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

Till Seuring1,2,3\*, Pieter Serneels3, Marc Suhrcke4 , 5

1. Research Group for Evidence-Based Public Health, Leibniz Institute for Prevention

Research and Epidemiology (BIPS), Bremen, Germany

1. Health Sciences Bremen, Institute for Public Health and Nursing, University of Bremen,

Bremen, Germany

1. School of International Development, University of East Anglia, Norwich, Norfolk, UK
2. Centre for Health Economics, University of York, York, UK
3. Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette/Belval, Luxembourg

\* seuring@leibniz-bips.de

# Abstract

There is at best scarce evidence on the economic consequences of diabetes, especially in a context where diabetes often remains undiagnosed, as is the case in low- and middleincome countries. We estimated the impact of self-reported diabetes on labour outcomes applying a within-between effects model to Mexican panel data and self-reported diabetes, allowing us to account for potential unmeasured time-invariant confounders for incident diabetes cases while simultaneously assessing general differences between those with diabetes and without. We also cross-sectionally explored the role of undiagnosed diabetes using biomarker information to assess the impact of measurement error when using self-reported data. We found a reduction of 5 percentage points in the probability of being employed due to self-reported diabetes but no impact on wages or hours worked. The employment probability fell gradually with time since diagnosis. In the biomarker analysis we observed that 18% of all observations were false negatives (undiagnosed), i.e. did not report diabetes but exhibited glycated hemoglobin (HbA1c) levels above the clinical diabetes threshold.

However, regression analysis revealed that there was no effect of diabetes on labour outcomes for undiagnosed women or men. The results highlight the importance of the economic impact of diabetes and the heterogeneity in the population with diabetes that suggests that an earlier diagnosis may help to reduce the employment burden of diabetes.

# 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 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, since only a minority of patients in Mexico achieves adequate blood glucose control [2]. Further, the diabetes burden in Mexico coexist with high levels of infectious diseases, exposing the health system to a ’double-disease burden’ that increases the pressure to identify treatment priorities and to efficiently use existing resources [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 for individuals and households affected by diabetes have been found [12–18]. 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]. So far, most studies have relied on an instrumental variable (IV) strategy that used the genetic component of diabetes based on its family history as an instrument, 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 strategy remains therefore debatable. Panel data methods to account for time-invariant unobserved individual characteristics, have not yet been used. Timeinvariant unobservables such as health endowments or risk preferences could adversely affect health in general and 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 could be accounted for by the use of individual fixed effects (FE), which does not rely on the strong assumptions of an IV.

Apart from the methodological challenges, the heterogeneity in the economic consequences across the population with diabetes needs to be investigated. Recent evidence from Mexico points to a strong positive relationship of diabetes duration with mortality due to diabetes related complications [26]. It was also found that a longer disease duration was related with higher glycated hemoglobin (HbA1c) levels and that those with 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 specifically models individual level fixed effects and between effects separately, we took account of time-invariant heterogeneity when assessing the impact of self-reported diabetes and of time since diagnosis on labour outcomes. We also made use of rich and novel biomarker data from the most recent wave of data, to explore the role of undiagnosed diabetes.

# DATA

This paper used 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 this may bias the estimated impact of diabetes on female employment status. We therefore 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 is higher. This reduced direct comparability to the panel results so that biomarker based results were compared to the analysis of the self-reported data for this specific subsample.

Our outcome variables were employment status, hourly wage, weekly working hours 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 renumeration, 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.

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  *Diabetes variables* | 0.16 | 0.26 | 0.00 | 0.09 | 0.11 | 0.04 |
| Diabetes duration (years)  *Control variables* |  | 7.40 |  |  | 7.79 |  |
| 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  Education | 0.19 | 0.15 | 0.00 | 0.19 | 0.19 | 0.86 |
| 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.

