Article:
Seuring, Till, Serneels, Pieter and Suhrcke, Marc orcid.org/0000-0001-7263-8626 (2019) The impact of diabetes on labour market outcomes in Mexico:A panel data and biomarker analysis. Social Science & Medicine. pp. 252-261. ISSN 1873-5347

https://doi.org/10.1016/j.socscimed.2019.05.051

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

Till Seuring*[a,b,c], Pieter Serneels[c], and Marc Suhrcke[d,e]

[a] *Leibniz Institute for Prevention Research and Epidemiology - BIPS, Research Group for Evidence-Based Public Health, Achterstr. 30, 28359 Bremen, Germany, Email: seuring@leibniz-bips.de, Phone: +49 421 218 569 23*

[b] *Health Sciences Bremen, Institute for Public Health and Nursing, University of Bremen, Bremen, Germany*

[c] *University of East Anglia, Norwich Research Park, Norwich, Norfolk, UK*

[d] *University of York, Heslington, York, UK*

[e] *Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette/Belval, Luxembourg*

Funding statement: No funding was received to support this project.

Declarations of interest: none

Manuscript number: SSM-D-18-02375R2

Acknowledgements: We thank Max Bachmann for his valuable comments on earlier drafts of the manuscript.

*Corresponding author.

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 and biomarker data we estimated the relationship between diabetes, as well as its time since diagnosis, and employment probabilities, wages and working hours. We further explored how these relationships differ for those with diagnosed and undiagnosed diabetes. For the longitudinal analyses, nationally representative data from 11995 men and 13858 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 7.7 and 6.3 percentage points for men and women, respectively, but no significant relationship with hours worked or wages. Employment probabilities fell gradually with each year since diagnosis for men but not for women. Using cross-sectional biomarker data, 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 (rather large) part of the population with undiagnosed diabetes, likely because of a selection of people in worse health and with a longer diabetes duration into the diagnosed population. Earlier diagnosis and improved treatment of diabetes therefore may prevent adverse health effects and related economic hardship.

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 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 (Lim et al., 2011; Gregg et al., 2012).

The observed increase in diabetes incidence has been attributed to a deterioration in diet and a reduction in physical activity (Barquera et al., 2008; Basu, Yoffe, Hills, & Lustig, 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, Rojas-Martinez, Aguilar-Salinas, & Hernández-Avila, 2017), increasing the risk of complications occurring during the productive lifespan. Only a minority of patients in Mexico achieves adequate blood glucose control (Barquera 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, & Suhrcke, 2015). In high-income countries substantial economic losses have been observed (Brown, Pagán, & Bastida, 2005; Brown, 2014; Brown et al., 2011; Minor, 2011, 2013; Minor

& MacEwan, 2016; Latif, 2009). A rare LMIC study exploited a natural experiment in China and found a significant reduction in income due to a recent diabetes diagnosis (Liu & 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, & Suhrcke, 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 (van Ewijk, 2011; Sotomayor, 2013; Li et al., 2010). Additionally, there may also be long-term effects on labour outcomes—either directly through reductions in contemporaneous productivity (Currie & Vogl, 2013), or indirectly by limiting educational attainment and human capital accumulation (Ayyagari, Grossman, & Sloan, 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, Guariguata, Weil, & Motala, 2014). So far, however, little evidence exists on the economic impact according to diabetes severity, its duration or for those with undiagnosed diabetes.

The objective of this study is to provide new evidence on the relationship of diabetes with 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 & 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 bias 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, exploiting 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), the average age in this subsample and hence the self-reported diabetes prevalence are 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. To investigate if the relationships of diabetes with our labour outcomes differed according to the year diabetes was diagnosed and thereby potentially distinguishing between type 1 and type 2 diabetes, we categorized self-reported diabetes into early-onset and late-onset cases. 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 of 35 had likely type 1 diabetes while those above had likely type 2 diabetes. Nonetheless, because we could not warranty that this was 100% accurate (as it was unlikely that both populations consisted exclusively of one type of diabetes) and lacked information about the use of insulin, we preferred to think of the groups as early- and late-onset groups. This separation also provided information about the relationships 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 around the survey, showing a prevalence of diagnosed diabetes of 7.5% in 2006 in a sample also including people over the age of 64 (Barquera 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 time since diagnosis 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 HbA1c levels that allowed us to identify those with undiagnosed diabetes 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 [INSERT LINK TO ONLINE FILE A], 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). We did not seek ethical approval for this study as we used publicly available secondary data.

3. Estimation strategy

3.1. Labour outcomes and self-reported diabetes

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 relationship for cases that received a diagnosis throughout the survey.

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

The fixed effects model 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 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 association of diabetes with employment.

To estimate the relationship 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.

In the main fixed effects (FE) models we only included 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 & Pischke, 2009). Hence, it is conceivable 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 could have affected the overall wealth of the person and its household. Neither did we 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 led 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 (Slade, 2012; De Fine Olivarius, Siersma, Køster-Rasmussen, Heitmann, & Waldorff, 2015; Seuring, Suhrcke, Serneels, & Bachmann, 2018). 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 and household wealth approximated by a household asset index. The household asset index was created using principal component analysis of household assets and housing following 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 an additional robustness check we also controlled for obesity by including an indicator that was one for a BMI ≥ 30 and zero for a BMI < 30. Stata 15 was used for all analyses (StataCorp, 2017).

In an additional analysis we split the diabetes indicator D_{it} into an early- and late onset diabetes group. These groups were defined by the age a diabetes diagnosis was first reported in the survey. The early onset group was comprised of people with diabetes who were below the age of 35 when they first reported to have diabetes, while the late onset group was comprised of people with diabetes who were 35 and older when they first reported to have diabetes. The two diabetes onset variables were then used instead of D_{it}, with no diabetes as the reference group, in order to estimate the relationship of early- and late onset diabetes with the respective labour outcome.

