The Economic Costs of Diabetes in Mexico: A Labour Market Perspective using Longitudinal and Biomarker data

Till Seuring[1,2,3*], Pieter Serneels[3], Marc Suhrcke[4,5]

1 Research Group for Evidence-Based Public Health, Leibniz Institute for Prevention Research and Epidemiology (BIPS), Bremen, Germany

2 Health Sciences Bremen, Institute for Public Health and Nursing, University of Bremen, Bremen, Germany

3 School of International Development, University of East Anglia, Norwich, Norfolk, UK

4 Centre for Health Economics, University of York, York, UK

5 Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette/Belval, Luxembourg

* seuring@leibniz-bips.de

Abstract

Background 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 biomarker data we estimated the impact of self-reported diabetes and diabetes duration on employment probabilities, wages and working hours. We further explored how these effects differed for those aware and those unaware of their diabetes.

Methods and findings For the longitudinal analyses data from 11836 men and 13745 women between the age of 15 to 64 were taken from three waves of the Mexican Family Life Survey. We estimated a within-between effects model to account for unmeasured time-invariant confounders of diabetes incidence (705 cases) while simultaneously assessing differences between subjects with and without diabetes. The within-effects indicated a

reduction in the probability of being employed of 5.4 (standard error (SE) 2.5) percentage points (p.p.) for men and 6.0 (SE 2.4) p.p. for women with incident diabetes, but no effects on hours worked or wages. The between effects showed similar results. Employment probabilities fell gradually with each year since diagnosis (1.6 p.p, SE 0.6 for men; 0.9 p.p, SE 0.05 for women). Using cross-sectional biomarker data of glycated hemoglobin (HbA1c) from 2785 men and 3620 women, we observed 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, likely due to considerably better health in the undiagnosed population. A limitation of this study is its observational nature and the lack of a quasi-experimental study design, making it possible that the results may still be confounded.

Conclusion Diabetes in Mexico is mainly affecting people's employment probabilities, where both women and men experience important reductions. Given measurement error when using self-reported diabetes, these results cannot be extended to the considerable population with undiagnosed diabetes, likely because of a selection of people in worse health and with a longer diabetes duration into the diagnosed population. This suggests that earlier diagnosis and improved treatment of diabetes could prevent adverse health effects and related economic hardship in Mexico.

Author summary

STILL MISSING

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 problem for low- and middle-income countries (LMICs) as well as high-income countries (HICs), with over two-thirds of people with diabetes living in the developing world [1]. In Mexico, diabetes prevalence is estimated to have grown from 6.7% in 1994 to 14.4% in 2006 [2] and 15.8% in 2015. Diabetes has become the number one contributor to mortality [1], by increasing the

risk for heart disease and stroke, blindness, kidney disease and nerve problems, food ulcers and amputations [3]. However, via effective self-management of the disease, many if not all of the complications can be avoided [4, 5].

The observed increase in diabetes incidence has been attributed to a deterioration in diet and a reduction in physical activity [6, 7], while genetic predisposition among Mexicans with pre-Hispanic ancestry may also play a role [8]. The onset of diabetes has been occurring at an ever earlier age in Mexico [9], increasing the risk of complications occurring during the productive lifespan. Only a minority of patients in Mexico achieves adequate blood glucose control [2]. Moreover, diabetes in Mexico coexist with high levels of infectious diseases, exposing the health system to a 'double-disease burden' increasing the pressure to identify treatment priorities and use existing resources efficiently [10].

Despite the catastrophic impact of diabetes on health, its economic consequences, in particular in LMICs have received less attention, especially its effects on labour outcomes [11]. The latter have been studied predominantly in high-income countries, where substantial economic losses have been observed [12–18].]. Supplemtary Table 13 provides an overview of the existing literature in this field. For LMICs less evidence is available. One study exploited a natural experiment in China and found a significant reduction in income due to a recent diabetes diagnosis [19]. 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. ($p<0.1$) [20]. Most existing studies rely on instrumental variable (IV) estimation to address the potential endogeneity of diabetes using the genetic component of diabetes based on its family history as an instrument. 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 have not yet been used by our knowledge. Such unobservables, like for instance health endowments could adversely affect health as well as the propensity to develop type 2 diabetes in particular [21–23]; they may also affect labour outcomes—either directly through their effects on contemporaneous productivity [24], or indirectly by limiting educational attainment and human capital accumulation [25]. These unobservables thereby present a major source of a potential bias

that can be accounted for by the use of 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 [26]. 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 as those who self-report, and hence tend to be diagnosed, are in worse health than those undiagnosed, potentially leading to an overestimation of the economic effects of diabetes, in particular in populations with a large undiagnosed population, such as in many LMICs [27]. So far, however, little evidence exists on the economic impact according to diabetes severity and 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 the population unaware of their condition (i.e. the 'undiagnosed'). We used three waves of Mexican panel data, covering the period 2002–2012. Applying a within-between model, which models individual fixed effects and between effects separately, we accounted for time-invariant heterogeneity when assessing the impact of self-reported diabetes and time since diagnosis on labour outcomes. We also used rich and novel biomarker data from the most recent wave of data, we assessed the role of undiagnosed diabetes.

DATA

This paper uses 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. Data was collected on a wide range of social, demographic, economic and health characteristics [28]. Our samples were restricted to the working age population (15–64) and excluded pregnant women and those in school. 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). The first part of the analysis used all three waves and 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 [29], 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 if 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 and 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 86% of men reported some form of employment compared to 37% of women (see Table 1). Interestingly, men did not report considerably higher hourly wages than women but worked more hours per week. Men also more often worked 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.

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-Diaz

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			7.79	
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	994		25664	1666	

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

et al. [30], 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 likely type 1 diabetes while those above had likely type 2 diabetes. Nonetheless, because we cannot warranty that this is 100% accurate as it may be 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 (in the case of the within-coefficient, which only takes into account incident cases). Because we did not have information about the exact age at diagnosis for all diabetes cases in all three waves, the between coefficient may also stratify people with diabetes into the late onset group, even though they actually had early onset diabetes but only joined the sample after already having been diagnosed for several years. In the pooled data, which combines all three waves, diabetes was self-reported by 5% of men and 6% of women, respectively. 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 [2]. 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 measurement error associated with self-reported diabetes. This was done by also considering those with undiagnosed diabetes, i.e. the false negatives. The biometrically measured blood glucose value that allowed us to identify those with undiagnosed diabetes, was available for over 6000 respondents in the third wave.