The first part of the analysis focused on the relationship of labour outcomes with selfreported 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 reestimated 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 [30], who assumed that everybody diagnosed before 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 it was unlikely that both populations consisted only 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 that actually had early onset diabetes but only joined the sample after already having been diagnosed for several years). For the pooled data of 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 in Mexico of 7.5% in 2006 in a sample also including people over the age of 64 [2]. Apart from self-reported diabetes that 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 allowed us to construct a measure of time since diagnosis. Since the timing of diagnosis was only reported in the third wave, only those that were present in the third wave could be included in this analysis (but this still allowed a panel data analysis).

The second part of the analysis assessed the role of measurement error associated with self-reported diabetes, by also considering those with undiagnosed diabetes, i.e. the false negatives. These were identified by their biometrically measured blood glucose value, which was available for over 6000 respondents in the third wave and allowed us to test for differences between the effects of self-reported and undiagnosed diabetes on labour outcomes.

# 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 the output of a FE model, accounting for the potential bias introduced by time-invariant unobservables, however it only provides an estimate of the effect for incident diabetes cases (705 incident cases compared to 970 non-changing diabetes cases in the used sample). Modeling the between effect also allowed us to use information from those that already had diabetes at baseline.

*Yit* = *β*0+ *β*1(*Dit* − *Di*)+ *β*2(*Xit* − *Xi*)+ *β*3*Di* + *β*4*Xi* +(*ui* + *eit*)*,* (1)

To arrive at the within-effect, we demeaned the diabetes indicator *Dit* by its cluster mean

*Di*), so that *β*1 represented the within-person variation of diabetes over time. The same was done for the other time-varying covariates *Xit*. Further, the cluster means of diabetes and of all other time-varying covariates were included in the model to capture the between effect. The error terms *ui* and *eit* captured the errors for the within and between variation, respectively. *Yit* was a binary variable taking a value of 1 if respondent *i* reported being in employment at time *t* and 0 otherwise. We used the user written Stata command xthybrid to estimate the WB model [31]. The command estimates the WB model using 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 were almost identical 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 relationship of diabetes with wages and working hours, our empirical models were estimated conditional on being in employment. *Yit* 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.

## Labour outcomes and time since diagnosis

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

*Yit* = *β*0+ *β*1(*Dyearsit* − *Dyearsi*)+ *β*2(*Xit* − *Xi*)+ *β*3*Dyearsi* + *β*4*Xi* +(*ui* + *eit*)*,* (2)

where *β*1*Dyearsit* 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. Models excluding the calendar year dummies provided similar results.

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

*Yit* = *β*0+*β*1(*Dsplinesit*−*Dsplinesi*)+*β*2(*Xit*−*Xi*)+*β*3*Dsplinesi*+*β*4*Xi*+(*ui*+*eit*)*,* (3)

Based on visual inspection (Fig 1 on page 35) 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:

*Yi* = *β*0 + *β*1*Dsri* + *β*2*Xi* + *ci* + *vi* (4)

where *vi* 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, as most households had only one member providing biomarker information.

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

*,* (5)

where *Dbiod* was equal to 1 if HbA1c ≥ 6*.*5%.

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

*Yi* = *β*0+ *β*1*Dsri* + *β*2*Dbioi* + *β*3*Dsri* ∗ *Dbioi* + *β*4*Xi* + *vi* + *ui.* (6)

The interaction term changed 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; *β*2 measured 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 showed the effect for those with self-reported diabetes and HbA1c levels above the threshold. We then tested if *β*1+*β*3 = *β*2, i.e. if the effect of self-reported diabetes was significantly different from that of undiagnosed diabetes.

We further investigated the effect of the severity of diabetes on labour outcomes, replacing *Dbiod* with *Dbioc*, a variable that was 0 for HbA1c < 6.5% and took the actual value of HbA1c for those with an *HbA*1*c* ≥ 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).

*Yi* = *β*0+ *β*1*Dsri* + *β*2*Dbioci* + *β*3*Dsri* ∗ *HbA*1*ci* + *β*4*Xi* + *vi* + *ui.* (7)

Stata 15 was used for all analyses [36].

