

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

Till Seuring<sup>1,2</sup>, Pieter Serneels<sup>2</sup>, and Marc Suhrcke<sup>3</sup>

<sup>1</sup>*Leibniz Institute for Prevention Research and Epidemiology - BIPS*

<sup>2</sup>*University of East Anglia*

<sup>3</sup>*University of York*

May 15, 2017

## Abstract

Diabetes has become one of the most common chronic diseases in middle- and high-income countries. Yet, our understanding of its economic consequences remains limited. Applying fixed-effects estimation to rich panel data for Mexico, where diabetes is the number one cause of death, we find evidence for adverse effects of self-reported diabetes on the probability of being employed, but not on wages or hours worked. Self-reported diabetes reduces the probability of employment with 5 percentage points for men and close to 6 percentage points for women. Examining the effects on wages and working hours by different types of work, the association remains weak, but the results also suggest occupational selection with women who

---

\*Till Seuring is a post-doctoral researcher at the Leibniz Institute for Prevention Research and Epidemiology - BIPS and his email is seuring@leibniz-bips.de, Pieter Serneels is Associate Professor at University of East Anglia and can be contacted at p.serneels@uea.ac.uk, Marc Suhrcke is Professor at the University of York and his email is marc.suhrcke@york.ac.uk. Till Seuring is corresponding author.

self-report diabetes less likely to work in agriculture. Our dynamic analysis finds that the employment probability falls gradually as time since diagnosis progresses. We then contrast self-reported with biomarker data on glycated hemoglobin (HbA1c), available for a large subsample, and observe that 18% of all observations are false negatives, i.e. do not report diabetes but exhibit HbA1c levels above the commonly used clinical threshold. The estimated employment impact of tested versus self-reported diabetes is similar for women, but higher for men. We find no effect for undiagnosed women or men (who tested positive but did not self-report), which seems to be related to their better general health rather than less severe diabetes. The results highlight both the importance of the economic impact of diabetes, and the need to take into account (especially female) undiagnosed patients.

## 1 Introduction

Diabetes, and particularly its most common variant, type 2 diabetes, has increased worldwide and is expected to continue to rise over the next decades (NCD Risk Factor Collaboration, 2016). The condition has become a problem for middle-income countries and high-income countries alike, with over two-thirds of people with diabetes living in the developing world (International Diabetes Federation, 2015). Mexicans and Mexican-Americans appear to be particularly affected by diabetes, also in comparison to other Latino populations (Schneiderman, Llabre, et al., 2014). In Mexico, diabetes prevalence has grown from 6.7% in 1994 to 14.4% in 2006 (Barquera, Campos-Nonato, et al., 2013) and is currently estimated to be around 15.8% in 2015 (International Diabetes Federation, 2015). Already now, diabetes is the number one cause of death in Mexico.

The observed trend has been attributed to a deterioration in diet and a reduction in physical activity (Barquera, Hernandez-Barrera, et al., 2008; Basu et al., 2013), while genetic predisposition among Mexicans with pre-Hispanic ancestry may also play a role (Williams et al., 2013). Recent evidence indicates that the onset of diabetes has been

occurring at an ever earlier age in Mexico (Bello-Chavolla et al., 2017). With treatment as ineffective as it currently is—only a minority achieves adequate blood glucose control (Barquera, Campos-Nonato, et al., 2013)—the earlier onset will increase the likelihood of complications during the productive lifespan.

Diabetes describes various conditions characterized by high blood glucose values, with the predominant disease being type 2 diabetes accounting for about 90% of all diabetes cases (Sicree et al., 2011). The elevated blood glucose levels that are a result of the body’s inability to use insulin properly to maintain blood glucose at normal levels, can entail a range of adverse health effects for the individual concerned. However, via effective self-management of the disease much if not all of the complications can be avoided (Gregg et al., 2012; Lim et al., 2011). In the absence of this management—or in the case of inadequate treatment—over time the constantly elevated blood glucose levels can lead to heart disease and stroke, blindness, kidney disease and nerve problems, foot ulcers and amputations (Reynoso-Noverón et al., 2011). Consequently, diabetes can reduce an individual’s economic activity, including its productivity and labor market participation.

The effect of diabetes on labor outcomes has been studied predominantly in high-income countries, where diabetes was associated with reductions in employment probabilities as well as wages and labor supply (Brown, Pagán, et al., 2005; Brown, Perez, et al., 2011; Brown, 2014; Latif, 2009; Minor, 2011, 2013; Minor and MacEwan, 2016; Seuring, Archangelidi, et al., 2015).<sup>1</sup> While these studies have provided useful evidence, many of the complexities of the relationship between diabetes and labor outcomes remain unaddressed. Especially time-invariant unobserved individual characteristics (e.g. health endowments or risk preferences) may adversely affect health in general and the propensity to develop type 2 diabetes in particular (Ewijk, 2011; Li et al., 2010; Sotomayor, 2013); they may also affect employment probabilities, wages or working hours—either directly through their effects on contemporaneous productivity (Currie and Vogl, 2013), or indi-

---

<sup>1</sup>We know of only two studies for middle-income countries, one for Mexico (Seuring, Goryakin, et al., 2015) and one for China (Liu and Zhu, 2014), which are discussed later in greater detail.

rectly by limiting educational attainment and human capital accumulation (Ayyagari et al., 2011). Further, given the chronic nature of the condition, a better understanding of the time and severity of the potential labor market penalties is important.

Especially in a middle-income country setting, large parts of the population remain undiagnosed (Beagley et al., 2014) and studies relying on self-reported diabetes data may leave undetected important differences between those with diagnosed and undiagnosed diabetes. It is at least conceivable, for instance, that a diabetes diagnosis in itself may exert an effect independently of that resulting from the actual disease. People aware of their condition could be less inclined to continue working if this interferes with their disease management; or they may be suffering from psychological stress, depression, or anxiety, caused by the sudden awareness of being sick; they may also use the diagnosis as a justification for decreasing their labor supply (Kapteyn et al., 2009). Hence labor market effects might well be distinct for people with self-reported diabetes versus those unaware of their condition.

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes, critically improving upon previous work by paying close attention to the challenges of unobserved heterogeneity, to the chronic nature of diabetes and to those undiagnosed. To this end we use three waves of panel data from the Mexican Family Life Survey (MxFLS), covering the period 2002–2012. Applying individual level fixed effects (FE) for the first time in this literature, we take account of time-invariant heterogeneity when assessing the impact of self-reported diabetes and its duration on labor outcomes.<sup>2</sup> We also make use of rich and novel biomarker data from the most recent wave of the MxFLS to also explore the role of undiagnosed diabetes—an issue that has remained unexplored in the existing literature despite its considerable importance.

Our results of the panel data analysis for self-reported diabetes suggest an economically important decrease in the employment probability of those who are aware of their disease.

---

<sup>2</sup>A recent review of the economic cost of diabetes confirms the scarcity of evidence for low- and middle-income countries (Seuring, Archangelidi, et al., 2015).

Wages and working hours, however, remain unaffected. Further analysis for the long term indicates that employment probabilities are reduced in particular immediately after diagnosis and then further after the first 10 years.

The biomarker analysis reveals that clinically diagnosed diabetes entails a significant employment penalty for women but not for men. Jointly assessing the effects of both clinical and self-reported diabetes provides an insight into the labor market impact for those with undiagnosed diabetes—people who are tested positive but did not self-report: in contrast to those self-reporting diabetes, men and women unaware of their condition do not experience adverse labor market effects.

## 2 Diabetes and labor outcomes—what do we know?

A limited number of studies provides insights on the relationship between diabetes and labor outcomes. Table 9 in appendix summarises the main findings of these studies, the characteristics of the sample, the estimation method they use and the approach to measure diabetes. To the best of our knowledge only two studies exist for low- and middle-income countries (LMICs). Liu and Zhu (2014) exploit a natural experiment in China and find a significant reduction in income for those with a recent diagnosis of diabetes. A study for Mexico using cross-sectional data from 2005, finds a significant ( $p < 0.01$ ) reduction in employment probabilities for males by about 10 percentage points and for females by about 4.5 percentage points ( $p < 0.1$ ), using parental diabetes as an instrumental variable (IV) (Seuring, Goryakin, et al., 2015).

More studies have investigated the effects of diabetes on labor outcomes in high-income countries. Brown, Pagán, et al. (2005) consider an elderly population of Mexican-Americans living close to the Mexican border in the US, and find 7 percentage points lower employment rates for men with self-reported diabetes, while for women, the negative relationship becomes insignificant when using IV estimation. In a similar vein, Brown, Perez,

et al. (2011), again considering a cross-section of Mexican-Americans, detect a negative relationship between the level of HbA1c on one hand and the probability of employment or male wages on the other hand. Women remain again unaffected.

Slightly different results are obtained in two other studies, this time for a more representative US population: using a sample comprised exclusively of women, Minor (2011) finds a significant negative effect of self-reported diabetes on female employment and earnings but not on working hours. In this study self-reported diabetes turns out as endogenous and the simple probit estimates are downward biased compared to IV estimates. In a subsequent study by the same author, employment probabilities decline shortly after diagnosis for men and after about 10 years for women, while wages are not affected by the time since diagnosis.