Estimation strategy

To investigate the relationship between self-reported diabetes and three labour outcomes—employment, wages and weekly working hours—we estimated a within-between (WB) effects model that explicitly models both within and between effects. The within effects are identical to a fixed effects (FE) model, accounting for the potential bias introduced by time-invariant unobservables, providing an estimate of the effect for cases that received a diagnosis throughout the survey (705 incident cases compared to 970 non-changing diabetes cases in the used sample). Modeling the between effect allowed us to also use information from those that already had diabetes at baseline.

$$Y_{it} = \beta_0 + \beta_1(D_{it} - \overline{D}_i) + \beta_2(X_{it} - \overline{X}_i) + \beta_3\overline{D}_i + \beta_4\overline{X}_i + (u_i + e_{it}), \quad (1)$$

The within-effect is arrived at by 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 to the other time-varying covariates X_{it}. Cluster means of diabetes and of all other time-varying covariates were also included to capture the between effect. The error terms u_i and e_{it} capture the errors for the within and between variation, respectively. Y_{it} was a binary variable taking a value of 1 if respondent i reported being in employment at time t and 0 otherwise. Making use of the user written Stata command xthybrid WB model [31], we estimated WB effects applying a multilevel mixed-effects generalized linear model. For ease of interpretation we chose to estimate a linear probability model for the effects of diabetes

on employment. As a robustness check we also estimated binary within-between models using a logit link function and calculating marginal effects. These results suggested similar effects compared to the linear probability models and are presented in the Supplementary material (use actual reference here where to find them. currently towards the end of the supplementary material).

For the outcomes 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.

The regressions controlled for level of urbanization, education, state, marital status, the number of children below the age of 6 in the household, a quadratic age term, calendar year dummies as well as household wealth. We composed an indicator using principal component of household assets and housing following Filmer et al. (2001) [32]. The assets indicators 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 body mass index (BMI). While part of the effect of diabetes may be due to potential adverse effects of obesity, including BMI as a control variable in the model would have led to biased estimates if the diagnosis of diabetes also had an effect on BMI, which was likely to be the case [33, 34]. In general, control variables should not also be potential outcome variables [35], hence we similarly did not control for other chronic diseases that may have been caused by diabetes, such as hypertension or cardiovascular disease. Stata 15 was used for all analyses [36].

Labour outcomes and time since diagnosis

The chronic nature and irreversibility of diabetes provide good reason to explore the long term effects post diagnosis. To do this we estimated the following model:

$$Y_{it} = \beta_0 + \beta_1(Dyears_{it} - \overline{Dyears}_i) + \beta_2(X_{it} - \overline{X}_i) + \beta_3\overline{Dyears}_i + \beta_4\overline{X}_i + (u_i + e_{it}), \quad (2)$$

where $\beta_1 Dyears_{it}$ was continuous indicating years since the diagnosis was first reported. While 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 of time since diagnosis in Eq 2 and Eq 3, identification here relied on the presence of people without diabetes in the sample, for which diabetes duration did not increase.

We also considered a spline function that allowed for non-linear effects over time.

$$Y_{it} = \beta_0 + \beta_1(Dsplines_{it} - \overline{Dsplines_i}) + \beta_2(X_{it} - \overline{X}_i) + \beta_3\overline{Dsplines_i} + \beta_4\overline{X}_i + (u_i + e_{it}), \quad (3)$$

Based on visual inspection (Fig 1 on page 36) we chose three nodes located at 3, 7 and 12 years after diagnosis. The first three years should capture any immediate effects of the diagnosis, the years four to seven any effects during time of adaptation to the disease and the later terms the long term effects. 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 was not available for those who were not interviewed in the third round. A reported diagnosis in the year of the interview was counted as 'one year since diagnosis'.

Labour outcomes and biometrically measured diabetes

The biomarker analysis consisted of three steps. We first re-estimated Eq 4 to assess the relationship between self-reported diabetes with labour outcomes, but this time for the cross-sectional biomarker sample only, using the following specification:

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 X_i + c_i + v_i \quad (4)$$

where v_i were community fixed effects which reflected local unobserved characteristics, such as access to healthcare, poverty and unemployment in the community. 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 labour outcomes, using the following equation:

$$Y_i = \beta_0 + \beta_1 Dbio_i^d + \beta_2 X_i + v_i + u_i, \quad (5)$$

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

To estimate the effect of undiagnosed diabetes, we added self-reported diabetes and interacted it with the biomarker (Eq 6).

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i + \beta_3 Dsr_i * Dbio_i + \beta_4 X_i + v_i + u_i. \quad (6)$$

Note that the interaction term changes the interpretation of β_1 and β_2, with β_1 now representing the effect on those aware of their condition but with HbA1c levels below the diabetes threshold; while β_2 reflects the effect on those with undiagnosed diabetes, i.e. the respondents not self-reporting diabetes but with HbA1c levels equal to or above the threshold. The interaction term β_3 shows the effect for those with self-reported diabetes and HbA1c levels above the threshold.

We further investigated the effect of the severity of diabetes on labour outcomes, replacing $Dbio^d$ with $Dbio^c$, a variable that was 0 for HbA1c $< 6.5\%$ and took the actual value of HbA1c for those with an $HbA1c \geq 6.5\%$ (Eq 7). This allowed us to investigate the effect of a one percentage point increase in HbA1c levels for people with undiagnosed diabetes (β_2) as well as for those with self-reported diabetes above the diabetes threshold (β_3).