# 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 similar between effect suggested that the effect is generalizable to the entire self-reporting diabetes population, i.e. not only for incidence cases. Additionally, it provided some evidence that time-invariant unmeasured confounders may play a less prominent 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 and wages and working hours, though the between effect suggested that men with diabetes in general earned more than their counterparts without diabetes (Columns 3–6 of Table 2). Splitting diabetes 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 wages, we found a positive effect of diabetes incidence on women. However, it may be that the found effects using the within estimator for early onset cases were spurious due to the relatively low incidence rates of diabetes before the age of 40. The betwee effects show that especially older men with diabetes received higher wages than those without diabetes. For working hours, we only found effects for the within-effects estimator for early onset, which again may have been spurious due to small rates of diabetes incidence.

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 Table 2. Labour outcomes and self-reported diabetes

Employment

Log

hourly

wages

Weekly

work

hours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Males | Females | Males | Females | Males | Females |
| Diabetes (within) | −0.054∗∗ | −0.060∗∗ | 0.063 | 0.074 | −0.582 | −1.990 |
|  | (0.025) | (0.024) | (0.067) | (0.160) | (1.501) | (2.513) |
| Diabetes (between) | −0.077∗∗∗ | −0.066∗∗∗ | 0.097∗∗ | −0.078 | −0.804 | −1.032 |
|  | (0.016) | (0.015) | (0.046) | (0.064) | (0.848) | (1.306) |
| Within=Between (p-value) | 0.447 | 0.832 | 0.680 | 0.392 | 0.897 | 0.735 |
| 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.

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∗ |
| 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 non-parametric kernel-weighted local polynomial regression. For women, a first dropoff occurred right after diagnosis; though no consistent patter emerged thereafter. The dynamics for wages and working hours 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. Apart from that, the results indicated that the effects found in Table 4 are driven mainly by those with a diabetes onset after age 35.

The non-linear results presented in panels B and C for the spline and dummy estimates, respectively, 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. Reestimating the specifications with a random effects model that combined both types of variation into one estimate showed that, at least for the dummy estimates, coefficients suggested a more or less immediate reduction in employment probabilities that became stronger the longer a person had diabetes ( see Supplementary Table 10). We did not estimate models splitting diabetes in early and late onset groups due to the likely strong reductions in statistical power.

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

Employment

Log

hourly

wages

Weekly

work

hours

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

## Cross-sectional biomarker analysis

Supplementary Table 11 shows that 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, indicating a selection based on the severity of diabetes and potentially diabetes duration into the diagnosed diabetes group. Hence, if the adverse effects of diabetes were due to its health impact we would suspect worse labour market outcomes for the diagnosed compared to the undiagnosed population.

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

Males

Females

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Diagnosed | Undiagnosed | P value | Diagnosed | Undiagnosed | P value |
|  | diabetes | diabetes | (t-test) | diabetes | diabetes | ( 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  Subjective health | −0.174 | 0.117 | 0.000 | 0.012 | 0.103 | 0.157 |
| 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)  Blood pressure | 0.262 | 0.074 | 0.000 | 0.397 | 0.150 | 0.000 |
| 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 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 for male wages and female working hours.

Table 6. Biomarker results.