Results for Canada indicate a significant negative impact of self-reported diabetes on the employment probability of women, but not of men, using an IV strategy similar to Brown, Pagán, et al. (2005). The IV results suggest diabetes to be endogenous for men, resulting in upward biased probit estimates (Latif, 2009). For Australia, (Zhang et al., 2009) show reduced labor market participation for men and women as a result of diabetes, with the effects again appearing overstated in the 'naive' regression models.

As far as we know, only one labor market related study has considered undiagnosed diabetes, if for a small sample: Minor and MacEwan (2016) find no statistically significant relationship between undiagnosed diabetes employment probabilities, but negative and statistically significant results when using self-reported diabetes, for both men and women. When combining both the undiagnosed and self-reporting, the effect on employment remains statistically significant but decreases in magnitude.

While these studies suggest substantial economic losses for individuals and households being affected by diabetes, most of the existing evidence suffers from methodological limitations, in particular due to the use of cross-sectional data and the limited possibilities to account for unobserved characteristics. The papers attempting to address this bias rely on

the family history of diabetes as the identifying instrument to exploit the genetic component of the disease. However, it remains debatable whether the instrument indeed satisfies the exclusion restriction, as it may also proxy for other genetically transferred traits, including unobserved abilities, as well as intrahousehold or intergenerational dynamics that impact labor outcomes directly.<sup>3</sup>

Furthermore, most—but not all—studies use self-reported diabetes as a proxy for diabetes. Self-reported health data likely suffer from several short-comings, thus introducing non-classical measurement error due to systematic misreporting. In the context of labor market impact studies, such measurement error has been shown to potentially cause biased and overstated impact estimates (Cawley et al., 2015; O’Neill and Sweetman, 2013; Perks, 2015). With regards to diabetes, the concern is especially linked to possible false negatives. False positives might of far less concern since one would expect there to be limited incentive to report diabetes when one does not have it—although we cannot exclude this. A recent study from China confirms that those who self-report diabetes are highly likely to actually have diabetes ( $>98\%$ ), while only a minority of those who have diabetes ( $40\%$ ) according to clinical tests, actually self-report the disease (Yuan et al., 2015). This pattern is confirmed in our data, where the biomarker results support the majority of positive diabetes self-reports. Even of those reporting a diabetes diagnosis while the biomarker data suggest non-diabetic HbA1c levels, many likely have diabetes but treatment has pushed their HbA1c levels back below the threshold, leading to very few false positives. However, a much larger proportion reports false negatives ( $18\%$ ), suggesting a large undiagnosed population with diabetes. This population may have a distinct profile that prevented them from getting diagnosed so far, for instance they may not be able to afford health care, live further away from a health facility, or their diabetes has remained mostly asymptomatic so far, all potentially influencing the effect of diabetes on

---

<sup>3</sup>It is conceivable that diabetes might deteriorate parental health in such a way that the offspring either has to give up its employment to provide care, or has to increase labor supply to compensate for lost income, as also argued by Seuring, Goryakin, et al. (2015).

their labor outcomes.

This paper makes headway to overcome these key limitations in two ways. First, we apply fixed effects (FE) estimations to three waves of panel data, allowing to control for unobserved time invariant characteristics. Second, we use biomarker data for a large subsample of the population to carry out a comparison between the effect of self-reported and clinically tested diabetes. Given the relatively precise measurement of self-reported diabetes it also allows inference about the effects for undiagnosed patients, who suffer from diabetes according to a clinical test, but are unaware of this. To further our understanding of the relationship between diabetes and labor outcomes, we also consider its effect on the type and sector of employment, and the long term effects in the years after diagnosis.

### 3 Context and Data

Mexico, a middle-income country, has among the highest levels of obesity and diabetes prevalence in the world (International Diabetes Federation, 2015; World Health Organization, 2011). A recent study showed that diabetes accounted for one-third of all deaths among those 35 to 74 years old in Mexico City, with renal disease, cardiac disease, infection, acute diabetic crisis, and other vascular diseases being the biggest contributors to the elevated mortality risk of people with diabetes (Alegre-Díaz et al., 2016). The high 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 the existing resources. (Recent work has suggested that the treatment of non-communicable diseases is not well integrated into the health care system (Gutiérrez-delgado and Guajardo-barrón, 2009)). Earlier studies point to a meager performance of the health system in diagnosing and effectively treating diabetes in Mexico, with most patients achieving only poor glucose control and suffering from other untreated risk factors such as hypertension (Alegre-Díaz et al., 2016; Flores-Hernández



et al., 2015). In addition, about half of the diabetes population has been estimated to be unaware of the condition (Barquera, Campos-Nonato, et al., 2013).

This paper uses the Mexican Family Life Survey (MxFLS), a nationally representative longitudinal household survey containing three waves conducted in 2002, 2005–2006 and 2009–2012. All household members aged 15 and above were interviewed, covering information on a wide range of social, demographic, economic and health characteristics (Rubalcava and Teruel, 2013). Throughout the analysis, the samples we use are restricted to the working age population (15–64). Our first part of the analysis uses all three waves, taking advantage of the large amount of observations and the panel structure of the data. The second part uses a biomarker subsample of the third wave (2009–2012). It is important to note that the age distribution of the biomarker sample is somewhat older than the entire sample, as it includes everybody above the age of 44 but only a random subsample of those aged 44 or below (Crimmins et al., 2015). Hence the self-reported diabetes prevalence is higher for this subsample. The biomarker analysis will therefore be compared to the analysis of the self-reported data for this subsample specifically.

Our outcome variables of interest comprise the labor outcomes employment, hourly wage, weekly working hours and occupation.<sup>4</sup> About half of the respondents in the sample live in rural areas. Descriptive statistics for the entire panel sample show that 86% of men report some form of employment compared to 37% of women (see Table 1). Interestingly,

---

<sup>4</sup>Employment status is defined as having worked or 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 include those employed informally, for instance people working in a family business. We tested if changing the definition of being employed to having worked at least ten hours per week affects the results, and this only leads to marginal changes in the coefficients and standard errors, keeping the interpretation of the results unchanged. The hourly wage was calculated by adding up the reported monthly income from the first and second job (if any) and dividing it by the average number of weeks per month providing us with an estimate of the average earnings per week, which is then divided by the weekly working hours to arrive at the hourly wage. Labor income was reported in two ways: either responding to questions on wages, income from piecework, tips, income from extra hours, meals, housing, transport, medical benefits and other earnings, or by reporting the aggregate labor income for the whole month. We adjusted the calculated wage for inflation from the year of the interview up to 2013 and take the log of real wages. Due to a considerable number of missing or zero income reports, the sample used for the wage estimation is smaller than the sample for working hours. Working hours reflect working hours of the first and second job (if applicable).

men do not report considerably higher hourly wages than women but work more hours per week. Men also work more often in agricultural jobs while women are more likely to be self-employed or in non-agricultural wage employment. Women also have lower educational attainment on average.

In the first part of the analysis we focus on the relationship of labor outcomes with self-reported diabetes<sup>5</sup>. For the pooled data of all three waves (Table 1), diabetes was self-reported by 5% of men and 6% of women, respectively. This is consistent with Barquera, Campos-Nonato, et al. (2013), who observe a prevalence of diagnosed diabetes in Mexico of 7.5% in 2006, using a slightly older sample that also included respondents beyond 64 years of age. Apart from self-reported diabetes information that is available in all rounds, we also use information on the self-reported year of diagnosis as well as biometrically measured HbA1c levels for a subsample of respondents. Throughout, our analysis focuses on the working age population (15–64), and excludes pregnant women and those in school.<sup>6</sup>

Because self-reported diabetes reporting exhibited some inconsistency over time for some of the respondents, we apply corrections using disease information from earlier and subsequent waves to infer on the current, missing or inconsistent, diabetes status. Appendix 1 provides details on the procedure. Information on the self-reported year of diagnosis allows us to construct a measure of the time since diagnosis for all waves. Importantly, this limits the sample of the time since diagnosis analysis to those that were present in the third wave.

A further, and no less important, source of measurement error with self-reported diabetes is the omission of those with undiagnosed diabetes, i.e the false negatives. The information on biometrically measured blood glucose values for a subsample of the 2009–2012 wave, containing data for over 6000 respondents allows identification of respondents

---

<sup>5</sup>Self-reported diabetes is based on the survey question: “Have you ever been diagnosed with diabetes?”

<sup>6</sup>Pregnant women have an increased diabetes risk and this may bias the estimated impact of diabetes on female employment status. We dropped all observations of women reporting to be pregnant at the time of the survey (N=764). We also estimated models including a dummy variable for pregnant women. This only leads to minor changes in the diabetes coefficient for women and does not affect the interpretation of the results.

with undiagnosed diabetes.

The biomarker data indicate that a large share of the population (27%) have an HbA1c indicative of diabetes, defined by the World Health Organization (WHO) as levels equal to or above 6.5% (World Health Organization, 2011). These rates are high when compared to those obtained from similar surveys in the USA and China (Frankenberg et al., 2015). A second striking observation is the large proportion of false negatives, namely 18% of all observations, implying that 68% of males and females who test positive do not self-report and hence are unaware of their condition. This may lead to biased estimates based on self-reported diabetes.

