0.014)	(4.994 1.426)	(11.038 2.455)	(0.247 0.050)	(0.583 0.112)
13+	-0.208 -0.032	-0.073 -0.018	-1.966 -2.731**	17.459 -1.609***	0.168 -0.132*	-1.090 -0.159***
	(0.105 0.027)	(0.081 0.017)	(4.975 1.270)	(10.262 0.507)	(0.256 0.075)	(0.499 0.016)
N	13012 17459	19972 23238	11622 14457	6487 7709	9262 11411	5054 6003

Notes Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications. Robust standard errors in parentheses. All models include variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tables

Table 1. **Descriptive statistics for the panel sample (2002,2005,2009).**

	Males			Females		
	No diabetes	Diabetes	p (t-test)	No diabetes	Diabetes	p (t-test)
Dependent variables						
Employed	0.87	0.80	0.00	0.37	0.26	0.00
Hourly wage (in Mexican Peso)	42.29	46.79	0.83	40.67	36.33	0.61
Weekly working hours	46.83	46.51	0.60	39.06	37.51	0.09
Non-agricultural worker or employee	0.51	0.41	0.00	0.24	0.13	0.00
Agricultural worker	0.19	0.13	0.00	0.02	0.01	0.00
Self-employed	0.16	0.26	0.00	0.09	0.11	0.04
Diabetes variables						
Diabetes duration (years)		~~7.40~~6.94			~~7.79~~7.09	
Control variables						
Age	35.31	50.68	0.00	35.37	50.45	0.00
Any medical insurance	0.47	0.59	0.00	0.50	0.62	0.00
City of 2,5oo-15,000	0.11	0.09	0.03	0.11	0.13	0.00
City of 15,000-100,000	0.10	0.14	0.00	0.10	0.10	0.40
City of >100,000	0.34	0.39	0.00	0.35	0.34	0.47
Married	0.53	0.77	0.00	0.53	0.66	0.00
Number of children (age<6) in household	1.49	1.14	0.00	1.60	1.13	0.00
Indigenous group	0.19	0.15	0.00	0.19	0.19	0.86
Education						
Secondary	0.31	0.22	0.00	0.31	0.16	0.00
High school	0.16	0.07	0.00	0.14	0.03	0.00
Higher education	0.11	0.12	0.39	0.10	0.03	0.00
Wealth index	0.00	0.04	0.27	−0.01	0.01	0.36
N	~~20391~~20394	994		~~25664~~25672	~~1666~~1667	

Notes Mean values. Diabetes refers to self-reported diabetes.

Table 2. **Labour outcomes and self-reported diabetes.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Diabetes	~~-0.054~~ 0.077***	~~-0.059~~ 0.063***	~~-0.506~~ 0.940	~~-1.998~~ 1.941	~~0.055~~ 0.001	~~0.081~~ 0.065
	(0.025)	(0.024)	(~~1.499~~1.489)	(~~2.511~~2.531)	(~~0.068~~0.066)	(~~0.158~~0.162)
N	21388	27339	~~17616~~ 17618	~~9112~~ 9115	~~13828~~ 13830	~~7068~~ 7070

Notes Robust standard errors in parentheses. All models include ~~variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year~~ calendar year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. **Selection into types of work and self-reported diabetes.**

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Diabetes	~~-0.006~~ -0.022	~~-0.008~~ -0.014	~~-0.043~~ -0.045*	−0.001	~~-0.022~~-0.023**	~~-0.029~~ -0.032*
	(0.029)	(0.022)	(0.026)	(0.018)	(0.009)	(0.018)
N	~~20719~~ 20537	~~20719~~ 20537	~~20719~~ 20537	~~26575~~ 26478	~~26575~~ 26478	~~26575~~ 26478