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i^c + \beta_3 Dsr_i * HbA1c_i + \beta_4 X_i + v_i + u_i. \quad (7)$$

Results

Labour outcomes and self-reported diabetes

The results of estimating Eq 1 in Table 2 indicated significant and substantial reductions in the probability of employment for men and women with self-reported diabetes. The overall similarity of between estimates suggests that the within effect is generalizable to the entire self-reporting diabetes population, i.e. not only for incidence cases. Additionally, it provides suggestive evidence that time-invariant unmeasured confounders may play a limited role. Employment probabilities were reduced by over 5 p.p. for both genders, translating into relative reductions of 14% for women and of 6% for men. There was no significant relationship between diabetes on one hand working hours and wages on the other, though the between estimates suggested that men with diabetes earned generally more than their

counterparts without diabetes (Columns 3–6 of Table 2). Note that wages are often seen as a proxy of on-the job productivity. Overall these results thus suggests effects at the extensive margin (employment), but not at the intensive margin (labour supply and productivity).

Dividing the diabetes population into early and late onset groups indicated that men, and potentially also women, saw their employment probabilities negatively affected by a diabetes onset later in life (Supplementary Table 7). For women, a particularly strong effect was also found for early diabetes onset. The between estimator suggested that men and women with diabetes were less likely to be employed at older ages, but not at a younger age. For working hours, we only found effects for the within-effects estimator for early onset, which may have been spurious due to low incidence. For wages, we found a positive effect of diabetes incidence on women. However, the within estimates for early onset cases again may be spurious. The between estimates show that especially older men with diabetes received higher wages than those without diabetes.

Table 2. Labour outcomes and self-reported diabetes

	Employment		Weekly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Diabetes (within)	-0.054^{**}	-0.060^{**}	-0.582	-1.990	0.063	0.074
	(0.025)	(0.024)	(1.501)	(2.513)	(0.067)	(0.160)
Diabetes (between)	-0.077^{***}	-0.066^{***}	-0.804	-1.032	0.097^{**}	-0.078
	(0.016)	(0.015)	(0.848)	(1.306)	(0.046)	(0.064)
Within=Between (p-value)	0.447	0.832	0.897	0.735	0.680	0.392
N	21388	27339	17616	9112	13828	7068

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

To assess whether diabetes affected 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. The incidence of diabetes only reduced the probability to work in agriculture for women, while we found no statistically significant effects for men. The between effects suggested that men with diabetes were less likely to be employed in agriculture, but more likely to be self-employed. Women with diabetes were less likely to be employed in agricultural and non-agricultural jobs (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 8). Interestingly, for male self-employment, incidence of 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. However, especially the results for early onset diabetes in women should be interpreted carefully due to limited number of diabetes cases, in particular for the selection into agriculture, where the logistic within-between model did not converge due to limited within-variation in early onset diabetes.

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

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Diabetes (within)	−0.007	−0.008	−0.042	−0.001	−0.022**	−0.030*
	(0.029)	(0.022)	(0.026)	(0.018)	(0.009)	(0.018)
Diabetes (between)	−0.032	−0.090***	0.044**	−0.049***	−0.008***	−0.012
	(0.020)	(0.014)	(0.019)	(0.012)	(0.003)	(0.011)
Within=Between (p-value)	0.470	0.002	0.008	0.026	0.173	0.391
N	20719	20719	20719	26575	26575	26575

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

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.

Figure 1 about here

Table 4 panel A shows the results of estimating Eq 2, which indicated that male and female employment probabilities fell every year, with a larger effect shown by the within coefficient. For females, the within coefficient also suggested a reduction in wages as diabetes progressed while the between coefficient showed no association for women and a relatively small positive association with male wages. Using diabetes onset groups, there was no evidence of an effect of diabetes duration for early onset groups (see Supplementary Table 9). However, again the within results for early onset should be interpreted with caution due to the limited number of diabetes incidence cases in this group, which also prohibited the estimation of any within effects of early diabetes onset duration for wages and working hours. The results also indicated that the effects found in Table 4 were driven mainly by

those with a diabetes onset after age 35.

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

	Employment		Weekly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Years since diagnosis (within)	−0.016***	−0.009*	0.185	0.115	−0.016	−0.067**
	(0.006)	(0.005)	(0.334)	(0.652)	(0.018)	(0.029)
Years since diagnosis (between)	−0.008***	−0.005***	−0.021	−0.140	0.010**	−0.010
	(0.002)	(0.001)	(0.107)	(0.125)	(0.005)	(0.009)
Within=Between (p-value)	0.139	0.482	0.560	0.696	0.165	0.061
Panel B: splines						
Years since SR diagnosis						
Within-effects						
0–3	−0.013	−0.018	0.708	2.953	−0.005	0.047
	(0.014)	(0.016)	(0.857)	(2.700)	(0.054)	(0.124)
4–7	−0.011	−0.002	0.215	−2.517	−0.032	−0.131
	(0.014)	(0.014)	(0.761)	(1.752)	(0.046)	(0.101)
8–12	0.003	−0.003	−1.153	1.144	−0.009	−0.053
	(0.021)	(0.014)	(1.252)	(1.635)	(0.065)	(0.061)
13+	−0.039***	−0.015	0.720	0.184	−0.007	−0.096***
	(0.014)	(0.010)	(0.943)	(1.414)	(0.057)	(0.037)
Between effects						
0–3	0.005	0.010	−0.030	0.204	0.042**	0.032
	(0.008)	(0.009)	(0.465)	(0.653)	(0.020)	(0.032)
4–7	−0.003	−0.034	−0.444	0.727	−0.025	−0.034
	(0.021)	(0.023)	(1.195)	(1.471)	(0.052)	(0.065)
8–12	−0.055**	0.004	1.110	−1.525	−0.031	−0.109
	(0.023)	(0.024)	(1.367)	(1.647)	(0.060)	(0.075)
13+	0.004	−0.004	−0.432	−0.282	0.034*	0.015
	(0.008)	(0.003)	(0.476)	(0.245)	(0.019)	(0.017)
Within=Between (p-value)	0.028	0.404	0.507	0.597	0.781	0.083
Panel C: dummies						
Within-effects						
0–3	0.005	−0.007	0.352	17.309*	0.223	−0.447
	(0.052)	(0.059)	(3.123)	(9.975)	(0.186)	(0.549)
4–7	−0.031	−0.049	2.860	10.878	0.047	−0.568
	(0.042)	(0.050)	(2.664)	(9.504)	(0.127)	(0.544)
8–12	−0.066	−0.026	−0.709	13.733	−0.133	−0.873*
	(0.063)	(0.059)	(4.181)	(9.695)	(0.207)	(0.521)
13+	−0.134	−0.062	−3.379	13.309	0.164	−0.882**
	(0.098)	(0.068)	(4.715)	(9.239)	(0.284)	(0.446)
Between effects						
0–3	−0.002	0.033	−0.760	0.322	0.109*	0.023
	(0.027)	(0.026)	(1.370)	(1.944)	(0.057)	(0.091)
4–7	0.036	0.016	−1.004	3.157	0.154**	0.082
	(0.030)	(0.042)	(1.968)	(3.208)	(0.076)	(0.134)
8–12	−0.128**	−0.199***	2.231	−0.792	−0.069	−0.197
	(0.065)	(0.055)	(3.111)	(4.043)	(0.131)	(0.161)
13+	−0.187***	−0.076*	−1.443	−4.127	0.196	−0.342**
	(0.050)	(0.042)	(2.895)	(3.492)	(0.142)	(0.173)
Within=Between (p-value)	0.397	0.060	0.672	0.242	0.721	0.745
N	16298	22427	10771	5746	13583	7391