3.2. Labour outcomes and time since diagnosis

The chronic nature and irreversibility of diabetes motivate our exploration of the long term relationships post diagnosis. To this end, we replaced the binary diabetes indicator of Eq 1 with a continuous variable indicating years since the diagnosis was first reported. 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 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.

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) we chose three nodes located at 3, 7 and 12 years after diagnosis. The first three years should capture any immediate associations with the diagnosis, the years four to seven any associations during time of adaptation to the disease and the later terms the associations after a longer time has passed. We also estimated a non-linear model using dummy variables for time since diagnosis groups rather than splines, applying the same time since diagnosis cut-offs. Because the year of diagnosis was only reported in the third wave, for the previous waves we could only create a time since diagnosis variable for those that were also interviewed in the third wave. Also, because we used the years of diabetes at baseline for the interaction effect, we could not estimate the effect of time since diagnosis for diabetes cases that were diagnosed after entering the sample. A reported diagnosis in the year of the interview was counted as 'one year since diagnosis'. As a robustness check, we used only the data from wave three containing the original time since diagnosis variable and estimated ordinary least squares models of the association of time since diagnosis with labour outcomes. Further, fixed effects models accounting for additional time-variant control variables including obesity were estimated. We also re-estimated the time-since diagnosis models splitting the

diabetes population into early- and late onset groups.

3.3. Labour outcomes and biometrically measured diabetes

The biomarker analysis consisted of three steps. We first re-estimated Eq 1 to assess the relationship between self-reported diabetes 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. X_i contains control variables for age, age squared, education, being of an indigenous group and survey year dummies, to account for the fact that data collection for wave three took place between 2009 and 2012. For the working hours and wage models we also controlled for the type of work.

In a second step, we estimated the relationship 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 relationship 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 relationship 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 how the severity of diabetes contributed to the relationship of self-reported and undiagnosed diabetes with labour outcomes (Eq 5). Therefore we created for both self-reported and undiagnosed cases a variable that was 0 for HbA1c $< 6.5\%$ and increased continuously with HbA1c for those $\geq 6.5\%$, applying the transformation HbA1c $- 6.4$ for those with HbA1c $\geq 6.5\%$. We estimated

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 HbA1c_i + \beta_3 Dsr_i * HbA1c_i + \beta_4 X_i + v_i + u_i. \tag{5}$$

The coefficient for β_1 then provided an estimate of the relationship of self-reported diabetes with labour outcomes, irrespective of the precise HbA1c level above the threshold (i.e. severity), β_2 provided an estimate of the relationship of an increase in HbA1c with labour outcomes for those with diabetes but undiagnosed, and β_3 estimated the difference in the relationship of HbA1c with the outcomes for self-reported diabetes compared to undiagnosed diabetes.

We tested the robustness of the results by including additional time-variant control variables, including an indicator for obesity.

4. Results

4.1. Labour outcomes and self-reported diabetes

The results of estimating Eq 1 in Table 2 indicated lower probabilities of employment for men and women with self-reported diabetes of 7.7 p.p. for men and 6.3 p.p. for women. There were no statistically significant relationships between diabetes and working hours or wages.

Dividing the diabetes population into early and late onset groups, men, and potentially also women, with a later diabetes onset had lower employment probabilities (Supplementary Table S2)[INSERT LINK TO ONLINE FILE A]. In particular in women an early diabetes onset was associated with lower employment probabilities. For working hours but also for wages, the estimates were less precise and quite large, leaving a substantial part of the observed variation unexplained. Finally, we found higher wages for women with an early diabetes onset but no relationship for men.

To assess whether diabetes was related to 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 [INSERT LINK TO ONLINE FILE A]). 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) 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) [INSERT LINK TO ONLINE FILE A]. 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 of the years since diagnosis on the respective labour outcome. 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 the relationship of an additional year since diagnosis with labour market outcomes. They indicate a reduction in male but not female employment probabilities with every year since the diagnosis. We also found some indication that time since diagnosis was related with a reduction in wages for women. Using diabetes onset groups, there was only evidence of a negative relationship of time since diagnosis with employment for men in the late-onset group (see Supplementary Table S4 [INSERT LINK TO ONLINE FILE A]). For monthly working hours the results indicate a negative relationship with time since diagnosis for those with early-onset diabetes, but not with late-onset diabetes. Further, we found a large positive relationship of time since diagnosis with wages of women with early-onset diabetes, but also a negative relationship for women with late-onset diabetes.

The non-linear results for the spline function and dummy variable approach are presented in panels B and C, respectively. The spline function results were not precisely estimated for either men or women, and provide only suggestive evidence that male employment probabilities were lower during the first three years, and eight to twelve years after diagnosis. 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 model that uses spline functions, the model that uses dummy variables did not suggest higher female working hours 13 years after diagnosis. However, similar to the model that uses spline functions 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 into early and late-onset groups,

as this implied major reductions in statistical power.

Controlling for other time-variant variables or additionally for obesity, results remained similar (Supplementary Table S8 and Supplementary Table S11 [INSERT LINK TO ONLINE FILE A]). We also estimated the models using only the cross-sectional data from the third survey wave, where the year of diagnosis was first reported (see Supplementary Table S12 [INSERT LINK TO ONLINE FILE A]). The results of this analysis confirmed the linear negative relationship of time since diagnosis with employment probabilities for men, and found a linear relationship for women as well. In general, the results of the cross-sectional analysis for employment probabilities were mostly qualitatively similar to those of the fixed effects analysis, also in the non-linear models. However, the coefficients indicated more pronounced relationships than in the fixed effects models. For working hours and wages, fewer (statistically significant) relationships were found compared to the fixed effects results.