Notes Robust standard errors in parentheses. All models include ~~variables for states, urbanization, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year~~ calendar year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. **Relationship between self-reported years since diagnosis and employment probabilities using continuous duration and duration splines.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Survey year	0.003***	0.004***	-0.018	0.180**	0.014***	0.019***
	(0.001)	(0.001)	(0.050)	(0.080)	(0.003)	(0.004)
Years since diagnosis at baseline × Survey year	-0.016 0.003***	-0.009 -0.001	0.185 0.012	0.115 0.040	-0.016 −0.004	-0.067 0.010***
	(0.006 0.001)	(0.005 0.001)	(0.334 0.061)	(0.652 0.104)	(0.018 0.003)	(0.029 0.002)
Panel B: splines						
~~Years since SR diagnosis~~ Survey year	0.003***	0.004***	-0.014	0.179**	0.014***	0.020***
	(0.001)	(0.001)	(0.050)	(0.080)	(0.003)	(0.004)
Interaction: Years since diagnosis at baseline with survey year						
0–3	-0.013 0.006*	-0.018 -0.001	0.708 −0.038	2.952 0.262	-0.005 −0.004	0.047 0.032**
	(0.014 0.004)	(0.016 0.003)	(0.857 0.206)	(2.706 0.233)	(0.054 0.013)	(0.124 0.014)
4–7	-0.011 0.007	-0.002 −0.004	0.215 −0.302	-2.517 −0.288	-0.032 −0.021	-0.131 0.016	
	(0.014 0.008)	(0.011 0.004)	(0.761 0.469)	(1.752 0.466)	(0.046 0.024)	(0.101 0.024)
8–12	0.003 −0.015*	-0.003 −0.004	-1.153 0.400	1.144 0.954	-0.009 0.045*	-0.053 0.022
	(0.021 0.009)	(0.011 0.005)	(1.252 0.514)	(1.635 0.714)	(0.065 0.027)	(0.061 0.039)
13+	-0.039 0.003	0.001	0.180	2.131***	-0.015 −0.027**	0.720 −0.044
	0.134 (0.002)	-0.007 (0.001)	-0.096 (0.211)	(0.698)	(0.012)	(0.036)
Panel C: dummies						
Survey year	0.003***	0.004***	-0.012	0.181**	0.014***	0.020***
	(0.014 0.001)	(0.010 0.001)	(0.943 0.050)	(1.414 0.080)	(0.057 0.003)	(0.037 0.004)
Panel C: dummies Interaction: Years since diagnosis at baseline with survey year						
0–3	0.005 −0.024**	−0.007	0.352 −0.229	17.309 0.216	0.223 −0.033	-0.447 −0.072
	(0.052 0.011)	(0.059 0.009)	(2.123 0.667)	(0.975 0.617)	(0.186 0.040)	(0.749 0.044)
4–7	-0.031 −0.017	-0.049 -0.004	2.869 −1.167	10.878 -0.663	0.047 −0.033	-0.568 −0.074
	(0.042 0.015)	(0.059 0.009)	(2.664 0.760)	(9.504 1.174)	(0.137 0.047)	(0.544 0.047)
8–12	-0.066 −0.029	-0.026 −0.047***	-0.709 −0.427	13.733 −3.335*	-0.133 0.048	-0.878 0.005
	(0.063 0.020)	(0.059 0.012)	(4.134 1.162)	(0.695 1.846)	(0.267 0.069)	(0.521 0.113)
13+	-0.134 -0.062 −0.043**	-3.379 −0.010	13.309 1.741	9.164 1.559	-0.882 0.160**	-0.137***
	(0.098 0.021)	(0.068 0.014)	(4.715 1.426)	(9.239 2.093)	(0.284 0.070)	(0.146 0.013)
N	16208 20760	22427 26313	10771 17137	5746 8863	13583 13481	7391 6885

Notes ~~Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications.~~ Robust standard errors in parentheses. ~~All models include variables for states, urbanisation, level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year.~~ * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5. **Descriptive comparison of diagnosed and undiagnosed population with diabetes.**