Employment

Log

hourly

wages

Weekly

working

hours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
|  | Males | Females | Males | Females | Males | Females |
| Panel A: Diabetes ( self-reported ) |  |  |  |  |  |  |
| Self-reported diabetes | −.051∗∗ | −.044∗ | −.010 | −.040 | −.293 | −.751 |
| Panel B: Diabetes ( biomarker ) | (.026) | (.023) | (.065) | (.113) | (1.305) | (2.178) |
| Biomarker diabetes (HbA1c ≥ 6*.*5) | −.012 | −.031∗ | −.007 | −.057 | −.088 | 1.153 |
| Panel C: Interacting self-reported and biomarker diabetes | (.016) | (.018) | (.044) | (.070) | (.844) | (1.462) |
| Self-reported diabetes but tested negative (*β*1) | −.030 | −.001 | 0.328 | −.002 | 1.756 | 6.183 |
|  | (.056) | (.050) | ( ) | (.226) | (3.248) | (4.356) |
| Biomarker diabetes but not self-reported (HbA1c ≥ 6*.*5) (*β*2) | 0.006 | −.017 | 0.017 | −.054 | 0.168 | 2.577 |
|  | (.018) | (.020) | ( ) | (.078) | (.960) | (1.640) |
| Self-reported diabetes and biomarker diabetes (*β*3) | −.029 | −.042 |  | −.010 | −2.511 | −10.883∗∗ |
|  | (.062) | (.058) | ( ) | (.259) | (3.594) | (5.153) |
| All self-reported (*β*1+ *β*3) | −.059∗∗ | −.043 |  | −.012 | −.755 | −4.700∗ |
|  | (.029) | (.031) | ( ) | (.136) | (1.570) | (2.777) |
| F-test (p-value): *β*1+ *β*3 = *β*2  Panel D: HbA1c levels | 0.111 | 0.564 | 0.462 | 0.818 | 0.674 | 0.056 |
| Self-reported diabetes | −.050 | −.013 | 0.223∗ | 0.029 | 1.650 | 3.464 |
|  | (.065) | (.041) | (.117) | (.178) | (2.504) | (3.527) |
| HbA1c if ≥ 6*.*5 | 0.001 | −.003 | 0.002 | −.005 | −.005 | 0.256 |
|  | (.002) | (.002) | (.005) | (.008) | (.104) | (.192) |
| Self-reported diabetes × HbA1c if ≥ 6*.*5 | −.001 | −.002 | −.029∗∗ | −.005 | −.231 | −.746∗ |
|  | (.007) | (.005) | (.014) | (.022) | (.283) | (.408) |
| N | 2785 | 3623 | 1803 | 884 | 2302 | 1144 |

*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 as well as highincome countries, with the potential to severely impact 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 confirmed the alarming levels of clinically tested diabetes (27% prevalence) and indicated that a large proportion of the population (18%) was 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 regarding the employment impact of diabetes that used cross-sectional information. Nonetheless, these new results also suggested a comparatively larger impact of diabetes on female employment probabilities. Also, the evidence points towards the main effects being driven by those with a diabetes onset at a relatively later state, probably largely consisting of people with type 2 diabetes Analyses of the long term impact indicated that the employment probability fell gradually over the years after having been diagnosed. Using non-linear models, the findings were less clear, potentially due to reductions in statistical power, overall indicating that adverse effects became stronger with time since diagnosis. This contrasts with estimates for the USA, where no linear trend was found, but non-linear effects with employment probabilities falling after 11 to 15 years for females and after 2-5 years for males [16]. For wages and working hours, our results were more ambiguous and only specifications investigating the effect of the time since diagnosis indicated an adverse relationship for female wages.

The additional information provided by the within-between model might be used to explain why we did not find effects of diabetes on wages and working hours. It appears that in particular the between effects suffered from selection bias especially when estimating wages, which were positively associated with a diabetes diagnosis in men. Reasons for that may be that higher paid men were more likely to develop diabetes, or that higher paid men were able to remain employed without experiencing wage reductions due to their particular set of skills or value that they present to their employer. They may also have had access to better health care leading to better diabetes related health outcomes. Lower paid workers, on the other hand, were probably more likely to be in informal employment or less skill intensive occupations, where it was easier for them to be replaced with healthier employees. They may also lacked access to high quality diabetes care, making it more likely for them to develop debilitating complications earlier [37] that then affected their ability to work (see also Seuring et al. (2015) [20], who found that reductions in employment probabilities particularly affected those in informal employment). This would explain why most of the adverse effects of diabetes were found at what economists call the extensive margin, mainly affecting the probability to be employed rather than wages or working hours.