## 4 Estimation strategy

### 4.1 Panel data on self-reported diabetes

To investigate the relationship between self-reported diabetes and three labor outcomes: employment, wages and weekly working hours, respectively, we estimate the following fixed effects (FE) model.<sup>7</sup>

$$Y_{it} = \beta_0 + \beta_1 Diabetes_{it} + \beta_2 X_{it} + c_i + \gamma_t + u_{it}. \quad (1)$$

where  $Y_{it}$  is a binary variable taking a value of 1 if respondent  $i$  reports being in employment at time  $t$  and 0 otherwise,  $Diabetes_{it}$  is a binary variable taking a value of 1 at time  $t$  if the respondent reports having ever received a diagnosis of diabetes<sup>8</sup>,  $X_{it}$  is a vector of control variables,  $c_i$  represents an individual fixed effect,  $\gamma_t$  represents year dummies, while  $u_{it}$  is the error term.

For the relationship of diabetes with wages and working hours, our empirical models

---

<sup>7</sup>We also estimated random effects models but do not present them here as the Hausman test suggested the use of the FE model throughout. Results are obtainable upon request.

<sup>8</sup>The data at hand does not allow us to distinguish between type 1 and type 2 diabetes. Existing studies find no effect of type 1 diabetes on labor outcomes. Our estimates of type 2 diabetes impact on labor outcomes may therefore be attenuated and provide a lower bound (Minor, 2011; Minor and MacEwan, 2016).

Table 1: Descriptive statistics for panel and biomarker sample.

	Panel		Biomarker	
	Males	Females	Males	Females
<i>Dependent variables</i>				
Employed	0.86 (0.34)	0.37 (0.48)	0.86 (0.35)	0.34 (0.47)
Hourly wage (Mexican Peso)	42.47 (485.87)	40.49 (142.08)	36.30 (53.69)	35.23 (43.63)
Weekly working hours	46.82 (16.79)	38.99 (18.90)	46.00 (16.89)	38.15 (19.65)
Agricultural worker	0.22 (0.41)	0.04 (0.20)	0.25 (0.43)	0.03 (0.18)
Self-employed	0.19 (0.39)	0.28 (0.45)	0.21 (0.41)	0.32 (0.47)
Non-agricultural worker or employee	0.59 (0.49)	0.68 (0.47)	0.53 (0.50)	0.64 (0.48)
<i>Diabetes variables</i>				
Self-reported diabetes	0.05 (0.22)	0.06 (0.24)	0.09 (0.29)	0.12 (0.32)
Diabetes duration if self- reported diabetes (years)	7.49 (6.01)	7.83 (7.83)	7.48 (6.07)	7.99 (7.03)
Glycated hemoglobin (HbA1c)			6.46 (1.89)	6.58 (2.02)
HbA1c $\geq 6.5\%$			0.26 (0.44)	0.28 (0.45)
Undiagnosed diabetes			0.18 (0.39)	0.18 (0.39)
<i>Control variables</i>				
Age	36.03 (13.62)	36.29 (13.17)	42.78 (14.28)	42.79 (13.94)
Rural village of $< 2,500$	0.44 (0.50)	0.43 (0.50)	0.50 (0.50)	0.46 (0.50)
Married	0.54 (0.50)	0.54 (0.50)	0.60 (0.49)	0.56 (0.50)
Number of children (age $< 6$ ) in household	1.48 (1.45)	1.57 (1.47)	1.18 (1.29)	1.22 (1.32)
Indigenous group	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)	0.18 (0.39)
Secondary	0.30 (0.46)	0.30 (0.46)	0.26 (0.44)	0.26 (0.44)
High school	0.16 (0.36)	0.13 (0.34)	0.14 (0.34)	0.12 (0.33)
Higher education	0.11 (0.32)	0.09 (0.29)	0.12 (0.32)	0.09 (0.28)
Wealth index	-0.00 (1.02)	-0.02 (1.00)	0.09 (1.07)	-0.00 (1.02)
Observations	21388	27341	2785	3623

Notes Mean values, standard deviations in parenthesis.

are estimated conditional on being in employment.  $Y_{it}$  represents the log hourly wage or the weekly working hours over the last year, for respondent  $i$  at time  $t$ .

The control variables in both FE specifications include dummy variables to capture the effects of living in a small, medium or large city with rural as the reference category, and state dummies. They also include a marital status dummy and the number of children residing in the household below the age of 6, to control for both the impact of marriage and children. To account for the effect of changes in household wealth on diabetes and employment probabilities, we use standard principal component analysis of multiple indicators of household assets and housing conditions to create an indicator for household wealth (Filmer and Pritchett, 2001)<sup>9</sup>. The models also include a quadratic age term and calendar year dummies to capture the non-linear effect of age and a time trend, respectively.

While using individual level FE does not allow to fully identify a causal relationship, it does improve considerably on existing estimates which are typically obtained from cross-section analysis, or from IV estimation that tend to be weakly identified. The FE model does control for unobserved personal characteristics, although omitted time-variant variables and simultaneity may still affect the relationship of interest. With respect to employment status, one potential concern could be that job loss affects lifestyle choices leading to changes in the probability to develop diabetes that could in turn again affect labor outcomes. Existing work for high-income countries finds no evidence for this kind of reverse causality (Bergemann et al., 2011; Schaller and Stevens, 2015). Another possible channel might be that stress at work leads to a higher propensity of developing type 2 diabetes (Eriksson et al., 2013; Heraclides et al., 2012). However, while stress levels may change over time, a person’s coping mechanisms to deal with stress are typically considered to be time-invariant (Schneiderman, Ironson, et al., 2005). But while we cannot exclude a role of time-variant unobserved factors or simultaneity, time-invariant variables (includ-

---

<sup>9</sup>Our composite wealth index consists of owning a vehicle, a second house, a washing machine, dryer, stove, refrigerator or furniture, any electric appliances, any domestic appliances, a bicycle or farm animals. It further accounts for the physical condition of the house, proxied by the floor material of the house, and the type of water access.

ing genetic predisposition and stable personality traits) may be more important. The FE approach should then limit the bias resulting from these time-invariant confounding factors.

## 4.2 Labor outcomes and time since diagnosis

In light of the chronic nature and irreversibility of diabetes, there is likely a merit in examining the long term effects post diagnosis. We estimate the following model:

$$Y_{it} = \beta_0 + \beta_1 Dyears_{it} + \beta_2 X_{it} + c_i + u_{it}, \quad (2)$$

where  $\beta_1 Dyears_{it}$  is continuous indicating years since first reported diabetes diagnosis.<sup>10</sup> To capture possible non-linearities in the relationship we also consider a spline function that allows for the effect to vary over time.

$$Y_{it} = \delta_0 + g(Dyears_{it}) + \delta_2 X_{it} + c_i + u_{it}. \quad (3)$$

with  $g(Dyears_{it}) = \sum_{n=1}^N \delta_n \cdot \max\{Dyears_{it} - \eta_{n-1}\} I_{in}$  and  $I_{in} = 1[\eta_{n-1} \leq Dyears_{it} < \eta_n]$ , with  $\eta_n$  being the place of the  $n$ -th node for  $n = 1, 2, \dots, N$ . The coefficient  $\delta_n$  captures the effect of diabetes for the  $n$ -th interval. The effects are linear if  $\delta_1 = \delta_2 = \dots = \delta_n$ .<sup>11</sup> Based on visual inspection (Figure 1 on page 21) we choose three nodes that seem to capture possible non-linearities in the relationship between diabetes duration and labor outcomes best. These are located at 4, 11 and 20 years after diagnosis. The first four

---

<sup>10</sup>Note that while usually the simultaneous inclusion of year dummies and time since diagnosis, which varies by one unit in each time period, would not allow a separate identification of the coefficient of time since diabetes diagnosis in Eq. (2) and Eq. (3), identification relies on the presence of people without diabetes in the sample, for which diabetes duration does not increase. Models excluding the calendar year dummies provide similar results. As a further robustness check, we also estimate two models that only use between-individuals variation, i.e. a linear probability model (LPM) that uses only data from the third wave, the only wave where year of diagnosis was originally reported, and a pooled LPM that used data from all three waves.

<sup>11</sup>Because the year of diagnosis was only reported in the third wave, time since diagnosis is not available for those who were not interviewed in the third round. A reported diagnosis in the year of the interview is counted as 'one year since diagnosis'.

years capture any immediate effects of the diagnosis, the years five to eleven capture any effects during time of adaptation to the disease. The last term (beyond 11 years) captures the long-term effects.

### 4.3 Labor outcomes and biometrically measured diabetes

Since a sizeable number of individuals remain undiagnosed, and some, may also misreport, estimations based on self-reports may be biased. More specifically, we see three relevant scenarios, which we consider in more detail to assess whether we can sign the bias ex-ante. First, respondents may systematically overreport diabetes, leading to false positives. This may be unintentionally—for instance due to a misdiagnosis, either from a health professional or because of self-diagnosis, or intentionally—for instance with a view to justifying some other adverse event in their life, such as being unemployed. Second, respondents may systematically underreport, leading to false negatives. They may be concerned about negative stigma associated with the condition or, more importantly, diabetes may have remained undiagnosed, leaving them unaware of their condition. Third, a diagnosis is more likely to exist for those who are more probably to visit a doctor, for instance because they are either more affected by the condition, wealthier, or hypochondriac. As a result, self-reports may suffer from a selection bias.