4.3. Cross-sectional biomarker analysis

As reported in Supplementary Table S5 [INSERT LINK TO ONLINE FILE A], 18% of the observations in the biomarker sample were undiagnosed, which 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 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 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 a significant relationship of diabetes with working hours or wages, so that in what follows 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

indicate that the relationship with employment became weaker when diabetes was defined by HbA1c levels instead of self-reported diabetes, in particular for men. Results in Panel C indicate an absence of a (statistically significant) negative relationship between undiagnosed diabetes and labour outcomes. The results in panel D show that increasing HbA1c levels appear not to be related to the observed reductions in employment probabilities due to diabetes. Robustness checks, where we additionally controlled for other potential confounding factors as well as obesity, did not lead to qualitatively different results (see Supplementary Tables S13 and S14 [INSERT LINK TO ONLINE FILE A]).

5. Discussion

Diabetes is now one of the most common chronic diseases in LMICs, as well as high-income countries (HICs), with 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 an adverse relationship of self-reported diabetes with employment, working hours and wages. While earlier work showed evidence for Mexico for employment (Seuring, Goryakin, & Suhrcke, 2015), this paper presented, by our knowledge, the first evidence on the relationship of diabetes with working hours and wages. Furthermore, we added to the study of Seuring, Goryakin, and Suhrcke (2015) by using longitudinal instead of cross-sectional data. We provided first evidence of the relationship of diabetes with labour market outcomes over the longer term in Mexico and explored the role of undiagnosed diabetes. We confirmed earlier findings for Mexico by Seuring,

Goryakin, and Suhrcke (2015) insofar that we found a negative relationship of diabetes with male employment. We further showed more conclusive evidence that also for women diabetes contributes to lower employment probabilities. Taking into account the general differences in employment between men and women, the estimated relationships translate into relatively lower employment probabilities of almost 9% for men and 17% for women with diabetes. We also found that the relationships 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 is similar to the findings of Seuring, Goryakin, and Suhrcke (2015) in their stratified analysis of an older and younger age group. Analyses of the long term relationship indicated that employment probabilities fell gradually in the years following diagnosis, albeit only for men. The results using non-linear models were less clear, potentially due to reductions in statistical power. Those results suggested that the negative relationship with employment probabilities and wages appeared in particular immediately after the diabetes diagnosis and then again after a considerable time of living with the disease. The linear relationship found in our analysis contrasts with estimates for the USA, where such a linear effect on employment probabilities was absent; however, revealed falling employment probabilities after 11–15 years for females and after 2–5 years for males when non-linear models were used (Minor, 2013).

Overall, there was no consistent relationship of diabetes with working hours or wages. 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 the population with diabetes was 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 a statistically significant relationship was only observed between diagnosed diabetes and employment, but not between undiagnosed diabetes and employment (Minor & MacEwan, 2016). Our results further indicated that diabetes severity as proxied by current HbA1c levels was likely unrelated to labour outcomes in men and women. This is in line with findings for Mexican-Americans in the USA, where employment outcomes were unrelated to higher HbA1c levels (Brown et al., 2011). A possible explanation may be that HbA1c levels are primarily informative for the last three months, and are neither the only nor the best indicator for the severity of diabetes. It therefore appears 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 identified negative relationship 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, Grönqvist, & Gudbjörnsdottir, 2011; Schaller & Stevens, 2015), though they were focused on 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, indicating that the main results are robust to the inclusion of additional time-variant variables. A limitation of the time since diagnosis analysis, 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. The results of the time since diagnosis 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 led to different HbA1c levels, then 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 informative.

Despite these limitations, our findings bear important implications. First, the relationship of self-reported diabetes with labour outcomes in Mexico seemed mostly limited to its relationship with lower employment probabilities. Second, its effect on employment was much stronger for females, though the underlying reasons for this require further investigation. Potential explanations are that lower working hours or wages for women make a drop-out less costly. Other evidence suggests that women with diabetes are in worse metabolic health compared to men when they cross the diabetes threshold (Peters, Huxley, Sattar, & Woodward, 2015), increasing their odds 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 undiag-

nosed sub-populations, 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 very early. Earlier diagnosis and ensuing effective treatment may mitigate 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 until a medical diagnosis. This would allow for a better understanding of when adverse economic effects start to arise. 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

Alegre-Díaz, J., Herrington, W., López-Cervantes, M., Gnatiuc, L., Ramirez, R., Hill, M., ... Emberson, J. R. (2016). Diabetes and Cause-Specific Mortality in Mexico City. *New England Journal of Medicine*, *375*(20), 1961–1971. doi:10.1056/NEJMoa1605368

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

Ayyagari, P., Grossman, D., & Sloan, F. (2011). Education and health: evidence on adults with diabetes. *International Journal of Health Care Finance and Economics*, *11*(1), 35–54. doi:10.1007/s10754-010-9087-x

Barquera, S., Campos-Nonato, I., Aguilar-Salinas, C., Lopez-Ridaura, R., Arredondo, A., & Rivera-Dommarco, J. (2013). Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy. *Globalization and Health*, *9*(1), 3. doi:10.1186/1744-8603-9-3

Barquera, S., Hernandez-Barrera, L., Tolentino, M. L., Espinosa, J., Ng, S. W., Rivera, J. A., & Popkin, B. M. (2008). Energy Intake from Beverages Is Increasing among Mexican Adolescents and Adults. *Journal of Nutrition*, *138*(12), 2454–2461. doi:10.3945/jn.108.092163

Basu, S., Yoffe, P., Hills, N., & Lustig, R. H. (2013). The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data. *PLoS ONE*, *8*(2), e57873. doi:10.1371/journal.pone.0057873

Beagley, J., Guariguata, L., Weil, C., & Motala, A. a. (2014). Global estimates of undiagnosed diabetes in adults. *Diabetes Research and Clinical Practice*, *103*(2), 150–160. doi:10.1016/j.diabres.2013.11.001

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

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

Brown, H. S., Pagán, J. A., & Bastida, E. (2005). The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit. *Health Economics*, *14*(5), 537–544. doi:10.1002/hec.942

Brown, H. S., Perez, A., Yarnell, L. M., Pagan, J. a., Hanis, C. L., Fischer-Hoch, S. P., & McCormick, J. B. (2011). Diabetes and employment productivity: does diabetes management matter? *American Journal of Managed Care*, *17*(8), 569–576.