	Males			Females		
	Diagnosed diabetes	Undiagnosed diabetes	P value (t-test)	Diagnosed diabetes	Undiagnosed diabetes	P value (t-test)
Employed	0.811	0.877	0.019	0.233	0.329	0.002
Hourly wage	35.280	30.939	0.220	37.242	32.822	0.495
Usual weekly working hours	44.562	46.682	0.166	31.838	39.788	0.004
Age	53.258	45.530	0.000	53.544	45.388	0.000
Any medical insurance	0.691	0.589	0.009	0.717	0.645	0.025
City of 2,5oo-15,000	0.092	0.105	0.593	0.116	0.114	0.916
City of 15,000-100,000	0.147	0.090	0.021	0.079	0.093	0.447
City of >100,000	0.332	0.290	0.267	0.292	0.329	0.250
Married	0.751	0.663	0.018	0.629	0.588	0.221
Number of children (<15) in household	0.972	1.138	0.110	0.934	1.250	0.001
Indigenous group	0.171	0.216	0.159	0.192	0.209	0.534
Primary	0.484	0.450	0.406	0.635	0.479	0.000
Secondary	0.212	0.230	0.594	0.126	0.230	0.000
High school	0.060	0.115	0.022	0.031	0.105	0.000
Higher education	0.147	0.109	0.147	0.025	0.071	0.003
Wealth index	−0.213	0.141	0.000	0.033	0.104	0.314
Subjective health						
very good	0.014	0.092	0.000	0.013	0.044	0.010
good	0.184	0.431	0.000	0.173	0.370	0.000
fair	0.664	0.446	0.000	0.635	0.533	0.002
bad	0.129	0.027	0.000	0.170	0.047	0.000
very bad	0.009	0.004	0.374	0.009	0.004	0.344
Glycated hemoglobin (HbA1c)	9.635	8.531	0.000	9.781	8.699	0.000
Hypertension (self-reported)	0.258	0.078	0.000	0.384	0.157	0.000
Blood pressure						
Systolic	136.475	130.981	0.001	136.426	123.516	0.000
Diastolic	84.562	82.448	0.025	84.912	80.019	0.000
Heart disease (self-reported)	0.032	0.008	0.013	0.041	0.025	0.178
BMI	28.989	28.385	0.128	30.573	30.058	0.234
Obese (BMI $\geq$ 30)	0.374	0.333	0.301	0.500	0.470	0.388