Overreporting may attenuate the estimated impact of diabetes if the false positives are in fact in good health, or it may lead to an overestimation if they have other attributes that negatively affect labor outcomes, including general health, or another illness. Similarly, underreporting may lead to an overestimation if those with undiagnosed diabetes are generally healthier, hence more likely to have positive labor outcomes. However, if the undiagnosed and the diagnosed groups are similar in terms of health, then this would lead to an underestimation of the impact of diabetes.

The health information revealed at a diabetes diagnosis may also have an effect in itself. It might, for instance, affect the patient's psychology which in turn may affect his or her

economic decision making and behaviour. Two studies found evidence that patients with a diabetes diagnosis and subsequent treatment are more prone to psychological conditions, including depression and anxiety (Paddison et al., 2011; Thoolen et al., 2006) compared to people without diabetes.

Since undiagnosed diabetes is not found to be associated with psychological conditions, this suggests a possible causal relationship (Nouwen et al., 2011). Health information may also lead to a change in behaviour. Slade (2012) shows how patients, when learning about their diabetes diagnosis, change their consumption of alcohol and smoking and start to lose weight. This is in line with evidence for other chronic diseases (see Baird et al. (2014), Gong (2015), Thornton (2008), and Zhao et al. (2013)). Recent evidence for China also suggests that receiving a diabetes diagnosis in itself reduces labor income, possibly through psychological effects of the diagnosis (Liu and Zhu, 2014).<sup>12</sup>

The use of biomarker data allows to explore both the extent of under- and overreporting, and the possible bias in the estimate of the relationship of self-reported diabetes and labor outcomes; it also enables us to look at diabetes severity, as measured by HbA1c values. Since these data are only available for a subsample of the most recent wave, our analysis here is limited to cross-sectional data not directly comparable to the panel-based results reported earlier.

Our analysis of the biomarker sample consists of three steps. We first re-estimate Eq. 4 to assess the relationship between self-reported diabetes with labor outcomes, but this time for the biomarker sample only, using the following specification:

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

where  $c_i$  are community fixed effects which reflect community characteristics such as

---

<sup>12</sup>Further evidence on the effect of health news on labor outcomes is provided by Dillon et al. (2014), who in a very different context, and using a randomized intervention, find that the news stemming from a diagnosis of malaria affect productivity and income, but not labor supply among sugar cane cutters in Nigeria.



access to healthcare, poverty and unemployment in the community. These factors are included as they potentially affect the propensity to develop diabetes, receive a diagnosis, and the labor outcomes of the individuals in the community.<sup>13</sup>

In a second step we then estimate the relations between diabetes, as defined by our biomarker, and labor outcomes, using the following equation:

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

where  $Dbio_i$  is equal to 1 if  $HbA1c \geq 6.5\%$ .

To find the effect of undiagnosed diabetes we then include both variables at the same time and estimate:

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

## 5 Results

### 5.1 Labor outcomes and self-reported diabetes

Table 2 presents the estimation results of Eq. 1 which indicate significant and substantial reductions in the probability of employment for men and women with self-reported diabetes. The coefficients are similar for both sexes, showing a reduction in employment probabilities of over 5 percentage points. Taking into account the lower employment rates for women compared to men, these absolute reductions translate into a relative reductions in employment probabilities of 14% for women and of 6% for men, suggesting a stronger impact of diabetes on women than men.

The results in Columns 3–6 of Table 2 show no significant relationship between self-reported diabetes and wages or working hours. To assess whether this relationship differs

---

<sup>13</sup>We did not include 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.

by the type of work (i.e. work in agriculture requiring strenuous, physical efforts may be more affected by diabetes complications than more sedentary work) we include interaction terms between diabetes and agricultural employment, and between diabetes and self-employment, respectively, using non-agricultural wage employment as the comparison group (and restricting our sample to those employed only). The results in Table ?? panel B, show that the type of work appears to be of less importance, as none of the interaction terms shows up as significant. In the working hours regression, the interaction term of diabetes with agricultural work indicates reduced work supply for this group relative to non-agricultural workers and employees. However, the Wald test to assess the overall significance of the interaction term, cannot reject the null of no interaction effects ( $p = .15$ ).

In summary, we find no evidence for an association between diabetes and wages or working hours. One possible explanation is selection bias, with those with 'mild' or asymptomatic diabetes being more likely to maintain their productivity. Only once complications become increasingly severe would they switch activity (or drop out of the labor market), without going through a notable phase of reduced productivity and labor supply.

To assess whether diabetes affects the selection into different types of work, we estimate FE models of the probability of being in non-agricultural wage employment, agricultural employment or self-employment, respectively. The results in Table 3 only indicate statistically significant effects on the probabilities of females to work in agriculture, possibly due to the higher physical requirements.<sup>14</sup>

---

<sup>14</sup>We prefer this model to a multinomial logit because it allows full control of fixed effects, and it produces results that are straight forward to interpret. As a robustness check we also estimated a multinomial logit model that includes means of time-varying covariates to proxy for fixed effects (see Bell and Jones (2015) and Mundlak (1978)) (Bell and Jones, 2015). The results turn out very similar, both in size and statistical significance.

Table 2: Labor outcomes and self-reported diabetes

	Employment		Log hourly wages		Monthly work hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Panel A: all labour outcomes						
Self-reported	-.054** (.025)	-.059** (.024)	0.054 (.067)	0.081 (.158)	-.524 (1.499)	-1.955 (2.517)
Hausman test	255.260	388.822	278.355	904.858	4101.669	976.631
p-value	0.000	0.000	0.000	0.000	0.000	0.000
N	21388	27341	13828	7068	17616	9112
Panel B: interaction with labor types						
Agricultural worker			-.078* (.044)	-.280 (.186)	-3.577*** (.800)	-4.473* (2.702)
Self-employed			0.028 (.043)	-.144* (.087)	-1.452** (.704)	-4.713*** (1.388)
Self-reported diabetes			0.105 (.076)	0.064 (.169)	0.617 (1.606)	-.524 (2.252)
<i>Interaction terms</i>						
Diabetes $\times$ agriculture			-.242 (.188)	-.409 (.373)	-5.495* (2.833)	-3.535 (22.300)
Diabetes $\times$ self-employed			-.105 (.192)	0.125 (.326)	0.306 (2.503)	-4.149 (4.739)
Hausman test			280.491	912.537	4086.461	995.171
p-value			0.000	0.000	0.000	0.000
N			13828	7068	17616	9112

*Notes* Individual fixed effects regression. Robust standard errors in parentheses. Reference category: dependent non-agricultural worker or employee. 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 3: Selection into types of work and self-reported diabetes

	Males			Females		
	(1) Non-agric.	(2) Agric.	(3) Self-employed	(4) Non-agric.	(5) Agric.	(6) Self-employed
Self-reported diabetes	-.006 (.029)	-.008 (.022)	-.043 (.026)	-.001 (.018)	-.022** (.009)	-.029 (.018)
Hausman test	2196.390	2005.383	1249.080	1126.933		86.400
p-value	0.000	0.000	0.000	0.000		0.000
N	20719	20719	20719	26577	26577	26577

*Notes* Individual fixed effects regression. 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 \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## 5.2 Labor outcomes and time since diagnosis

Given the chronic (and life-long) nature of diabetes, we investigate how soon after the first diagnosis it impacts labor outcomes. Over time, ever more severe complications develop if diabetes remains poorly treated, and the economic effects may therefore increase as time passes. These effects may also vary over time as problems that have led to the diagnosis, as well as psychological effects related to the diagnosis may have stronger impacts immediately post diagnosis. Similarly, management of the disease may cause complications to only appear after a number of years, reducing health, labor supply and productivity years after the initial diagnosis.

Using non-parametric kernel-weighted local polynomial regression, Figure 1 shows that the probability of employment for men shows a more or less steady decline that becomes more pronounced as time progresses. For women, a first drop-off occurs right after diagnosis; thereafter no consistent pattern is observed.<sup>15</sup> A similar analysis for wages shows somewhat less clear dynamics, with a possibly long term negative trend for women but not for men. An analogous picture is obtained for working hours.

Table 4 panel A shows the results of estimating Eq. 2, which indicate that male employment probabilities fall every year, with the biggest effects being observed in the FE model. For women, the coefficient shows a reduction of close to 1 percentage point per year in the FE model, though its statistical significance is lower than in the ordinary least squares (OLS) models.