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

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

Currie, J. & Vogl, T. (2013). Early-Life Health and Adult Circumstance in Developing Countries. *Annual Review of Economics*, *5*(1), 1–36. doi:10.1146/annurev-economics-081412-103704

De Fine Olivarius, N., Siersma, V. D., Køster-Rasmussen, R., Heitmann, B. L., & Waldorff, F. B. (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. *PLoS ONE*, *10*(4), 1–14. doi:10.1371/journal.pone.0122219

Filmer, D. & Pritchett, L. (2001). Estimating wealth effects without expenditure data-Or tears: An application to educational enrollments in states of India. *Demography*, *38*(1), 115–132.

Flores-Hernández, S., Saturno-Hernández, P. J., Reyes-Morales, H., Barrientos-Gutiérrez, T., Villalpando, S., & Hernández-Ávila, M. (2015). Quality of Diabetes Care: The Challenges of an Increasing Epidemic in Mexico. Results from Two National Health Surveys (2006 and 2012). *Plos One*, *10*(7), e0133958. doi:10.1371/journal.pone.0133958

Gregg, E. W., Chen, H., Wagenknecht, L. E., Clark, J. M., Delahanty, L. M., Bantle, J., ... Bertoni, A. G. (2012). Association of an Intensive Lifestyle Intervention With Re-

mission of Type 2 Diabetes. *Journal of the American Medical Association*, *308*(23), 2489. doi:10.1001/jama.2012.67929

Herrington, W. G., Alegre-Díaz, J., Wade, R., Gnatiuc, L., Ramirez-Reyes, R., Hill, M., ... Emberson, J. R. (2018). Effect of diabetes duration and glycaemic control on 14-year cause-specific mortality in Mexican adults: a blood-based prospective cohort study. *The Lancet Diabetes & Endocrinology*, *8587*(18), 1–9. doi:10.1016/S2213-8587(18)30050-0

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

Latif, E. (2009). The impact of diabetes on employment in Canada. *Health Economics*, *18*(5), 577–589. doi:10.1002/hec.1390

Li, Y., He, Y., Qi, L., Jaddoe, V. W., Feskens, E. J. M., Yang, X., ... Hu, F. B. (2010). Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood. *Diabetes*, *59*(10), 2400–2406. doi:10.2337/db10-0385

Lim, E. L., Hollingsworth, K. G., Aribisala, B. S., Chen, M. J., Mathers, J. C., & Taylor, R. (2011). Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia*, *54*(10), 2506–2514. doi:10.1007/s00125-011-2204-7

Liu, X. & Zhu, C. (2014). Will knowing diabetes affect labor income? Evidence from a natural experiment. *Economics Letters*, *124*(1), 74–78. doi:10.1016/j.econlet.2014.04.019

Minor, T. (2011). The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey. *Health Economics*, *20*(12), 1468–1486. doi:10.1002/hec.1685

Minor, T. (2013). An investigation into the effect of type I and type II diabetes duration on employment and wages. *Economics & Human Biology*, *11*(4), 534–544. doi:10.1016/j.ehb.2013.04.004

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

Peters, S. A. E., Huxley, R. R., Sattar, N., & Woodward, M. (2015). Sex Differences in the Excess Risk of Cardiovascular Diseases Associated with Type 2 Diabetes: Potential Explanations and Clinical Implications. *Current Cardiovascular Risk Reports*, *9*(7), 1–7. doi:10.1007/s12170-015-0462-5

Reynoso-Noverón, N., Mehta, R., Almeda-Valdes, P., Rojas-Martinez, R., Villalpando, S., Hernández-Ávila, M., & Aguilar-Salinas, C. a. (2011). Estimated incidence of cardiovascular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey. *Cardiovascular Diabetology*, *10*(1), 1. doi:10.1186/1475-2840-10-1

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

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

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

Seuring, T., Archangelidi, O., & Suhrcke, M. (2015). The Economic Costs of Type 2 Diabetes: A Global Systematic Review. *PharmacoEconomics*, *33*(8), 811–831. doi:10.1007/s40273-015-0268-9

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

Seuring, T., Suhrcke, M., Serneels, P., & Bachmann, M. (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. *IZA Discussion Papers*, (11817).

Slade, A. N. (2012). Health Investment Decisions in Response to Diabetes Information in Older Americans. *Journal of Health Economics*, *31*(3), 502–520.

Sotomayor, O. (2013). Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes. *Economics & Human Biology*, *11*(3), 281–293. doi:10.1016/j.ehb.2012.02.009

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

van Ewijk, R. (2011). Long-Term Health Effects on the Next Generation of Ramadan Fasting during Pregnancy. *Journal of Health Economics*, *30*(6), 1246–1260.

Williams, A. L., Jacobs, S. B. R., Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., ... Tusié-Luna, T. (2013). Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*, *506*(7486), 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.

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

Yuan, X., Liu, T., Wu, L., Zou, Z.-Y., & Li, C. (2015). Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study. *British Medical Journal Open*, *5*(4), e006633–e006633. doi:10.1136/bmjopen-2014-006633

Zajacova, A., Dowd, J., Schoeni, R. F., & Wallace, R. B. (2010). Consistency and precision of cancer reporting in a multiwave national panel survey. *Population Health Metrics*, *8*(1), 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, Dowd, Schoeni, and Wallace (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, Liu, Wu, Zou, & Li, 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.