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

The non-linear results for the spline function and dummy variable approach are presented in panels B and C, respectively. They suggest that the main adverse effects appeared after a prolonged time of living with diabetes; i.e. after more than seven years since diagnosis. The

same was true for female wages. The lack of a statistically significant effect for the earlier years of diabetes duration may have been due to a reduction in statistical efficiency, reduced by the separation into duration groups and into within and between variation. Re-estimating the specifications with a random effects model that combined both types of variation into one estimate showed that, at least for the models with dummies, there was a more or less immediate reduction in employment probabilities that became stronger the longer a person had diabetes (see Supplementary Table 10). Note that we did not estimate models splitting diabetes in early and late onset groups, as this implied strong reductions in statistical power.

Cross-sectional biomarker analysis

As reported in Supplementary Table 11, 18% of the observations in the biomarker sample were false negatives, i.e. undiagnosed. Further, 2% were false positives, though the latter may have included cases that received a diabetes diagnosis and managed to reduce their HbA1c to non-diabetes levels via medication and/or lifestyle changes [37]. Overall 80% of the self-reports were consistent with the biomarker data. Comparing the health status and diabetes risk factors of the diagnosed and undiagnosed diabetes populations suggested that those with self-reported diabetes were older and in worse health, both objectively and subjectively compared to those undiagnosed. This suggests a selection into the diagnosed group based on the severity and potentially duration of diabetes. This, if the adverse effects of diabetes are due to its health impact we would suspect worse labour market outcomes for the diagnosed compared to the undiagnosed population.

Table 6 presents the results from estimating Eq 4 – 7. 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 using diabetes defined by the biomarker instead of self-reported diabetes, in particular for men. Results in Panel C were obtained from estimating Eq 6 and indicate an absence of a (statistically significant) negative relationship between undiagnosed diabetes (expressed in the 'Biomarker diabetes but not self-reported' coefficient) and labour outcomes. The coefficients for the interaction term were negative throughout, though only statistically significant in the case of and female working hours and male wages.

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	0.807	0.875	0.012	0.241	0.331	0.002
Hourly wage	35.931	30.670	0.120	36.092	32.638	0.550
Usual weekly workinghours	44.341	46.682	0.104	34.708	39.681	0.046
Age	53.162	44.720	0.000	53.167	44.449	0.000
Any medical insurance	0.677	0.599	0.033	0.733	0.643	0.002
City of 2,5oo-15,000	0.096	0.107	0.643	0.112	0.112	0.994
City of 15,000-100,000	0.135	0.096	0.098	0.087	0.090	0.884
City of >100,000	0.331	0.297	0.325	0.294	0.333	0.185
Married	0.719	0.643	0.031	0.633	0.569	0.035
Number of children (<6) in household	0.950	1.122	0.073	0.960	1.257	0.001
Indigenous group	0.158	0.211	0.079	0.195	0.207	0.626
Primary	0.479	0.434	0.238	0.626	0.465	0.000
Secondary	0.212	0.231	0.554	0.132	0.233	0.000
High school	0.062	0.131	0.003	0.037	0.117	0.000
Higher education	0.139	0.113	0.288	0.030	0.073	0.003
Wealth index	−0.174	0.117	0.000	0.012	0.103	0.157
Subjective health						
very good	0.023	0.094	0.000	0.012	0.054	0.001
good	0.212	0.434	0.000	0.180	0.367	0.000
fair	0.619	0.442	0.000	0.643	0.528	0.000
bad	0.135	0.026	0.000	0.155	0.048	0.000
very bad	0.012	0.004	0.187	0.010	0.004	0.246
Glycated hemoglobin (HbA1c)	9.037	8.533	0.004	8.979	8.680	0.049
Hypertension (self-reported)	0.262	0.074	0.000	0.397	0.150	0.000
Blood pressure						
Systolic	136.688	130.506	0.000	136.070	122.835	0.000
Diastolic	84.677	82.063	0.003	84.495	79.689	0.000
Heart disease	0.035	0.007	0.004	0.050	0.024	0.021
BMI	28.868	28.311	0.135	30.640	29.778	0.032
Obese	0.338	0.311	0.440	0.469	0.431	0.225

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

Table 6. Biomarker results.