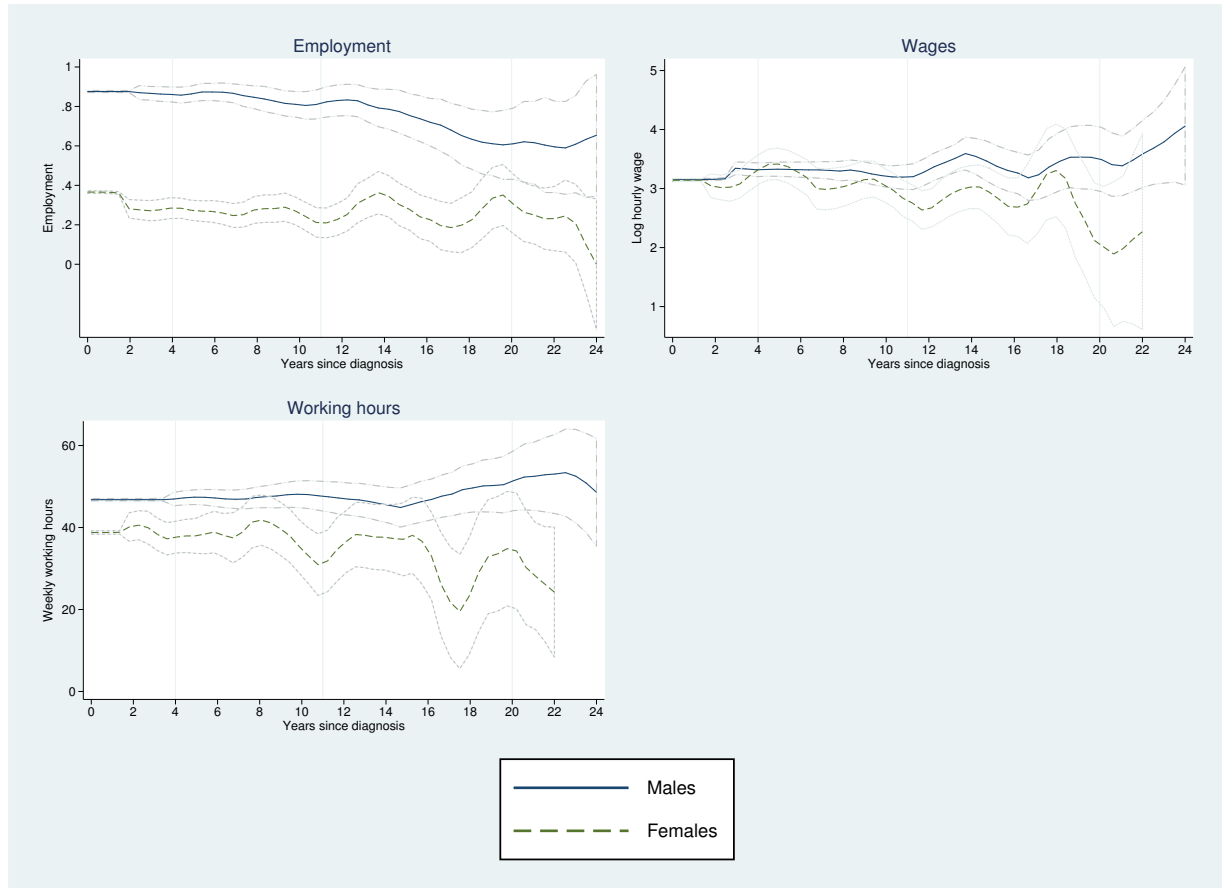
Panel B shows the estimates when using a spline function as described in Eq. 3. Focusing on the FE results, the coefficients provide some evidence for an immediate effect of diabetes, which then levels off for some time upon which it becomes stronger again. However, standard errors are quite large.

The results for wages indicate a reduction in female wages of about 7% per year after diagnosis in the FE model. For men we find no consistent effect. Panel B indicates that

---

<sup>15</sup>Since long run estimations suffer from large standard errors—as the sample size is strongly reduced—this limits its interpretation and we therefore truncate the graphs at a disease duration of 24 years.

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.

there may be a reduction in wages 5–11 years after the initial diagnosis for both men and women. We also find associations for women with more than 20 years of diabetes, but these estimates may be spurious due to the considerably reduced number of observations in this group.<sup>16</sup> Interestingly, the reductions in wages found in the non-linear specification appear exactly at the time where employment probabilities are less affected. This may suggest that at this time reductions in productivity affect wages but are not so severe that they would cause job loss. For working hours there appears to be no consistent relationship

<sup>16</sup>There are only 9 (3) observations for male (female) wages with more than 20 years since diagnosis in wave 3, and 17 (7) in the pooled sample. For male (female) working hours there are 12 (7) observations with more than 20 years since diagnosis in wave 3, and 20 (12) for the pooled sample.

with the time since diagnosis, neither for men nor for women.

In summary, the results suggest a fairly constant decrease in the probability of employment for both men and women and in earnings for women, which contrast with estimates for the USA (Minor, 2013), where no such relationship is observed. Minor (2013) finds a reduction in employment probabilities of 82 percentage points for females after 11 to 15 years and a reduction of 60 percentage points for males after 2-5 years, indicating very large employment penalties in comparison to our results for Mexico. <sup>17</sup>

---

<sup>17</sup>Note that our non-linear results are not directly comparable to Minor's as he used pooled cross-sectional data, made use of dummy variables to indicate time since diagnosis and used different categories of duration. Following the approach of Minor (2013), we find a significant reduction in employment probabilities throughout, regardless of whether we use our duration groups to construct the dummies or the duration groups used by Minor (2013).

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

	Males			Females		
	(1) OLS (Wave 3)	(2) OLS (Pooled)	(3) FE	(4) OLS (Wave 3)	(5) OLS (Pooled)	(6) FE
<b>Dependent variable: Employment status</b>						
<i>Panel A: linear</i>						
Years since SR diagnosis	-.008*** (.002)	-.007*** (.002)	-.017*** (.006)	-.005*** (.002)	-.004*** (.001)	-.009* (.005)
Hausman test			153.024			200.073
p-value			0.000			0.000
<i>Panel B: splines</i>						
Years since SR diagnosis						
0-4	-.007 (.007)	-.007 (.006)	-.026* (.014)	-.010 (.007)	-.015** (.006)	-.017 (.016)
5-11	0.000 (.009)	-.003 (.006)	-.003 (.009)	-.004 (.008)	0.004 (.006)	-.003 (.008)
12-20	-.030** (.012)	-.017* (.010)	-.029* (.016)	0.005 (.008)	-.004 (.006)	-.014 (.011)
> 20	0.011 (.016)	0.007 (.014)	-.046* (.028)	-.010* (.006)	-.003 (.003)	-.015 (.018)
Hausman test			161.953			198.692
p-value			0.000			0.000
N	8217	16292	16292	10467	22407	22407
<b>Dependent variable: Log hourly wages</b>						
<i>Panel A: linear</i>						
Years since SR diagnosis	0.001 (.006)	0.010** (.005)	-.019 (.018)	-.014* (.008)	-.009 (.008)	-.073** (.029)
Hausman test			838.213			93.232
p-value			0.000			0.000
<i>Panel B: splines</i>						
Years since SR diagnosis						
0-4	0.034* (.017)	0.046*** (.016)	0.033 (.055)	0.027 (.031)	0.030 (.026)	0.015 (.138)
5-11	-.041* (.021)	-.037** (.018)	-.055* (.033)	-.039 (.030)	-.034 (.024)	-.101* (.056)
12-20	0.015 (.033)	0.044 (.029)	0.062 (.056)	-.032 (.042)	-.071* (.039)	-.051 (.047)
> 20	0.053 (.054)	0.014 (.040)	-.111 (.104)	-.007 (.028)	0.041*** (.015)	-.204*** (.053)
Hausman test			1037.290			96.266
p-value			0.000			0.000
N	5509	10767	10767	2874	5741	5741
<b>Dependent variable: Weekly working hours</b>						
<i>Panel A: linear</i>						
Years since SR diagnosis	0.069 (.124)	0.048 (.102)	0.181 (.330)	-.020 (.187)	-.124 (.127)	0.208 (.652)
Hausman test			704.904			107.709
p-value			0.000			0.000
<i>Panel B: splines</i>						
Years since SR diagnosis						
0-4	-.033 (.421)	-.233 (.325)	0.709 (.938)	0.739 (.645)	0.470 (.586)	2.014 (2.947)
5-11	0.269 (.539)	0.338 (.399)	-.218 (.568)	-.410 (.728)	-.479 (.553)	-.508 (1.020)
12-20	0.209 (.730)	0.137 (.538)	0.698 (.945)	-.164 (.995)	-.051 (.700)	-.402 (1.207)
> 20	-1.300 (.944)	-.768 (.930)	0.039 (2.184)	-.499 (.930)	-.418 (.305)	8.117*** (1.612)
Hausman test			724.225			112.627
p-value			0.000			0.000
N	6807	13581	13581	3591	7383	7383

Notes The table presents the results of three estimation methods. 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 The OLS and pooled OLS models additionally control for age. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 5.3 Cross-sectional biomarker analysis

Table 5 presents a cross tabulation of self-reported diabetes and biomarker results. Overall, for 80% of the observations the self reports are consistent with the biomarker results (diagonal cell %), 18% are false negatives or undiagnosed and 2% false positives. Due to the nature of the condition and the possible efficacy of its management the false negatives may well include cases where the person received an actual diabetes diagnosis but consequently achieved blood glucose levels below the diabetes threshold due to successfully management of the disease (Flores-Hernández et al., 2015). There are no considerable differences between men and women (results not shown).

Table 5: 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%
Column %	97%	68%	89%
Cell %	<b>71%</b>	<b>18%</b>	-
Self-reported diabetes (N)	129	554	683
Row %	19%	81%	100%
Column %	3%	32%	11%
Cell %	<b>2%</b>	<b>9%</b>	-
Total (N)	4673	1735	6408

Table 6 presents the results from estimating Eq. 4, Eq. 5 and Eq. 6. The results in columns 1 and 2 of Table 6 show that the earlier longitudinal results using self-reported diabetes are robust for the biomarker sample. The coefficients in column 3 and 4 indicate that the relationship becomes much weaker when using diabetes defined by the biomarker instead of self-reported diabetes.<sup>18</sup> Results in columns 5 and 6, obtained from estimating

<sup>18</sup>We also created a dummy variable that additionally to measured diabetes accounted for those with a self-reported diabetes diagnosis but biomarker levels below the diabetes threshold. This is of interest because those who suffer from diabetes but manage to control their sugar levels may obtain test results outside the diabetes range. If one would choose to believe there is no misreporting, this can be seen as representing the entire diabetes population. The coefficients and their statistical significance are only



Eq. 6, the coefficient for the biomarker diabetes population now reflects the effect of undiagnosed diabetes, as the regression includes a control for self-reported diabetes, revealing now statistically significant reductions in any labor outcome for those with undiagnosed diabetes.