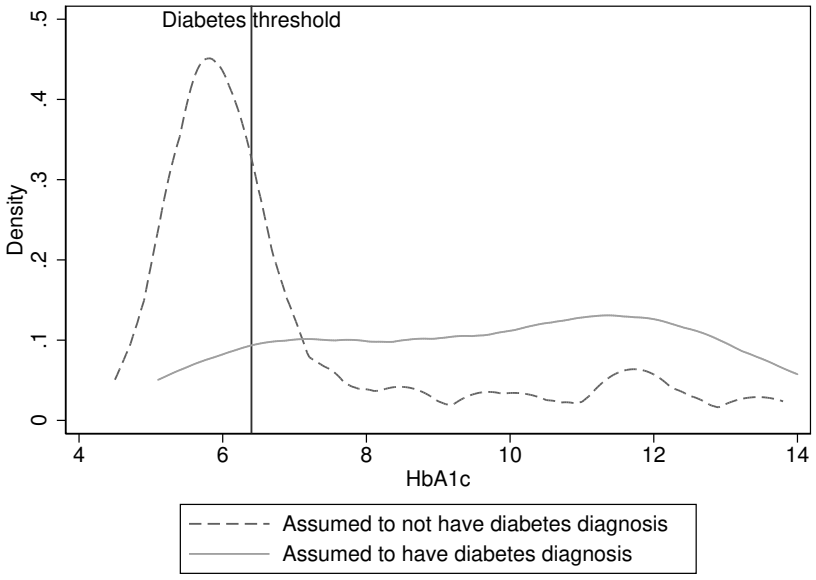
Table S1. Inconsistencies in diabetes self-report in MxFLS.

	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- versus 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	0.134	-0.195^{**}	14.395^{*}	-18.665^{*}	-0.513^{*}	0.362^{***}
	(0.176)	(0.086)	(8.377)	(9.650)	(0.311)	(0.039)
Late-onset	-0.082^{***}	-0.053^{**}	−1.360	−1.267	0.016	0.059
	(0.025)	(0.025)	(1.500)	(2.565)	(0.067)	(0.165)
N	21388	27339	17618	9115	13830	7070

Notes Robust standard errors in parentheses. All models include year dummies. The early-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time before age 35. The late-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time at or after age 35. Having no diabetes is the reference group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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	0.036	−0.105	−0.233*	−0.066	0.328**	−0.020
	(0.215)	(0.074)	(0.139)	(0.047)	(0.161)	(0.049)
Late-onset	−0.024	0.006	−0.008	−0.019**	−0.056**	−0.033*
	(0.029)	(0.019)	(0.022)	(0.009)	(0.026)	(0.019)
N	20537	26478	20537	26478	20537	26478

Notes Robust standard errors in parentheses. All models include year dummies. The early-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time before age 35. The late-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time at or after age 35. Having no diabetes is the reference group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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
Survey year	0.003***	0.004***	−0.019	0.181**	0.014***	0.019***
	(0.001)	(0.001)	(0.050)	(0.080)	(0.003)	(0.004)
Years since diagnosis at baseline (early-onset) × Survey year	−0.003	0.000	0.197	−2.933***	−0.008	0.264***
	(0.003)	(0.001)	(0.233)	(0.040)	(0.019)	(0.002)
Years since diagnosis at baseline (late-onset) × Survey year	−0.003***	−0.001	0.006	0.042	−0.004	−0.010***
	(0.001)	(0.001)	(0.062)	(0.108)	(0.003)	(0.003)
N	16329	22519	13614	7430	10815	5769

Notes Robust standard errors in parentheses. The early-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time before age 35. The late-onset group is comprised of people with diabetes reporting a diabetes diagnosis for the first time at or after age 35. Having no diabetes is the reference group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

Robustness checks

Additional time-variant controls

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

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Diabetes	-0.074^{***}	-0.069^{***}	-0.825	-2.132	0.012	0.074
	(0.025)	(0.024)	(1.493)	(2.513)	(0.067)	(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 year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $p < 0.01$.

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	−0.017	−0.012	−0.049*	−0.007	−0.023***	−0.032*
	(0.029)	(0.022)	(0.026)	(0.018)	(0.009)	(0.018)
N	20537	20537	20537	26478	26478	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, and year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $p < 0.01$.

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

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Survey year	0.003***	0.004***	−0.018	0.207**	0.013***	0.022***
	(0.001)	(0.001)	(0.052)	(0.092)	(0.003)	(0.005)
Years since diagnosis at baseline × survey year	−0.003***	−0.001*	0.011	0.031	−0.004	−0.009***
	(0.001)	(0.001)	(0.062)	(0.104)	(0.003)	(0.003)
Panel B: splines						
Survey year	0.003***	0.004***	−0.013	0.207**	0.013***	0.022***
	(0.001)	(0.001)	(0.052)	(0.092)	(0.003)	(0.005)
Interaction: Years since diagnosis at baseline with survey year						
0–3	−0.006*	−0.000	−0.045	0.240	−0.002	−0.032**
	(0.004)	(0.003)	(0.206)	(0.237)	(0.013)	(0.014)
4–7	0.006	−0.004	−0.299	−0.260	−0.022	0.018
	(0.008)	(0.004)	(0.469)	(0.469)	(0.023)	(0.024)
8–12	−0.015*	−0.003	0.399	−0.985	0.046*	0.021
	(0.009)	(0.005)	(0.519)	(0.727)	(0.027)	(0.040)
13+	0.003	0.001	0.188	2.101***	−0.026**	−0.042
	(0.002)	(0.001)	(0.213)	(0.717)	(0.012)	(0.037)
Panel C: dummies						
Survey year	0.003***	0.004***	−0.011	0.208**	0.013***	0.022***
	(0.001)	(0.001)	(0.052)	(0.093)	(0.003)	(0.005)
Interaction: Years since diagnosis at baseline with survey year						
0–3	−0.024**	−0.009	−0.248	0.199	−0.031	−0.076*
	(0.011)	(0.009)	(0.669)	(0.618)	(0.040)	(0.044)
4–7	−0.018	0.000	−1.173	0.611	−0.032	−0.069
	(0.016)	(0.009)	(0.770)	(1.179)	(0.047)	(0.047)
8–12	−0.030	−0.047***	−0.463	−3.313*	0.053	−0.008
	(0.020)	(0.012)	(1.177)	(1.840)	(0.071)	(0.115)
13+	−0.043**	−0.011	1.759	1.321	−0.154**	−0.126***
	(0.021)	(0.014)	(1.430)	(2.098)	(0.067)	(0.020)
N	16265	22435	13562	7402	10766	5744