	Employment		Weekly work hours		Log hourly wages	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Panel A: Diabetes (self-reported)						
Self-reported diabetes	−.055**	−.050**	−.308	−.323	−.026	−.005
	(.026)	(.024)	(1.321)	(2.190)	(.066)	(.105)
Panel B: Diabetes (biomarker)						
Biomarker diabetes (HbA1c $\geq$ 6.5)	−.012	−.032*	−.060	1.067	−.008	−.041
	(.016)	(.018)	(.844)	(1.470)	(.045)	(.071)
Panel C: Interacting self-reported and biomarker diabetes						
Self-reported diabetes but tested negative (β_1)	−.049	−.027	1.547	7.417	0.250	−.025
	(.059)	(.051)	(3.446)	(4.669)	(.205)	(.232)
Biomarker diabetes but not self-reported (HbA1c $\geq$ 6.5) (β_2)	0.005	−.020	0.174	2.304	0.015	−.049
	(.018)	(.020)	(.960)	(1.639)	(.050)	(.078)
Self-reported diabetes and biomarker diabetes (β_3)	−.011	−.015	−2.255	−11.484**	−.319	0.059
	(.065)	(.059)	(3.772)	(5.423)	(.220)	(.265)
All self-reported ($\beta_1 + \beta_3$)	−.060**	−.042	−.708	−4.067	−.069	0.034
	(.029)	(.031)	(1.571)	(2.813)	(.084)	(.137)
Panel D: HbA1c levels						
Self-reported diabetes	−.067	−.035	1.471	4.974	0.157	−.017
	(.072)	(.041)	(2.619)	(3.654)	(.119)	(.186)
HbA1c if $\geq$ 6.5	0.001	−.003	−.005	0.235	0.002	−.005
	(.002)	(.002)	(.104)	(.193)	(.005)	(.008)
Self-reported diabetes × HbA1c if $\geq$ 6.5	0.001	0.000	−.207	−.852**	−.023	0.005
	(.007)	(.005)	(.294)	(.424)	(.014)	(.022)
N	2785	3620	2302	1142	1803	882

Notes Results are based on community level fixed effects. 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 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). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To explore whether the adverse effects increased with higher HbA1c levels, we estimated Eq 7. The results in panel D support the findings from panel C, showing negative coefficients for a 1 percentage point increase in HbA1c for those with self-reported diabetes and HbA1c levels in the diabetes range, again, however, only statistically significant for male wages and female working hours. For undiagnosed diabetes, we again found no effects.

Discussion

Diabetes is now one of the most common chronic diseases in low- and middle income countries, as well as high-income countries, 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 population (18%) is unaware of their condition.

Overall, the paper found evidence for adverse effects of self-reported diabetes on the probability of being employed, confirming earlier findings for Mexico that used cross-sectional information. But, these new results also suggest a comparatively larger impact of diabetes on female employment probabilities. The evidence also points towards the main effects being driven by those with a diabetes onset at a relatively later state, consisting most likely of people with type 2 diabetes. Analyses of the long term impact indicated that the employment probability fell gradually in the years following diagnosis. The results for the non-linear models were less clear, potentially due to reductions in statistical power, but suggest that adverse effects became stronger with time since diagnosis. The linear effect contrasts with estimates for the USA, where such an effect seemed to be absent, however allowing for non-linearity revealed falling employment probabilities after 11 to 15 years for females and after 2-5 years for males [16]. For working hours and wages, our results were more ambiguous with adverse effects being observed for time since diagnosis on female wages only.

In summary, most of the adverse effects of diabetes were found at the extensive margin, mainly affecting the probability of employment, rather than the intensive margin, i.e. working

hours or wages.

Overall, a relationship of diabetes with working hours or wages is mostly absent, except for the between estimator in the case of wages for men. Contrasting the between and within results suggests that this significant result may have arisen from individual selection. Although any explanation at this point is speculative, it may be that higher paid men, who also tend to be more educated, 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 lack access to quality diabetes care, making it more likely that they develop severe complications earlier [37]. They are also more likely to be in informal employment and low skilled jobs, with less job security and are thus more prone to being laid off to be replaced with healthier workers.

Because self-reported diabetes may not adequately represent the entire diabetes population if the share of undiagnosed diabetes is relatively large and those undiagnosed differ from those diagnosed, estimates based on self-reported diabetes are likely to be biased. Our results using the cross-sectional biomarker data suggest indeed that those with undiagnosed diabetes were significantly healthier and younger, among others. Consequently, diabetes based on biomarkers was found to be less related to reduced employment, compared to self-reported diabetes, in particular for men. Further analysis showed that this was due to the absence of any association between undiagnosed diabetes and employment. These results are similar to those found for the USA, where no statistically significant relationship was observed between undiagnosed diabetes on employment, while a significant effect of diagnosed diabetes was observed [17]. Our results further indicate that the difference in the employment effects of diagnosed and undiagnosed diabetes was not mediated by current HbA1c levels, similar to findings for Mexican-Americans in the USA, where employment outcomes were unrelated to higher HbA1c levels [14]. This may stem from HbA1c levels primarily being informative for the last three months, and not being the only indicator for the severity of diabetes. Overall, it seems that general health differences related to a longer diabetes duration and selection into the diagnosed population, for instance based on emerging diabetes related health problems, seem to be driving the adverse economic effects among those with a self-reported diagnosis.

Our study had several limitations. Given the use of observational data and the lack of

a suitable quasi-experimental study design, we cannot claim that our findings are causal. While the within-coefficient accounts 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 play a role. Existing studies that looked at this particular direction of causality, however, not found strong evidence for an effect of employment status on diabetes [38, 39], though they were carried out in high-income countries. For 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, thereby excluding those that dropped out of the sample prior to the third wave.

Despite these limitations, our findings bear important implications. First, the impact of self-reported diabetes on labor outcomes in Mexico seems mostly limited to its effect on employment probabilities, though there is some indication that it could also reduce wages over time for women. 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 dropout less costly. Other evidence suggests that women with diabetes are in worse metabolic health compared to men when they cross the diabetes threshold [40], 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 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 effect of self-reported diabetes should be limited to this specific part of the population, in particular in environments where the share of undiagnosed diabetes is high, as is the case for most low- and middle income countries.

While we find no effect for undiagnosed diabetes in our cross section analysis, further inquiry over time is needed. The large proportion of previously undiagnosed cases 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. Further analysis is also needed to

explain why the adverse economic effects are so large for women. Ultimately, prevention of diabetes is of high importance. Taxation of sugar sweetened beverages may be one promising way forward [41], though their long-term effectiveness remains unclear. Given the established links between early life heath and later life incidence of diabetes as well as other chronic diseases [22, 23, 42],investments in maternal and child health may be particularly attractive to address non-communicable and communicable diseases at the same time.

Conflicts of interest

We report no conflicts of interest.

Ethical approval

No ethical approval was sought for this study that used publicly available data.