Table 6: Biomarker results

	Employment		Log hourly wages		Weekly working hours	
	(1)	(2)	(3)	(4)	(5)	(6)
	Males	Females	Males	Females	Males	Females
<b>Panel A: Diabetes (self-reported)</b>						
Self-reported diabetes	-.051** (.026)	-.044* (.023)	-.010 (.065)	-.040 (.113)	-.293 (1.305)	-.751 (2.178)
<b>Panel B: Diabetes (biomarker)</b>						
HbA1c $\geq 6.5$	-.012 (.016)	-.031* (.018)	-.007 (.044)	-.057 (.070)	-.088 (.844)	1.153 (1.462)
<b>Panel C: Interacting self-reported and biomarker diabetes</b>						
Self-reported diabetes	-.030 (.056)	-.001 (.050)	0.328* (.192)	-.002 (.226)	1.756 (3.248)	6.183 (4.356)
HbA1c $\geq 6.5$ (dummy)	0.006 (.018)	-.017 (.020)	0.017 (.050)	-.054 (.078)	0.168 (.960)	2.577 (1.640)
Self-reported diabetes $\times$ HbA1c $\geq 6.5$	-.029 (.062)	-.042 (.058)	-.396* (.209)	-.010 (.259)	-2.511 (3.594)	-10.883** (5.153)
Linear combinations (F-test)						
Self-reported	-.059** (.029)	-.043 (.031)	-.068 (.084)	-.012 (.136)	-.755 (1.570)	-4.700 (2.777)
Biomarker	-.024 (.060)	-.060 (.054)	-.378* (.202)	-.064 (.250)	-2.344 (3.463)	-8.306 (4.910)
N	2785	3623	1803	884	2302	1144

*Notes* 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 calender 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 this stems from selection into the diagnosed population of those with a more severe diabetes—proxied by higher HbA1c levels—and a therefore higher risk to lose their job, we test whether those with self-reported diabetes have higher levels of HbA1c than those who were not diagnosed, using a t-test, and find that in particular men do.<sup>19</sup> We then extend the model to take into account the severity of diabetes more marginally different to those presented in columns 3 and 4 of Table 8, which is why we do not present them here.

<sup>19</sup>Men: Self-reported diabetes HbA1c of 9.0% vs 8.5% in those undiagnosed ( $p < 0.01$ ); Women: Self-reported diabetes HbA1c of 8.9% vs 8.7% in those undiagnosed ( $p < 0.05$ ).

explicitly, using the measured HbA1c levels. If current severity would be related to labour outcomes and explain the difference in effects of self-reported and undiagnosed diabetes, one would expect (i) that those with higher HbA1c levels have a lower probability of employment, and (ii) that the inclusion of diabetes severity weakens the coefficient of self-reported diabetes. To investigate this, we replace the indicator variable for diabetes with a variable that takes the value zero for levels below and the actual value of HbA1c for those above the diabetes threshold. The results in Table 7, panel A do not find a consistent relationship between increased HbA1c levels and the probability of employment, suggesting that disease severity may not explain the different employment effects of diabetes.

Secondly, to assess whether the differences in effects are driven by differences in subjective health, we include additional controls for self-reported health. The results are reported in Table 7, panel B, and indicate that the relationship between employment and self-reported diabetes becomes weaker, resulting in reduced difference between self-reported and undiagnosed diabetes. For women, the point estimates for self-reported diabetes and undiagnosed diabetes are now virtually of the same size, suggesting that differences in general health could be driving the above results, though the difference was not very big to begin with. For men, it appears that differences in health between the two groups play a role, however, other unobserved factors may still be important.

Thirdly, instead of controlling for subjective health we include a battery of indicators for other chronic diseases that are often related to diabetes. In detail, we control for overweight and obesity (based on anthropometrically measured body mass index (BMI)) and self-reports of a diagnosis of hypertension and heart disease. If those diagnosed with diabetes are more likely to experience adverse employment outcomes because they are more likely to suffer from one of these diseases, then accounting for them should lead to a sizeable reduction in the coefficient of self-reported diabetes. In line with the results in panel B, we find a reduction in the coefficient for self-reported diabetes in both men and women, though the reduction here is much bigger for men than for women. Having had a

diagnosis of heart disease is significantly associated with lower employment probabilities for men, and overall the coefficient for self-reported diabetes in men is reduced by about one percentage point after the inclusion of these diabetes related diseases.<sup>20</sup>

Table 7: Self-reported diabetes, biomarkers, diabetes severity and self-reported health and their association with labor outcomes

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Self-reported diabetes	-.051** (.026)	-.044* (.023)	-.010 (.065)	-.040 (.113)	-.293 (1.305)	-.751 (2.178)
Biomarker tested diabetes	-.012 (.016)	-.031* (.018)	-.007 (.044)	-.057 (.070)	-.088 (.844)	1.153 (1.462)
Unidagnosed diabetes	0.006 (.018)	-.017 (.020)	0.017 (.050)	-.054 (.078)	0.168 (.960)	2.577 (1.640)
False positives	-.030 (.056)	-.001 (.050)	0.328* (.192)	-.002 (.226)	1.756 (3.248)	6.183 (4.356)
<b>Linear combinations (F-test)</b>						
Linear Combination: Self-reported	-.059 (.029)	-.043 (.031)	-.068 (.084)	-.012 (.136)	-.755 (1.570)	-4.700 (2.777)
Linear Combination: Biomarker	-.024 0.060	-.060 0.054	-.378 0.202	-.064 0.250	-2.344 3.463	-8.306 4.910
<b>Diabetes severity</b>						
False positives	-.050 (.065)	-.013 (.041)	0.223* (.117)	0.029 (.178)	1.650 (2.504)	3.464 (3.527)
Undiagnosed diabetes	0.001 (.002)	-.003 (.002)	0.002 (.005)	-.005 (.008)	-.005 (.104)	0.256 (.192)
<b>Controlling for other chronic diseases</b>						
Diagnosed diabetes=1	0.008 (.050)	0.008 (.050)	0.340* (.195)	0.033 (.227)	2.021 (3.277)	6.769 (4.390)
HbA1c $\geq 6.5\%$ =1	-.015 (.021)	-.015 (.021)	0.013 (.050)	-.036 (.079)	0.124 (.962)	2.372 (1.649)
<b>Controlling for self-reported health</b>						
HbA1c $\geq 6.5\%$ =1	0.003 (.017)	-.023 (.019)	-.004 (.049)	-.051 (.075)	-.066 (.926)	1.829 (1.569)
Diagnosed diabetes=1	-.036 (.026)	-.023 (.027)	0.002 (.079)	0.060 (.121)	0.123 (1.433)	-2.191 (2.386)

*Notes* 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 calender 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). \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

It is interesting to contrast these results with those obtained from Brown, Perez, et al. (2011), one of two other studies that use biomarker data. Using data for a Mexican

<sup>20</sup>Further analysis indicates that the reductions in the diabetes coefficient for men appear after the inclusion of heart disease as well as hypertension, while overweight and obesity play a minor role.

American population the study finds that once diabetes is diagnosed blood glucose levels themselves have little effect on labour outcomes. This is not surprising given that HbA1c levels only provide a picture of blood glucose levels over the last three months. They therefore may not be representative of blood glucose levels in the years before and after the diabetes diagnosis which ultimately determine how soon complications appear and how severe they will be.

In the same vein, Minor and MacEwan (2016) find for a USA population that people with undiagnosed diabetes experience smaller employment penalties than those self-reporting the condition. The study finds, however, much bigger effects than we do when estimating the impact of biometrically measured diabetes. One possible explanation for the difference is that the undiagnosed population made up a much smaller share of the overall diabetes population compared to our context, and is therefore likely to have a distinct profile.