Notes Mean values. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

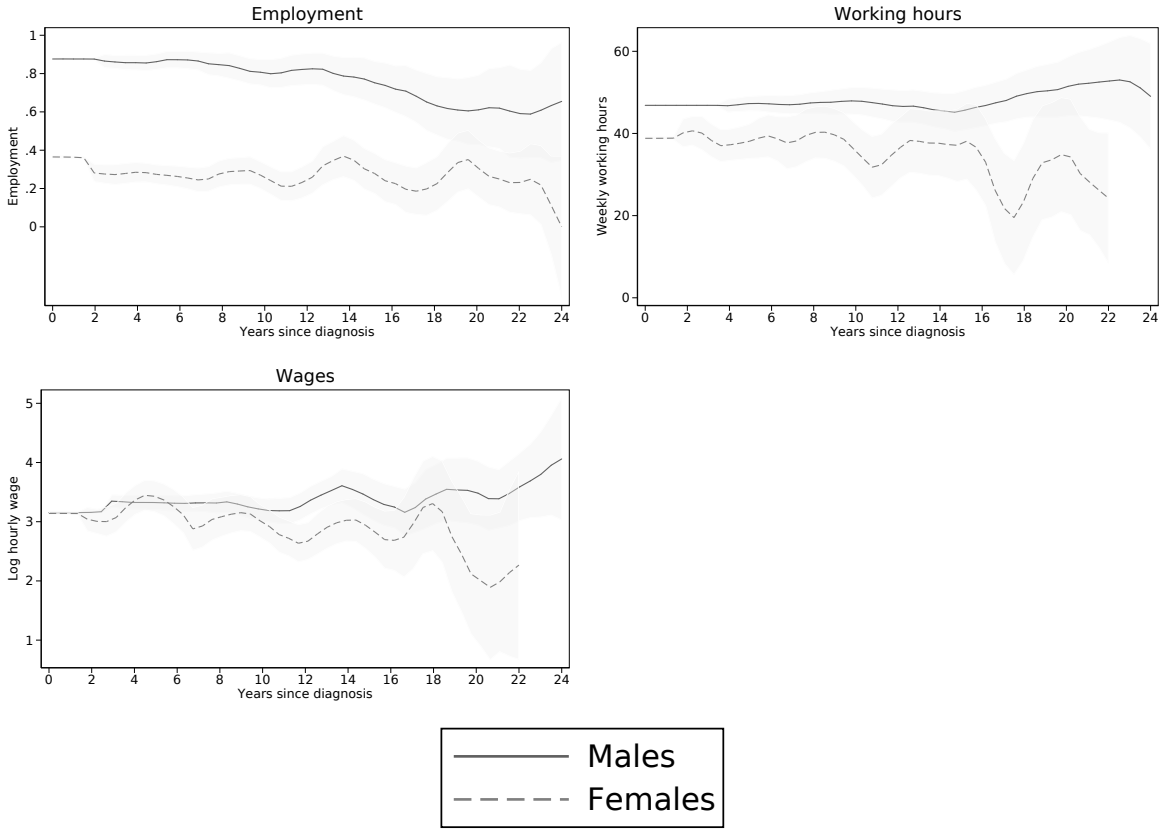
Table 6. **Biomarker results.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: Diabetes (self-reported)						
Self-reported diabetes	−.057**	−.057**	-.543~.728	-2.154~1.843	-.057~.033	-.005~.007
	(.025)	(-.026,.027)	(-1.427,1.408)	(-2.433,2.413)	(-.070,.067)	(-.121,.117)
Panel B: Diabetes (biomarker)						
Biomarker diabetes (HbA1c ≥ 6.5)	-.013~.015	-.034~.032*	.018~.134	1.382~1.392	-.005~.000	-.045~.039
	(.016)	(.018)	(-.849,.848)	(-1.480,1.472)	(.045)	(-.071,.070)
Panel C: Self-reported and undiagnosed diabetes						
Self-reported diabetes (β_1)	-.061~.059**	-.042~.045	-.715~.805	-3.954~3.574	-.067~.042	.034~.043
	(.028)	(-.031,.032)	(-1.574,1.571)	(-2.823,2.805)	(.085)	(-.137,.136)
Undiagnosed diabetes (HbA1c ≥ 6.5) (β_2)	.006~.003	-.020~.017	.224~.100	2.394~2.306	.014~.011	-.053~.049
	(.018)	(-.020,.021)	(.962)	(-1.647,1.638)	(-.050,.051)	(-.078,.077)
Panel D: HbA1c levels						
Self-reported diabetes	-.080~.008	-.066~.012*	.084~.237	-4.463~.080	-.061~.006	.011~.002
	(-.046,.006)	(-.046,.006)	(-2.409,.370)	(-4.592,.511)	(-.107,.018)	(-.227,.025)
~~HbA1c if ≥ 6.5~~ Undiagnosed diabetes	0.005	-.009~.008	-.150~.186	.318~.345	.004~.003	-.005~.004
	(.005)	(.006)	(-.253,.249)	(-.463,.457)	(.014)	(-.019) (-.052,.020)
N	2749	3537	2276	1121	1787	866

Notes Results are based on community level fixed effects. Robust standard errors in parentheses. All models include variables for ~~states, urbanization,~~ the level of education, ~~marital status, number of children < 6, wealth, health insurance status,~~ age and age squared and one dummy variable for each calendar year to account for the multiple years of data collection for the third wave. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural wage employment as the base). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures

Figure 1: Employment, wages, working hours and years since self-reported diabetes: Kernel-weighted local polynomial regression



Notes The shaded areas indicate the 95% confidence intervals.