Appendix

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 [43]. 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 false negatives. False positives are deemed less of a problem since incentives to report diabetes when one does not have it seem to be very limited—although we cannot exclude this. A study from China finds that the vast majority (98%) of those who self-report diabetes are tested positive for diabetes, while only a minority of those who are tested positive for diabetes (40%) actually self-report the disease [44]. Our data showed a similar pattern, with a low proportion (3%) of the respondents being tested negative while self-reporting diabetes, while the majority of those who are tested positive (68%) do not self-report diabetes.

We used the above information to infer the "true" diabetes status for those with incon-

sistent reports. For respondents present in all three waves, we corrected inconsistencies as reported in Supplementary Table 12. 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 12) and that the person never had received a diagnosis. If a diabetes diagnosis was however 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 12). 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 12), given that most diabetes self-reports tend to be correct.

Table 12 about here

We then tested if those 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 12) and 49 with one self-report of diabetes (from lines 1 and 2 in Table 12)). Fig 2 illustrates the difference between both groups and suggests that indeed those with two self-reports of diabetes are much more likely to have HbA1c values above the diabetes threshold. A t-test comparing the mean HbA1c for the two groups indicates that those with two self-reports also have 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% have 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.

Alternatively we also test if using an alternative strategy, i.e. assuming that everybody who reported a diabetes diagnosis once had diabetes in any later wave, would lead to different estimation results. We do not find this to be the case and find only minor differences in the point estimates of the coefficients (results available on request).

Figure 2 about here

Table 13 about here

References

1. International Diabetes Federation. Diabetes Atlas. 7th ed. International Diabetes Federation; 2015.

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

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

4. Lim EL, Hollingsworth KG, Aribisala BS, Chen MJ, Mathers JC, Taylor R. Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. Diabetologia. 2011;54(10):2506–2514. doi:10.1007/s00125-011-2204-7.

5. Gregg EW, Chen H, Wagenknecht LE, Clark JM, Delahanty LM, Bantle J, et al. Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes. Journal of the American Medical Association. 2012;308(23):2489. doi:10.1001/jama.2012.67929.

6. Barquera S, Hernandez-Barrera L, Tolentino ML, Espinosa J, Ng SW, Rivera JA, et al. Energy Intake from Beverages Is Increasing among Mexican Adolescents and Adults. Journal of Nutrition. 2008;138(12):2454–2461. doi:10.3945/jn.108.092163.

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

8. Williams AL, Jacobs SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, Márquez-Luna C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature. 2013;506(7486):97–101. doi:10.1038/nature12828.

9. Bello-Chavolla OY, Rojas-Martinez R, Aguilar-Salinas CA, Hernández-Avila M. Epidemiology of diabetes mellitus in Mexico. Nutrition Reviews. 2017;75(suppl 1):4–12. doi:10.1093/nutrit/nuw030.

10. Gutiérrez-delgado C, Guajardo-barrón V. The double burden of disease in developing countries: The Mexican experience. In: Advances in Health Economics and Health Services Research. vol. 21; 2009. p. 3–22.

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

12. Brown HS, Pagán JA, Bastida E. The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit. Health Economics. 2005;14(5):537–544. doi:10.1002/hec.942.

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

14. Brown HS, Perez A, Yarnell LM, Pagan Ja, Hanis CL, Fischer-Hoch SP, et al. Diabetes and employment productivity: does diabetes management matter? American Journal of Managed Care. 2011;17(8):569–576.

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

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

17. Minor T, MacEwan JP. A comparison of diagnosed and undiagnosed diabetes patients and labor supply. Economics & Human Biology. 2016;20:14–25. doi:10.1016/j.ehb.2015.10.003.

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

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

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

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

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

23. Li Y, He Y, Qi L, Jaddoe VW, Feskens EJM, Yang X, et al. Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood. Diabetes. 2010;59(10):2400–2406. doi:10.2337/db10-0385.

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

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

26. Herrington WG, Alegre-Díaz J, Wade R, Gnatiuc L, Ramirez-Reyes R, Hill M, et al. 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. 2018;8587(18):1–9. doi:10.1016/S2213-8587(18)30050-0.

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

28. Rubalcava L, Teruel G. User's Guide for the Mexican Family Life Survey Third Round; 2013.

29. Crimmins E, McDade T, Rubalcava L, Seeman T, Teruel G, Thomas D. Health of the Mexican population: Results from the Mexican Family Life Survey (MxFLS); 2015.Available from: http://gero.usc.edu/CBPH/files/4{_}30{_}2014{_}PAA/Thomas{_}Health{_}of{_}the{_}Mexican{_}population{_}in{_}MxFLS.pdf.

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

31. Schunck R, Perales F. Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. Stata Journal. 2017;17(1):89–115. doi:The Stata Journal.

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

33. Slade AN. Health Investment Decisions in Response to Diabetes Information in Older Americans. Journal of Health Economics. 2012;31(3):502–520.

34. De Fine Olivarius N, Siersma VD, Køster-Rasmussen R, Heitmann BL, Waldorff FB. 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. 2015;10(4):1–14. doi:10.1371/journal.pone.0122219.

35. Angrist JD, Pischke JS. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press; 2009.

36. StataCorp. Stata Statistical Software: Release 15; 2017.

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

38. Bergemann A, Grönqvist E, Gudbjörnsdottir S. The effects of job displacement on the onset and progression. Netspar Discussion Paper. 2011;(25).

39. Schaller J, Stevens AH. Short-run effects of job loss on health conditions, health insurance, and health care utilization. Journal of Health Economics. 2015;43:190–203. doi:10.1016/j.jhealeco.2015.07.003.

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

41. Colchero MA, Popkin BM, Rivera JA, Ng SW. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. British Medical Journal. 2016;352:h6704. doi:10.1136/bmj.h6704.

42. Hanson MA, Gluckman PD, Ma RC, Matzen P, Biesma RG. Early life opportunities for prevention of diabetes in low and middle income countries. BMC Public Health. 2012;12(1):1025. doi:10.1186/1471-2458-12-1025.

43. Zajacova A, Dowd J, Schoeni RF, Wallace RB. Consistency and precision of cancer reporting in a multiwave national panel survey. Population Health Metrics. 2010;8(1):20. doi:10.1186/1478-7954-8-20.