Table 8: Self-reported diabetes, biomarkers, diabetes severity and self-reported health and their association with labor outcomes

	Employment		Log hourly wages		Weekly working hours	
	(1)	(2)	(3)	(4)	(5)	(6)
	Males	Females	Males	Females	Males	Females
<b>Panel D: Interacting self-reported and HbA1c (if HbA1c <math>\geq 6.5</math>)</b>						
Self-reported diabetes	-.050 (.065)	-.013 (.041)	0.223* (.117)	0.029 (.178)	1.650 (2.504)	3.464 (3.527)
HbA1c (if HbA1c $\geq 6.5$ )	0.001 (.002)	-.003 (.002)	0.002 (.005)	-.005 (.008)	-.005 (.104)	0.256 (.192)
Self-reported diabetes $\times$ HbA1c (if HbA1c $\geq 6.5$ )	-.001 (.006)	-.002 (.004)	-.029** (.013)	-.005 (.021)	-.231 (.282)	-.746* (.348)
N	2785	3623	1803	884	2302	1144
<b>Panel C (self-reported health)</b>						
HbA1c $\geq 6.5\%$	0.005 (.018)	-.018 (.020)	0.020 (.050)	-.052 (.078)	0.116 (.961)	2.728* (1.639)
Self-reported diabetes	-.013 (.056)	0.007 (.050)	0.353* (.193)	0.043 (.225)	2.189 (3.257)	4.672 (4.374)
Self-reported diabetes $\times$ HbA1c (if HbA1c $\geq 6.5$ )	-.028	-.041	-.416**	0.024	-2.545	-9.725*
Self-reported health status						
good	0.023 (.025)	0.057* (.034)	0.059 (.074)	-.114 (.124)	-1.137 (1.376)	3.203 (2.501)
fair	-.006 (.026)	0.005 (.034)	0.027 (.076)	-.156 (.129)	-1.588 (1.424)	4.404* (2.606)
bad	-.127*** (.043)	-.023 (.046)	-.019 (.135)	-.371* (.189)	-6.204** (2.521)	6.999* (3.853)
very bad	-.166 (.110)	0.117 (.116)	-.384 (.301)	0.314 (.440)	-2.155 (6.447)	-15.878* (9.029)
N	2785	3621	1803	883	2302	1143
<b>Panel B (other chronic diseases)</b>						
Self-reported diabetes	-0.0160 (-0.28)	0.00791 (0.16)	0.340 (1.75)	0.0330 (0.15)	2.021 (0.62)	6.769 (1.54)
HbA1c $\geq 6.5$ (dummy)	0.00367 (0.20)	-0.0148 (-0.72)	0.0127 (0.25)	-0.0362 (-0.46)	0.124 (0.13)	2.372 (1.44)
Self-reported diabetes $\times$ HbA1c (if HbA1c $\geq 6.5$ )	-0.0347 (-0.56)	-0.0504 (-0.87)	-0.398 (-1.89)	0.0297 (0.11)	-2.662 (-0.74)	-11.52* (-2.23)
Overweight	0.0204 (1.27)	-0.00266 (-0.14)	0.0890 (1.93)	0.0354 (0.49)	-0.140 (-0.16)	-1.582 (-1.04)
Obese	0.0222 (1.19)	-0.0306 (-1.53)	0.0713 (1.32)	-0.0200 (-0.26)	1.141 (1.12)	-0.535 (-0.34)
Hypertension	-0.0354 (-1.43)	-0.0205 (-0.94)	-0.128 (-1.74)	-0.0796 (-0.87)	-0.234 (-0.17)	-2.465 (-1.30)
Heart disease	-0.166** (-2.80)	-0.0465 (-0.92)	0.0259 (0.14)	-0.0172 (-0.08)	-2.940 (-0.80)	-5.319 (-1.06)
N	2785	3621	1803	883	2302	1143

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

## 6 Conclusion

Diabetes is now one of the most common chronic diseases in middle- and high-income countries, with the potential to severely impact the health and economic well-being of those affected. Yet rigorous evidence on the economic consequences for these countries remains scarce.

To address key methodological challenges this paper uses rich longitudinal panel data from Mexico that also contain biomarker data. The biomarker data confirm the alarming levels of clinically tested diabetes, 27%, and indicate that a large proportion of these (68% of those with diabetes or 18% of the population) are unaware that they have the condition.

The paper finds evidence for adverse effects of self-reported diabetes on the probability of being employed, but not on wages or hours worked, using fixed effects estimation. Considering different types of work, the relationship between self-reported diabetes, and wages and hours worked remains weak, but the results also suggest occupational selection with women with self-reported diabetes less likely to work in agriculture. A dynamic analysis finds that the employment probability falls gradually over the years after having been diagnosed with this chronic condition. Overall, in particular female employment chances and potentially also wages are reduced, taking into account their lower baseline employment rate.

Making use of the biomarker data allows us to both test reporting error and the effects of undiagnosed diabetes. Results suggest lower employment for the diagnosed, especially for men, while there is no statistically significant employment effect for undiagnosed men or women. This lack of employment effect for undiagnosed men seems to stem from better general health and fewer complications rather than less severe diabetes. The results highlight both the importance of the economic impact of diabetes, and the need to take into account (especially female) undiagnosed patients.

Our findings bear several implications. First, when interpreting labor market impact estimates relying on self-reported diabetes, one cannot assume that the results extend to

those with undiagnosed diabetes. However, simply combining self-reported and undiagnosed in one diabetes category may not be ideal either, as doing so will fail to account for the heterogeneity between the groups in terms of health information, their actual time of living with diabetes and consequently their subjective as well as true health status, leading to a potentially important loss of information. By contrast, accounting for both groups separately acknowledging their inherent differences, allows to gain information about the distribution of the economic burden across the two groups.

Our results add further weight to the case for reducing the incidence and progression of diabetes. On top of the well-documented health benefits, it appears there are considerable gains to be had by increasing the productive lifespan of people. This is of particular importance in low- and middle-income countries, where parental health shocks, related job loss and increasing health expenditures can have repercussions across the entire household. Other family members, including children, may be forced to increase their labor supply and to reduce non-health expenditures in order to prevent a deterioration of the household's economic situation. This can lead to forgone investments into child education, showcasing the potential for adverse long-term effects of health shocks due to diabetes (Bratti and Mendola, 2014). Moreover, the large proportion of previously undiagnosed cases indicates that diagnosis—at least in Mexico—still happens too late or not at all. This reduces the possibilities to prevent complications via treatment and self-management, consequently increasing the risk of severe complications appearing earlier. Hence, much of the health and economic burden may be prevented by earlier diagnosis and, given the generally limited success in achieving good blood-glucose control in Mexico, better treatment of those already diagnosed with diabetes. In particular exploring why women experience such strong economic effects could help to reduce the diabetes burden. Ultimately, there is a need to invest in the prevention of diabetes. Taxation of sugar sweetened beverages may be one promising way forward (Colchero et al., 2016), though the long-term effects remain to be demonstrated. Further, considering the double-disease burden of non-communicable

and communicable diseases and malnutrition in many low- and middle-income countries, investments in maternal and child health may not only reduce the current disease burden but would likely reduce the future incidence of diabetes, given the established links between early life health status and later life incidence of diabetes and other chronic diseases (**Hanson2012**; Li et al., 2010; Sotomayor, 2013).

Our results indicate a significant economic burden of diabetes and it is unlikely that it will be reduced in the near future given that diabetes has started appearing at an increasingly younger age in many low- and middle-income countries (LMICs), causing people to live with the disease for larger parts of their productive lifespan, possibly exacerbating the economic effects of reduced employment due to diabetes (Hu, 2011; Villalpando et al., 2010). Therefore, population level measures as well as efforts to improve early-life health are needed to prevent a further increase in diabetes, as is a better integration of diabetes care in the existing health system, to reduce its burden now and in the future.



# Appendix

## A 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 labor 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'. If they answered 'yes', they were asked if they received treatment for diabetes and the type of treatment they received.

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. However, Zajacova et al. (2010) assess the consistency of a self-reported cancer diagnosis over time in the USA. The study found that 30% of those who had reported a cancer diagnosis at an earlier point failed to report the diagnosis at a later point in time. A more recent diagnosis was found to be reported with greater consistency possibly due to increasing recall problems as time since diagnosis advanced.

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

As discussed at the end of section 3, 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 (Yuan et al., 2015). Our data showed a similar pattern, with a negligible proportion (3%) of the respondents who are tested negative self-reporting to suffer from diabetes, while the majority of those who are tested positive (68%) do not self-report suffering from 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 Table 8. 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 Table 8) 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 Table 8). 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 Table 8), given that most diabetes self-reports tend to be correct.

Table 9: Inconsistencies in diabetes self-report in MxFLS.

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

We then tested if those respondents we categorized as not having a diabetes diagnosis based on above rules were actually more likely to not have 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 8) and 49 with one self-report of diabetes (from lines 1 and 2 in Table 8)). Figure 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  $\text{HbA1c} \geq 6.5\%$  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: Kernel density of HbA1c values for those with one inconsistent and two inconsistent reports.

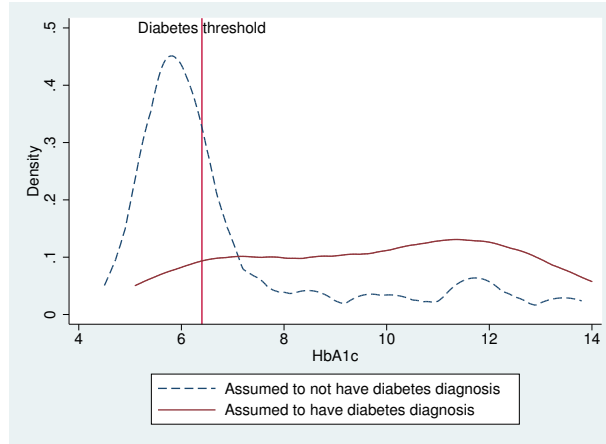


Table 10: Studies estimating the relationship between diabetes and employment

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	Liu and Zhu (2014)
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	Seuring, Goryakin, et al. (2015)

Table 10: Studies estimating the relationship between diabetes and employment

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	Brown, Pagán, et al. (2005)
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 the male wages. No effects found for women.	Exogeneity assumed	Probit and Heckman selection model	Brown, Perez, et al. (2011)
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	Minor (2011)

Table 10: Studies estimating the relationship between diabetes and employment

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	Minor (2013)

Table 10: Studies estimating the relationship between diabetes and employment

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 HbA1c levels for subsample	and 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	Minor and MacEwan (2016)

Table 10: Studies estimating the relationship between diabetes and employment

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	Latif (2009)



Table 10: Studies estimating the relationship between diabetes and employment

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
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	Zhang et al. (2009)

## Acknowledgements

We are grateful to the participants at the European Health Economics Association PhD-Supervisor conference September 2015 in Paris, the Health, Education and Labor Market Outcomes Workshop at the WifOR Institute in October 2015 in Darmstadt, Germany, seminar participants at the Centre for Health Economics at York University, and Max Bachmann for helpful comments.