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

Additionally controlling for time-variant variables and obesity

Table S9. **Labour outcomes and self-reported diabetes using additional control variables including obesity.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Obese (BMI $\geq$ 30)	0.009	−0.002	−0.073	−1.105	0.026	0.082
	(0.012)	(0.013)	(0.772)	(1.187)	(0.038)	(0.061)
Diabetes	−0.064**	−0.076***	−1.021	−0.300	−0.010	0.040
	(0.028)	(0.027)	(1.765)	(2.909)	(0.076)	(0.181)
N	17992	24145	14867	7931	11712	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, and year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S10. **Selection into types of work and self-reported diabetes using additional control variables including obesity.**

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Obese (BMI $\geq$ 30)	0.007	−0.031**	0.035**	−0.019*	0.003	0.011
	(0.017)	(0.013)	(0.014)	(0.011)	(0.004)	(0.009)
Diabetes	0.002	−0.001	−0.068**	−0.018	−0.022**	−0.029
	(0.034)	(0.023)	(0.028)	(0.020)	(0.010)	(0.021)
N	17261	17261	17261	23377	23377	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, and year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S11. **Relationship between self-reported years since diagnosis and employment probabilities using additional control variables including obesity.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Obese (BMI $\geq$ 30)	0.010	−0.003	0.054	−0.288	0.020	0.023
	(0.013)	(0.014)	(0.835)	(1.236)	(0.041)	(0.065)
Survey year	0.002**	0.004***	−0.036	0.164	0.012***	0.019***
	(0.001)	(0.001)	(0.061)	(0.106)	(0.003)	(0.006)
Years since diagnosis at baseline × survey year	−0.004***	−0.001**	0.039	−0.022	−0.004	−0.011**
	(0.001)	(0.001)	(0.064)	(0.092)	(0.003)	(0.005)
Panel B: splines						
Obese (BMI $\geq$ 30)	0.009	−0.002	0.086	−0.270	0.020	0.019
	(0.013)	(0.014)	(0.828)	(1.239)	(0.041)	(0.065)
Survey year	0.002**	0.004***	−0.034	0.160	0.012***	0.019***
	(0.001)	(0.001)	(0.061)	(0.106)	(0.003)	(0.006)
0–3	−0.008*	0.003	0.100	0.057	−0.009	−0.033**
	(0.004)	(0.003)	(0.243)	(0.214)	(0.015)	(0.014)
4–7	0.000	−0.009*	−0.489	0.283	−0.002	0.015
	(0.009)	(0.005)	(0.530)	(0.438)	(0.024)	(0.026)
8–12	−0.007	−0.002	0.589	−1.074	0.016	0.028
	(0.010)	(0.006)	(0.544)	(0.727)	(0.021)	(0.045)
13+	0.003	0.000	0.140	0.953	−0.017	−0.063
	(0.005)	(0.003)	(0.180)	(1.267)	(0.011)	(0.051)
Panel C: dummies						
Obese (BMI $\geq$ 30)	0.010	−0.003	0.069	−0.264	0.019	0.020
	(0.013)	(0.014)	(0.831)	(1.237)	(0.041)	(0.065)
Survey year	0.002**	0.004***	−0.031	0.163	0.013***	0.019***
	(0.001)	(0.001)	(0.061)	(0.106)	(0.003)	(0.006)
0–3	−0.029**	−0.001	−0.132	−0.302	−0.046	−0.082
	(0.013)	(0.011)	(0.819)	(0.534)	(0.048)	(0.052)
4–7	−0.034*	0.003	−0.446	0.924	−0.043	−0.078
	(0.018)	(0.010)	(0.735)	(1.199)	(0.044)	(0.048)
8–12	−0.048**	−0.062***	−0.852	−2.406	0.042	0.002
	(0.020)	(0.014)	(1.428)	(2.460)	(0.050)	(0.112)
13+	−0.032	−0.018	2.746**	−1.600***	−0.130*	−0.148***
	(0.027)	(0.017)	(1.262)	(0.461)	(0.075)	(0.016)
N	13880	19978	11601	6499	9254	5051

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Cross-sectional analysis for time since diagnosis

Table S12. **Relationship between self-reported years since diagnosis and employment probabilities (only wave three, ordinary least squares).**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Years since diagnosis	−0.009***	−0.004***	0.090	0.014	0.003	−0.014*
	(0.002)	(0.002)	(0.126)	(0.187)	(0.007)	(0.008)
Panel B: splines						
0–3	−0.014*	−0.011	0.362	0.625	0.019	0.033
	(0.008)	(0.008)	(0.447)	(0.694)	(0.019)	(0.034)
4–7	0.029**	−0.005	−0.838	0.311	0.006	−0.037
	(0.014)	(0.016)	(0.920)	(1.642)	(0.041)	(0.071)
8–12	−0.054***	0.004	1.766	−1.019	−0.085	−0.074
	(0.019)	(0.018)	(1.136)	(1.804)	(0.056)	(0.069)
13+	−0.007	−0.003	−0.513	−0.223	0.041	−0.005
	(0.009)	(0.004)	(0.545)	(0.613)	(0.031)	(0.026)
Panel C: dummies						
0–3	−0.052*	−0.042	0.423	1.364	0.028	−0.001
	(0.027)	(0.027)	(1.431)	(2.153)	(0.064)	(0.099)
4–7	0.021	−0.041	0.022	2.951	0.120*	0.101
	(0.024)	(0.032)	(1.580)	(3.158)	(0.063)	(0.138)
8–12	−0.104**	−0.075*	2.419	2.206	−0.117	−0.254
	(0.051)	(0.042)	(3.008)	(4.531)	(0.127)	(0.165)
13+	−0.193***	−0.056	1.693	−0.868	0.023	−0.275
	(0.055)	(0.037)	(2.838)	(4.011)	(0.165)	(0.170)
N	8233	10501	6819	3602	5516	2881