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 likely suffer from measurement error. Our results using the cross-sectional biomarker data suggest that indeed those with undiagnosed diabetes were significantly more healthy and younger, among other factors, than those aware of their diabetes. Consequently, we found that diabetes based on biomarkers was less related to reduced employment compared to self-reported diabetes, in particular for men. Further analysis showed that this was due to the non-relationship between undiagnosed diabetes and employment. For the USA, a study found—similar to us—no statistically significant effects of undiagnosed diabetes on employment, while the effect of diagnosed diabetes was significant [17]. Our results further suggest 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 not be surprising given that HbA1c levels are only informative for the last three months, and are not the only indicator for the severity of diabetes. More likely overall health differences related to a longer diabetes duration and selection into the diagnosed population based on emerging diabetes related health problems, were driving the adverse economic effects.

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 still have been affected by unobserved time-variant confounders or reverse causality, where employment status affected the propensity to develop or to be diagnosed with diabetes. However, studies that particularly looked at this did not find strong evidence for the existence of an effect of employment status on diabetes [38,39], though they were carried out in high-income countries. Further, because diabetes affects the development of potentially important confounders that were not predetermined, such as BMI, blood pressure, heart disease and consequently health status, in order to capture the full effect of diabetes on our outcomes we did not include these An additional limitation of the duration analysis 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. Additionally, the panel data analysis suffered from the use of self-reported diabetes information which, as shown in the biomarker section of this study, limited its generalizable to those self-reporting their diabetes which were likely to be considerably different to the undiagnosed diabetes population. A limitation of the biomarker section was its cross-sectional nature which limited the possibilities to account for any unobserved confounding. However, at least for the models estimating the effects on employment it appeared that confounding due to time-invariant unobservables could be a minor concern, given the small differences in the within- and between-estimates found in the panel data analysis.

Our findings bear several implications. First, the impact of self-reported diabetes on labor outcomes in Mexico was 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 wages or working hours for women make a dropout less costly. Other evidence suggests that women with diabetes are in worse metabolic health compared to men when they cross the diabetes threshold [40] , making it more likely for them to drop out due to a deterioration in health. Third, it cannot be assumed that estimates based on self-reported diabetes extend to those with undiagnosed diabetes. Studies should therefore, whenever possible, account for both groups separately, acknowledging their inherent differences, to gain information about the distribution of the economic burden in the diabetes population. If this is not possible, conclusions about the effect of self-reported diabetes should be limited to this specific population, in particular in environments where the share of undiagnosed diabetes is high such as in most low- and middle income countries.

Our results add further weight to the case for reducing the incidence and progression of diabetes. The large proportion of previously undiagnosed cases indicates that diagnosis—at least in Mexico—happens too late or not at all. This reduces the possibilities to prevent complications via treatment and self-management, increasing the risk of severe complications to appear earlier. Earlier diagnosis and ensuing effective treatment may lighten the health and economic burden. Further, there is a need to explore why women experience such strong economic effects. Ultimately, prevention of diabetes is of high importance. Taxation of sugar sweetened beverages may be one promising way forward [41], though their longterm effectiveness remains unclear. Therefore, addressing the double-disease burden of non-communicable and communicable diseases via investments in maternal and child health may be particularly promising, in particular given the established links between early life heath and later life incidence of diabetes and other chronic diseases [22,23,42].

## 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 selfreported 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 inconsistent 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≥ 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

# Studies on diabetes and labour market outcomes

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.

1. 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.
2. 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.
3. 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.
4. 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.

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.

1. 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.

1. Latif E. The impact of diabetes on employment in Canada. Health Economics.

2009;18(5):577–589. doi:10.1002/hec.1390.

1. 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.
2. 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.
3. 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.
4. 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.

1. 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.

1. 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-081412103704.
2. 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.