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

45. Zhang X, Zhao X, Harris A. Chronic Diseases and Labour Force Participation in Australia. Journal of Health Economics. 2009;28(1):91–108. doi:10.1016/j.jhealeco.2008.08.001.

Supplementary material

Early versus late onset of diabetes

Table 7. Labour outcomes and self-reported diabetes by diabetes onset.

	Employment		Weekly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Early onset (within)	0.133	−0.206**	14.712*	−18.636*	−0.523	0.388***
	(0.176)	(0.086)	(8.370)	(9.668)	(0.340)	(0.057)
Late onset (within)	−0.059**	−0.048*	−1.008	−1.322	0.079	0.067
	(0.025)	(0.025)	(1.513)	(2.553)	(0.067)	(0.163)
Early onset (between)	0.016	−0.052	1.483	−2.640	−0.102	0.286
	(0.037)	(0.047)	(2.581)	(4.034)	(0.099)	(0.232)
Late onset (between)	−0.087***	−0.067***	−1.077	−0.869	0.121**	−0.111*
	(0.017)	(0.016)	(0.895)	(1.375)	(0.050)	(0.066)
N	21388	27339	13828	7068	17616	9112

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 8. 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 (within)	0.030	−0.105	−0.225	−0.068	0.328**	−0.027
	(0.216)	(0.074)	(0.139)	(0.047)	(0.161)	(0.048)
Late onset (within)	−0.008	0.007	−0.002	−0.019**	−0.053**	−0.030
	(0.029)	(0.018)	(0.022)	(0.009)	(0.026)	(0.019)
Early onset (between)	0.011	−0.050	−0.078***	−0.006	0.087	0.005
	(0.055)	(0.043)	(0.030)	(0.007)	(0.056)	(0.029)
Late onset (between)	−0.037*	−0.048***	−0.091***	−0.009***	0.039*	−0.014
	(0.022)	(0.012)	(0.015)	(0.003)	(0.020)	(0.011)
N	20719	26575	20719	26575	20719	26575

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

	Employment		Monthly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Early onset (within)	−0.009	0.002				
	(0.017)	(0.013)				
Late onset (within)	−0.012**	−0.007*	0.086	0.308	−0.006	−0.059***
	(0.005)	(0.004)	(0.275)	(0.504)	(0.013)	(0.022)
Early onset (between)	0.011***	−0.016***	−0.016	0.029	0.187	−0.401
	(0.004)	(0.005)	(0.012)	(0.051)	(0.333)	(0.896)
Late onset (between)	−0.009***	−0.005***	−0.060	−0.144	0.011**	−0.010
	(0.002)	(0.002)	(0.117)	(0.129)	(0.006)	(0.009)
N	16308	22450	13592	7394	10778	5748

Notes The within estimator for the effects of early onset diabetes on wages and working hours could not be estimates 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10. Relationship between self-reported years since diagnosis and employment probabilities using continuous duration and duration splines (random effects).

	Employment		Weekly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Panel A: linear effect						
Years since diagnosis	−0.007***	−0.004***	0.039	−0.130	0.010**	−0.009
	(0.002)	(0.001)	(0.102)	(0.127)	(0.005)	(0.008)
Panel B: splines						
Years since SR diagnosis						
0–3	−0.008	−0.015**	−0.035	0.507	0.038**	0.034
	(0.006)	(0.006)	(0.346)	(0.614)	(0.017)	(0.029)
4–7	0.001	0.004	0.242	−0.570	−0.032	−0.048
	(0.011)	(0.011)	(0.665)	(1.062)	(0.032)	(0.052)
8–12	−0.008	0.002	−0.116	−0.080	−0.003	−0.074
	(0.015)	(0.011)	(0.855)	(1.098)	(0.041)	(0.050)
13+	−0.012	−0.004	0.035	−0.339	0.029	0.011
	(0.008)	(0.003)	(0.410)	(0.241)	(0.018)	(0.017)
Panel C: dummies						
0–3	−0.036*	−0.041**	−0.821	1.091	0.134**	0.021
	(0.021)	(0.021)	(1.154)	(1.826)	(0.054)	(0.083)
4–7	−0.014	−0.056**	0.877	1.200	0.093	−0.003
	(0.022)	(0.023)	(1.375)	(2.530)	(0.059)	(0.118)
8–12	−0.069*	−0.043	0.427	0.302	−0.070	−0.148
	(0.037)	(0.030)	(2.288)	(2.995)	(0.101)	(0.117)
13+	−0.121***	−0.043	−0.568	−2.104	0.242*	−0.279*
	(0.045)	(0.031)	(2.280)	(3.088)	(0.126)	(0.153)
N	16308	22450	13592	7394	10778	5748

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

Table 11. 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

Table 12. 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 no diabetes	44
6	Diabetes self-report in 2005, but not in 2009. Not in survey in 2002	Has no diabetes	23

Table 13. Studies estimating the relationship between diabetes and labour market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
China	2009, 2011	Employed population	Panel	HbA1c	Find a significant reduction of 16.3 % in income for those with a recent diagnosis in China.	NA	Use difference-in-difference model, exploiting a recent diagnosis of diabetes as a result of biomarker collection within the used survey, as a natural experiment to measure how income developed between those who were newly diagnosed and those without diabetes in the years following diagnosis	[19]
Mexico	2005	Working age population	Cross-section	Self reported	A significant ($p<0.01$) reduction in employment probabilities for males by about 10 % points and for females by about 4.5 % points ($p<0.1$)	Diabetes exogenous for men and women based on Hausman test ($p>.10$)	Probit and bivariate probit model using parental diabetes as IV	[20]

Table 13. Studies estimating the relationship between diabetes and labour market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	1996-1997	Elderly population of Mexican Americans close the Mexican border	Cross-section	Self reported	Significant adverse relationship, with 7 % points lower employment rates for men - for women, the negative relationship becomes insignificant when using instrumental variable (IV) estimation	Diabetes endogenous for women but not men based on Hausman test	Bivariate probit	[12]
USA	2008	Mexican-American working age adults	Cross-section	HbA1c levels	Find a negative relationship between HbA1c levels and the probability of employment as well as male wages. No effects found for women.	Exogeneity assumed	Probit and Heckman selection model	[14]
USA	2006	Women 20 - 65	Cross-section	Self reported	Exogenous: 25.2 % points less likely to be employed, endogenous: 45.1 % points less likely to be employed.	Self-reported diabetes endogenous and estimates upward biased compared to IV estimates	Probit and Heckman selection model; unclear which model is used for IV estimates	[15]