Notes Robust standard errors in parentheses. All models include variables for the level of education, age and age squared and year dummies 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) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Biomarker robustness checks

Table S13. **Biomarker results using additional control variables.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: Diabetes (self-reported)						
Self-reported diabetes	−.057**	−.057**	−.543	−2.154	−.057	−.005
	(.025)	(.026)	(1.427)	(2.433)	(.070)	(.121)
Panel B: Diabetes (biomarker)						
Biomarker diabetes (HbA1c $\geq$ 6.5)	−.013	−.034*	0.018	1.382	−.005	−.045
	(.016)	(.018)	(.849)	(1.480)	(.045)	(.071)
Panel C: Self-reported and undiagnosed diabetes						
Self-reported diabetes (β_1)	−.061**	−.042	−.715	−3.954	−.067	0.034
	(.028)	(.031)	(1.574)	(2.823)	(.085)	(.137)
Undiagnosed diabetes (HbA1c $\geq$ 6.5) (β_2)	0.006	−.020	0.224	2.394	0.014	−.053
	(.018)	(.020)	(.962)	(1.647)	(.050)	(.078)
Panel D: HbA1c levels for self-reported and undiagnosed diabetes						
Self-reported diabetes	−.080*	−.066	0.084	−4.463	−.061	0.011
	(.046)	(.046)	(2.409)	(4.592)	(.107)	(.227)
HbA1c if $\geq$ 6.5	0.005	−.009*	−.150	0.318	0.004	−.005
	(.005)	(.006)	(.253)	(.463)	(.014)	(.019)
Self-reported diabetes × HbA1c if $\geq$ 6.5	0.003	0.010	−.064	0.375	−.002	−.000
	(.012)	(.012)	(.668)	(1.043)	(.030)	(.052)
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 the level of education, age and age squared, indigenous, state, urbanisation, marital status, number of children < 6, wealth and year dummies 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). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S14. **Biomarker results using additional control variables including obesity.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: Diabetes (self-reported)						
Obese (BMI $\geq$ 30)	0.009	−.023	0.873	0.494	0.021	−.057
	(.016)	(.016)	(.798)	(1.347)	(.043)	(.062)
Self-reported diabetes	−.062**	−.052*	−.759	−1.901	−.047	−.003
	(.026)	(.028)	(1.444)	(2.566)	(.070)	(.136)
Panel B: Diabetes (biomarker)						
Obese (BMI $\geq$ 30)	0.009	−.022	0.858	0.280	0.020	−.052
	(.016)	(.017)	(.855)	(1.370)	(.045)	(.064)
Biomarker diabetes (HbA1c $\geq$ 6.5)	−.017	−.029	−.006	1.802	0.001	−.053
	(.016)	(.018)	(.883)	(1.570)	(.046)	(.075)
Panel C: Self-reported and undiagnosed diabetes						
Obese (BMI $\geq$ 30)	0.009	−.022	0.862	0.306	0.020	−.053
	(.016)	(.017)	(.855)	(1.370)	(.045)	(.064)
Self-reported diabetes (β_1)	−.062**	−.041	−.958	−4.014	−.061	0.041
	(.030)	(.032)	(1.653)	(2.984)	(.087)	(.143)
Undiagnosed diabetes (HbA1c $\geq$ 6.5) (β_2)	0.001	−.016	0.260	2.827	0.017	−.062
	(.018)	(.021)	(.995)	(1.744)	(.052)	(.082)
Panel D: HbA1c levels for self-reported and undiagnosed diabetes						
Obese (BMI $\geq$ 30)	0.009	−.023	0.875	0.555	0.021	−.058
	(.016)	(.016)	(.796)	(1.340)	(.043)	(.062)
Self-reported diabetes	−.104**	−.061	−.296	−4.421	−.075	0.084
	(.050)	(.049)	(2.495)	(4.859)	(.109)	(.249)
HbA1c if $\geq$ 6.5	0.002	−.009	−.126	0.501	0.005	−.006
	(.006)	(.006)	(.273)	(.570)	(.014)	(.020)
Self-reported diabetes × HbA1c if $\geq$ 6.5	0.012	0.010	−.033	0.277	0.004	−.020
	(.014)	(.012)	(.710)	(1.061)	(.028)	(.055)
N	2606	3362	2158	1065	1699	824

Notes Results are based on community level fixed effects. Robust standard errors in parentheses. All models include variables for the level of education, age and age squared, indigenous, state, urbanisation, marital status, number of children < 6, wealth and year dummies 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). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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)		6.94			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	20394	994		25672	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	-0.077^{***}	-0.063^{***}	-0.940	-1.941	0.001	0.065
	(0.025)	(0.024)	(1.489)	(2.531)	(0.066)	(0.162)
N	21388	27339	17618	9115	13830	7070

Notes Robust standard errors in parentheses. All models include 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	−0.022	−0.014	−0.045*	−0.001	−0.023**	−0.032*
	(0.029)	(0.022)	(0.026)	(0.018)	(0.009)	(0.018)
N	20537	20537	20537	26478	26478	26478

Notes Robust standard errors in parentheses. All models include year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. **Relationship between self-reported years since diagnosis and employment probabilities.**