1. 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.
2. 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.
3. Rubalcava L, Teruel G. User’s Guide for the Mexican Family Life Survey Third Round; 2013.
4. 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.](http://gero.usc.edu/CBPH/files/4%7b_%7d30%7b_%7d2014%7b_%7dPAA/Thomas%7b_%7dHealth%7b_%7dof%7b_%7dthe%7b_%7dMexican%7b_%7dpopulation%7b_%7din%7b_%7dMxFLS.pdf)

1. 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.
2. 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.
3. 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.
4. Slade AN. Health Investment Decisions in Response to Diabetes Information in Older Americans. Journal of Health Economics. 2012;31(3):502–520.
5. 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.

1. Angrist JD, Pischke JS. Mostly Harmless Econometrics: An Empiricist’s Companion.

Princeton University Press; 2009.

1. StataCorp. Stata Statistical Software: Release 15; 2017.
2. 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.
3. Bergemann A, Grönqvist E, Gudbjörnsdottir S. The effects of job displacement on the onset and progression. Netspar Discussion Paper. 2011;(25).
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Yuan X, Liu T, Wu L, Zou ZY, Li C. Validity of self-reported diabetes among middleaged and older Chinese adults: the China Health and Retirement Longitudinal Study. British Medical Journal Open. 2015;5(4):e006633–e006633. doi:10.1136/bmjopen-2014006633.
10. 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

Log

hourly

wages

Weekly

work

hours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Males | Females | Males | Females | Males | Females |
| Early onset (within) | 0.133 | −0.206∗∗ | −0.523 | 0.388∗∗∗ | 14.712∗ | −18.636∗ |
|  | (0.176) | (0.086) | (0.340) | (0.057) | (8.370) | (9.668) |
| Late onset (within) | −0.059∗∗ | −0.048∗ | 0.079 | 0.067 | −1.008 | −1.322 |
|  | (0.025) | (0.025) | (0.067) | (0.163) | (1.513) | (2.553) |
| Early onset (between) | 0.016 | −0.052 | −0.102 | 0.286 | 1.483 | −2.640 |
|  | (0.037) | (0.047) | (0.099) | (0.232) | (2.581) | (4.034) |
| Late onset (between) | −0.087∗∗∗ | −0.067∗∗∗ | 0.121∗∗ | −0.111∗ | −1.077 | −0.869 |
|  | (0.017) | (0.016) | (0.050) | (0.066) | (0.895) | (1.375) |
| Within=Between (p-value) | 0.521 | 0.130 | 0.248 | 0.671 | 0.132 | 0.113 |
| 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) |
| Early onset: Within=Between (p-value) | 0.934 | 0.538 | 0.302 | 0.247 | 0.139 | 0.573 |
| Late onset: Within=Between (p-value) | 0.419 | 0.011 | 0.001 | 0.316 | 0.006 | 0.460 |
| 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

Log

hourly

wages

Monthly

work

hours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Males | Females | Males | Females | Males | Females |
| Early onset (within) | −0.009 | 0.002 |  |  |  |  |
| Early onset: Within=Between (p-value) | 0.218 | 0.195 |  |  |  |  |
| Late onset: Within=Between (p-value) 0.575 0.556 0.208 0.036 0.628 0.379 | | | | | | |

*Notes* The within estimator for the early onset diabetes for 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