Table 13. Studies estimating the relationship between diabetes and labour market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	1979 - 2010	Follows young adults in 1979 throughout their adult life	Panel	Self-reported year of diagnosis	Average reduction of employment probability of 28 % points for men and 36 % points for women; employment probabilities decline shortly after diagnosis for men and after about 10 years for women, while wages are not affected by the duration of diabetes	Exogeneity assumed	Uses sibling and job fixed effects model (no individual fixed effects) using logit model for selection into employment and ordinary least squares for wages	[16]

Table 13. Studies estimating the relationship between diabetes and labour market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	2001 - 2008	Men and women 18 - 65	Panel	Self-reported and HbA1c levels for subsample	No statistically significant relationship between undiagnosed diabetes and the probability of employment. Self-reported diabetes significantly related with lower employment probabilities for men (-11 % points) and women (-19 % points). Using only biomarker information (HbA1c >6.4 %), statistically significant reductions in employment probabilities for men (-8.3 % points) and women (-11 % points). No significant effects of undiagnosed diabetes on hours worked. Increase in HbA1c by 1 % point related to 1.3 % points lower employment probabilities for men. No effect for women.	Exogeneity assumed	Probit model for binary outcomes, OLS for continuous outcomes; all applied to pooled data	[17]

Table 13. Studies estimating the relationship between diabetes and labour market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
Canada	1998	Men and women 15 - 64	Cross-section	Self reported	For men: Exogenous 19 % points less likely to be employed; endogenous: not significant and positive; test indicates endogeneity For women: Exogenous: 17 % points less likely to be employed; endogenous: not significant and positive and test indicates exogeneity	Diabetes endogenous for men, resulting in upwards biased estimates; exogenous for women	Instrumental variable strategy using bivariate probit model and family history of diabetes as the instrument	[18]
Australia	1999 - 2000	Men and women age >24	Cross-section	Self reported	Reduced labour market participation for men (-7.1 % points) and women (-9 % points) as a result of diabetes, with the effects appearing overstated (-10.8 % points for men and -10 % points for women) if the endogeneity of diabetes is unaccounted for	Overestimation if endogeneity unaccounted for	Endogenous multivariate probit model	[45]

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

Notes The dashed lines show 95% confidence intervals.

Figure 2. Kernel density of HbA1c values for those with one inconsistent and two inconsistent reports.

Logit results

Table 14. Labour outcomes and self-reported diabetes (using logistic regression for employment models).

	Employment		Weekly work hours		Log hourly wages	
	Males	Females	Males	Females	Males	Females
Diabetes (within)	−0.051**	−0.062**	0.063	0.074	−0.582	−1.990
	(0.025)	(0.026)	(0.067)	(0.160)	(1.501)	(2.513)
Diabetes (between)	−0.067***	−0.067***	0.097**	−0.078	−0.804	−1.032
	(0.013)	(0.016)	(0.046)	(0.064)	(0.848)	(1.306)
Within=Between (p-value)	0.576	0.854	0.680	0.392	0.897	0.735
N	21388	27339	13828	7068	17616	9112

Notes Marginal effects presented in the employment models. 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 15. Selection into types of work and self-reported diabetes (logistic regression).

	Males			Females		
	Non-agric.	Agric.	Self-employed	Non-agric.	Agric.	Self-employed
Diabetes (within)	−0.004	−0.009	−0.037*	0.004	−0.026***	−0.027*
	(0.029)	(0.022)	(0.022)	(0.024)	(0.010)	(0.016)
Diabetes (between)	−0.027	−0.078***	0.031**	−0.062***	−0.011**	−0.007
	(0.020)	(0.017)	(0.014)	(0.016)	(0.005)	(0.009)
Within=Between (p-value)	0.525	0.012	0.007	0.022	0.091	0.275
N	20719	20719	20719	26575	26575	26575

Notes Marginal effects. 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 16. Labour outcomes and self-reported diabetes by diabetes onset (using logistic regression for employment models).

	Employment		Log hourly wages		Weekly work hours	
	Males	Females	Males	Females	Males	Females
Early onset (within)	0.153	−0.219**	−0.523	0.388***	14.712*	−18.636*
	(0.178)	(0.094)	(0.340)	(0.057)	(8.370)	(9.668)
Late onset (within)	−0.056**	−0.049*	0.079	0.067	−1.008	−1.322
	(0.025)	(0.027)	(0.067)	(0.163)	(1.513)	(2.553)
Early onset (between)	0.039	−0.055	−0.102	0.286	1.483	−2.640
	(0.053)	(0.051)	(0.099)	(0.232)	(2.581)	(4.034)
Late onset (between)	−0.076***	−0.069***	0.121**	−0.111*	−1.077	−0.869
	(0.014)	(0.017)	(0.050)	(0.066)	(0.895)	(1.375)
N	21388	27339	13828	7068	17616	9112

Notes Marginal effects for employment models. 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 17. Selection into types of work and self-reported diabetes by diabetes onset (logistic regression).

	Non-agric.		Agriculture		Self-employed	
	Males	Females	Males	Females	Males	Females
Early onset (within)	0.023	−0.117	−0.269**	−0.263***	0.355**	−0.034
	(0.205)	(0.087)	(0.108)	(0.014)	(0.156)	(0.065)
Late onset (within)	−0.005	0.015	−0.002	−0.023**	−0.046**	−0.027*
	(0.030)	(0.024)	(0.022)	(0.010)	(0.022)	(0.016)
Early onset (between)	0.020	−0.062	−0.172***	−0.236***	0.070*	0.011
	(0.057)	(0.048)	(0.063)	(0.025)	(0.042)	(0.031)
Late onset (between)	−0.032	−0.062***	−0.073***	−0.010*	0.026*	−0.009
	(0.022)	(0.017)	(0.018)	(0.005)	(0.014)	(0.009)
N	20719	26575	20719	26575	20719	26575

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