	Employment		Weekly working hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Survey year	0.003***	0.004***	−0.006	0.200**	0.012***	0.018***
	(0.001)	(0.001)	(0.050)	(0.081)	(0.003)	(0.004)
Years since diagnosis at baseline × Survey year	−0.003***	−0.001	0.011	0.041	−0.004	−0.010***
	(0.001)	(0.001)	(0.061)	(0.108)	(0.003)	(0.003)
Panel B: splines						
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	−0.006*	0.001	−0.038	0.262	−0.004	−0.032**
	(0.004)	(0.003)	(0.206)	(0.233)	(0.013)	(0.014)
4–7	0.007	−0.004	−0.302	−0.288	−0.021	0.016
	(0.008)	(0.004)	(0.469)	(0.466)	(0.024)	(0.024)
8–12	−0.015*	−0.004	0.400	−0.954	0.045*	0.022
	(0.009)	(0.005)	(0.514)	(0.714)	(0.027)	(0.039)
13+	0.003	0.001	0.180	2.131***	−0.027**	−0.044
	(0.002)	(0.001)	(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***
	(0.001)	(0.001)	(0.050)	(0.080)	(0.003)	(0.004)
Interaction: Years since diagnosis at baseline with survey year						
0–3	−0.024**	−0.007	−0.229	0.216	−0.033	−0.072
	(0.011)	(0.009)	(0.667)	(0.617)	(0.040)	(0.044)
4–7	−0.017	0.004	−1.167	0.663	−0.033	−0.074
	(0.015)	(0.009)	(0.760)	(1.174)	(0.047)	(0.047)
8–12	−0.029	−0.047***	−0.427	−3.335*	0.048	−0.005
	(0.020)	(0.012)	(1.162)	(1.846)	(0.069)	(0.113)
13+	−0.043**	−0.010	1.741	1.559	−0.160**	−0.137***
	(0.021)	(0.014)	(1.426)	(2.093)	(0.070)	(0.013)
N	16329	22519	13614	7430	10815	5769

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

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

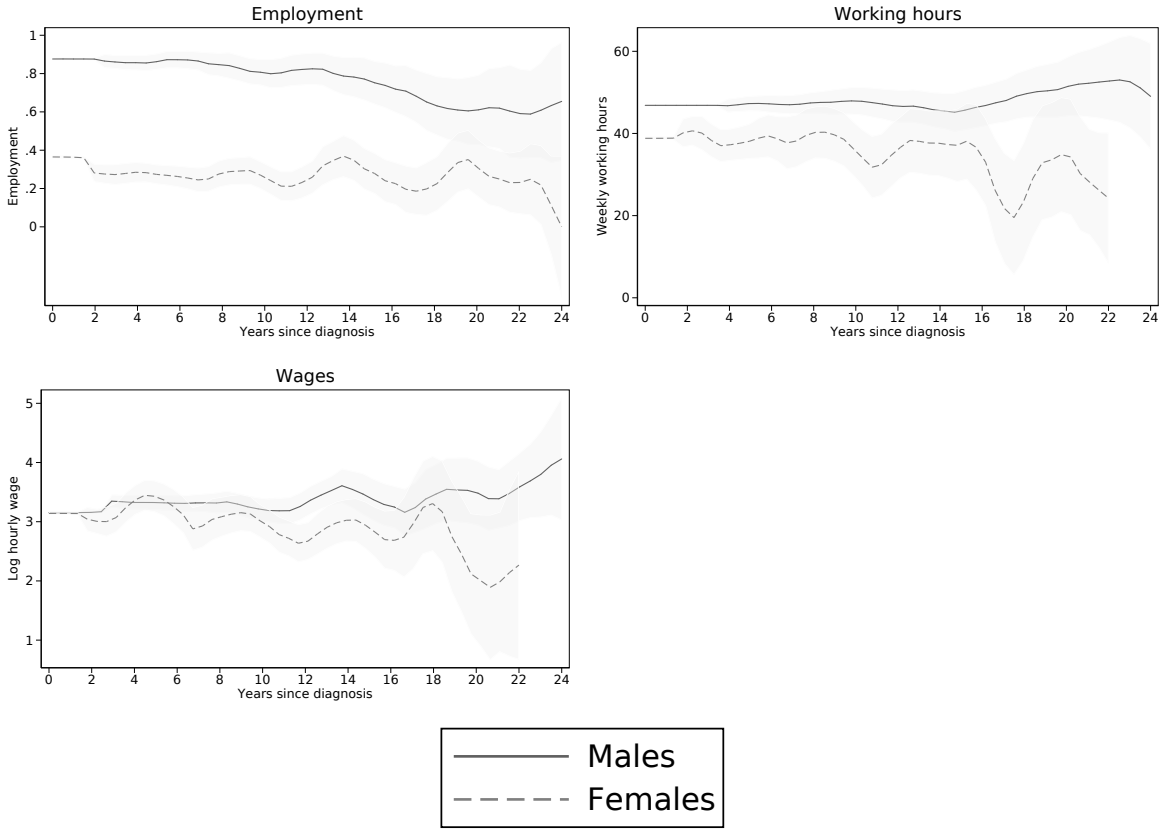
Table 6. **Biomarker results.**

	Employment	
	Males	Fema
Panel A: Diabetes (self-reported)		
Self-reported diabetes	−.057**	−.0
	(.025)	(.0
Panel B: Diabetes (biomarker)		
Biomarker diabetes (HbA1c $\geq$ 6.5)	−.015	−.0
	(.016)	(.0
Panel C: Self-reported and undiagnosed diabetes		
Self-reported diabetes (β_1)	−.059**	−.0
	(.028)	(.0
Undiagnosed diabetes (HbA1c $\geq$ 6.5) (β_2)	0.003	−.0
	(.018)	(.0
Panel D: HbA1c levels		
Self-reported diabetes	−.081*	−.0
	(.046)	(.0
HbA1c if $\geq$ 6.5	0.004	−.0
	(.005)	(.0
Self-reported diabetes × HbA1c if $\geq$ 6.5	0.004	0.0
	(.012)	(.0
N	2749	353

Notes
Results are based on community level fixed effects. Robust standard errors in parentheses. All models include variables for the level of education, age and age squared and year dummies to account for the multiple years of data collection for the third wave. The

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.