Log

hourly

wages

Weekly

work

hours

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Males | Females | Males | Females | Males | Females |
| *Panel A: linear effect* | |  |  |  |  |  |  |
| Years since diagnosis | | −0.010∗∗∗ | −0.007∗∗∗ | 0.009∗ | −0.012 | −0.000 | −0.139 |
| *Panel B: splines*  Years since SR diagnosis | | (0.002) | (0.001) | (0.005) | (0.009) | (0.102) | (0.126) |
| 0–3 |  | −0.010 | −0.018∗∗∗ | 0.039∗∗ | 0.029 | −0.024 | 0.499 |
|  |  | (0.006) | (0.006) | (0.017) | (0.030) | (0.348) | (0.615) |
| 4–7 |  | −0.002 | 0.001 | −0.036 | −0.058 | 0.147 | −0.593 |
|  |  | (0.011) | (0.011) | (0.032) | (0.053) | (0.667) | (1.061) |
| 8–12 |  | −0.010 | −0.003 | −0.003 | −0.074 | −0.142 | −0.088 |
|  |  | (0.015) | (0.011) | (0.041) | (0.051) | (0.860) | (1.097) |
| 13+ |  | −0.017∗ | −0.005 | 0.026 | 0.011 | −0.031 | −0.342 |
|  |  | (0.009) | (0.003) | (0.017) | (0.018) | (0.414) | (0.241) |
| *Panel* | *C: dummies* |  |  |  |  |  |  |
| 0–3 |  | −0.038∗ | −0.046∗∗ | 0.134∗∗ | 0.006 | −0.810 | 1.062 |
|  |  | (0.022) | (0.021) | (0.054) | (0.084) | (1.155) | (1.827) |
| 4–7 |  | −0.030 | −0.073∗∗∗ | 0.087 | −0.043 | 0.684 | 1.119 |
|  |  | (0.023) | (0.023) | (0.059) | (0.120) | (1.390) | (2.531) |
| 8–12 |  | −0.084∗∗ | −0.079∗∗∗ | −0.083 | −0.217∗ | 0.218 | 0.123 |
|  |  | (0.037) | (0.030) | (0.101) | (0.120) | (2.293) | (2.981) |
| 13+ |  | −0.175∗∗∗ | −0.093∗∗∗ | 0.206 | −0.335∗∗ | −1.506 | −2.291 |
|  |  | (0.046) | (0.033) | (0.126) | (0.150) | (2.279) | (3.081) |
| 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.

Table 11. Number of observations with diabetes (HbA1c ≥ 6*.*5%) and self-reported diabetes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *HbA*1*c <* 6*.*5% | HbA1c ≥ 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 |

or cross- betes endogeneity

sections

China 2009, 2011 Employed Panel HbA1c Find a significant re NA Use difference-in [19]

population duction of 16.3 % in difference model,

income for those with exploiting a recent dia recent diagnosis in agnosis of diabetes as a

China. result of biomarker col-

lection 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

Mexico 2005 Working Cross- Self reported A significant (p<0.01) Diabetes exogenous for Probit and bivariate [20]

age popula- section reduction in employ men and women based probit model using tion ment probabilities for on Hausman test (p parental diabetes as IV

males by about 10 % >.10)

points and for females by about 4.5 % points

(p<0.1)

or cross- betes endogeneity

sections

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| USA | 1996-1997 | Elderly population of Mexican Americans close the  Mexican border | Crosssection | 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 | MexicanAmerican working age adults | Crosssection | 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 | Crosssection | 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] |

or cross- betes endogeneity

sections

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| USA | 1979 - 2010 | Follows young adults  in 1979  throughout their adult  life | Panel | Self-reported year of diagnosis | Average reduction of employment probabil-  ity of 28 % points for men and 36 %  points for women; employment probabilities decline shortly after di-  agnosis 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] |

or cross- betes endogeneity

sections

USA 2001 - 2008 Men and Panel Self-reported and women 18 - HbA1c levels for

65 subsample

No statistically sig- Exogeneity assumed Probit model for binary [17]

|  |  |
| --- | --- |
| nificant relationship between undiagnosed diabetes and the prob- | outcomes, OLS for continuous outcomes; all applied to pooled data |

ability 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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Year | Population | | Panel  or cross-  sections | Measurement betes | of | dia- | Main finding | Finding on bias due to endogeneity | | | Estimation method | Reference |
| Canada | 1998 | Men women  64 | and  15 - | Crosssection | 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 for men, resulting upwards biased mates; exogenous women | endogenous | in estifor | Instrumental variable strategy using bivariate probit model and family history of diabetes as the  instrument | [18] |
| Australia | 1999 - 2000 | Men and  women age  >24 | | Crosssection | 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

Log

hourly

wages

Weekly

work

hours

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 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.