## References

- Alegre-Díaz, J., Herrington, W., López-Cervantes, M., Gnatiuc, L., Ramirez, R., Hill, M., Baigent, C., McCarthy, M. I., Lewington, S., Collins, R., Whitlock, G., Tapia-Conyer, R., Peto, R., Kuri-Morales, P., and Emberson, J. R. (2016). “Diabetes and Cause-Specific Mortality in Mexico City.” *New England Journal of Medicine* 375 (20), 1961–1971. DOI: 10.1056/NEJMoa1605368.
- Ayyagari, P., Grossman, D., and Sloan, F. (2011). “Education and health: evidence on adults with diabetes.” *International Journal of Health Care Finance and Economics* 11 (1), 35–54. DOI: 10.1007/s10754-010-9087-x.
- Baird, S., Gong, E., McIntosh, C., and Özler, B. (2014). “The heterogeneous effects of HIV testing.” *Journal of Health Economics* 37, 98–112. DOI: 10.1016/j.jhealeco.2014.06.003.
- Barquera, S., Campos-Nonato, I., Aguilar-Salinas, C., Lopez-Ridaura, R., Arredondo, A., and Rivera-Dommarco, J. (2013). “Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy.” *Globalization and Health* 9 (1), 3. DOI: 10.1186/1744-8603-9-3.
- Barquera, S., Hernandez-Barrera, L., Tolentino, M. L., Espinosa, J., Ng, S. W., Rivera, J. A., and Popkin, B. M. (2008). “Energy Intake from Beverages Is Increasing among

- Mexican Adolescents and Adults.” *Journal of Nutrition* 138 (12), 2454–2461. DOI: 10.3945/jn.108.092163.
- Basu, S., Yoffe, P., Hills, N., and Lustig, R. H. (2013). “The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data.” *PLoS ONE* 8 (2), e57873. DOI: 10.1371/journal.pone.0057873.
- Beagley, J., Guariguata, L., Weil, C., and Motala, A. a. (2014). “Global estimates of undiagnosed diabetes in adults.” *Diabetes Research and Clinical Practice* 103 (2), 150–160. DOI: 10.1016/j.diabres.2013.11.001.
- Bell, A. and Jones, K. (2015). “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data.” *Political Science Research and Methods* 3 (01), 133–153. DOI: 10.1017/psrm.2014.7.
- Bello-Chavolla, O. Y., Rojas-Martinez, R., Aguilar-Salinas, C. A., and Hernández-Avila, M. (2017). “Epidemiology of diabetes mellitus in Mexico.” *Nutrition Reviews* 75 (suppl 1), 4–12. DOI: 10.1093/nutrit/nuw030.
- Bergemann, A., Grönqvist, E., and Gudbjörnsdóttir, S. (2011). “The effects of job displacement on the onset and progression.” *Netspar Discussion Paper* (25).
- Bratti, M. and Mendola, M. (2014). “Parental health and child schooling.” *Journal of Health Economics* 35 (1), 94–108. DOI: 10.1016/j.jhealeco.2014.02.006.
- Brown, H. S., Pagán, J. A., and Bastida, E. (2005). “The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit.” *Health Economics* 14 (5), 537–544. DOI: 10.1002/hec.942.
- Brown, H. S., Perez, A., Yarnell, L. M., Pagan, J. a., Hanis, C. L., Fischer-Hoch, S. P., and McCormick, J. B. (2011). “Diabetes and employment productivity: does diabetes management matter?” *American Journal of Managed Care* 17 (8), 569–576.
- Brown, T. T. (2014). “How effective are public health departments at preventing mortality?” *Economics & Human Biology* 13, 34–45. DOI: 10.1016/j.ehb.2013.10.001.

- Cawley, J., Maclean, J. C., Hammer, M., and Wintfeld, N. (2015). “Reporting error in weight and its implications for bias in economic models.” *Economics & Human Biology* 19, 27–44. DOI: 10.1016/j.ehb.2015.07.001.
- Colchero, M. A., Popkin, B. M., Rivera, J. A., and Ng, S. W. (2016). “Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study.” *British Medical Journal* 352, h6704. DOI: 10.1136/bmj.h6704.
- Crimmins, E., McDade, T., Rubalcava, L., Seeman, T., Teruel, G., and Thomas, D. (2015). “Health of the Mexican population: Results from the Mexican Family Life Survey (MxFLS).”
- Currie, J. and Vogl, T. (2013). “Early-Life Health and Adult Circumstance in Developing Countries.” *Annual Review of Economics* 5 (1), 1–36. DOI: 10.1146/annurev-economics-081412-103704.
- Dillon, A., Friedman, J., and Serneels, P. M. (2014). “Health information, treatment, and worker productivity: Experimental evidence from malaria testing and treatment among Nigerian sugarcane cutters.” *World Bank Policy Research Working Paper* (7120).
- Eriksson, A.-K., Donk, M. van den, Hilding, A., and Ostenson, C.-G. (2013). “Work Stress, Sense of Coherence, and Risk of Type 2 Diabetes in a Prospective Study of Middle-Aged Swedish Men and Women.” *Diabetes Care* 36 (9), 2683–2689. DOI: 10.2337/dc12-1738.
- Ewijk, R. van (2011). “Long-Term Health Effects on the Next Generation of Ramadan Fasting during Pregnancy.” *Journal of Health Economics* 30 (6), 1246–1260.
- Filmer, D. and Pritchett, L. (2001). “Estimating wealth effects without expenditure data—Or tears: An application to educational enrollments in states of India.” *Demography* 38 (1), 115–132.
- Flores-Hernández, S., Saturno-Hernández, P. J., Reyes-Morales, H., Barrientos-Gutiérrez, T., Villalpando, S., and Hernández-Ávila, M. (2015). “Quality of Diabetes Care: The Challenges of an Increasing Epidemic in Mexico. Results from Two National Health

- Surveys (2006 and 2012).” *Plos One* 10 (7), e0133958. DOI: 10.1371/journal.pone.0133958.
- Frankenberg, E., Ho, J. Y., and Thomas, D. (2015). “Biological Health Risks and Economic Development.” *NBER Working Paper* (21277).
- Gong, E. (2015). “HIV Testing and Risky Sexual Behaviour.” *The Economic Journal* 125 (582), 32–60. DOI: 10.1111/ecoj.12125.
- Gregg, E. W., Chen, H., Wagenknecht, L. E., Clark, J. M., Delahanty, L. M., Bantle, J., Pownall, H. J., Johnson, K. C., Safford, M. M., Kitabchi, A. E., Pi-Sunyer, F. X., Wing, R. R., Bertoni, A. G., and Look AHEAD Research Group, for the (2012). “Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes.” *Journal of the American Medical Association* 308 (23), 2489. DOI: 10.1001/jama.2012.67929.
- Gutiérrez-delgado, C. and Guajardo-barrón, V. (2009). “The double burden of disease in developing countries: The Mexican experience.” *Advances in Health Economics and Health Services Research*. Ed. by C. Gutiérrez-Delgado and V. Guajardo-Barrón. Vol. 21. Advances in Health Economics and Health Services Research. Bingley: Emerald Group Publishing, 3–22. DOI: 10.1108/S0731-2199(2009)0000021004.
- Heraclides, A. M., Chandola, T., Witte, D. R., and Brunner, E. J. (2012). “Work Stress, Obesity and the Risk of Type 2 Diabetes: Gender-Specific Bidirectional Effect in the Whitehall II Study.” *Obesity* 20 (2), 428–433. DOI: 10.1038/oby.2011.95.
- Hu, F. B. (2011). “Globalization of Diabetes: The role of diet, lifestyle, and genes.” *Diabetes Care* 34 (6), 1249–1257. DOI: 10.2337/dc11-0442.
- International Diabetes Federation (2015). *Diabetes Atlas*. 7th ed. International Diabetes Federation.
- Kapteyn, A., Smith, J. P., and Van Soest, A. (2009). “Work disability, work, and justification bias in Europe and the United States.” *Unpublished*.
- Latif, E. (2009). “The impact of diabetes on employment in Canada.” *Health Economics* 18 (5), 577–589. DOI: 10.1002/hec.1390.

- Li, Y., He, Y., Qi, L., Jaddoe, V. W., Feskens, E. J. M., Yang, X., Ma, G., and Hu, F. B. (2010). “Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood.” *Diabetes* 59 (10), 2400–2406. DOI: 10.2337/db10-0385.
- Lim, E. L., Hollingsworth, K. G., Aribisala, B. S., Chen, M. J., Mathers, J. C., and Taylor, R. (2011). “Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol.” *Diabetologia* 54 (10), 2506–2514. DOI: 10.1007/s00125-011-2204-7.
- Liu, X. and Zhu, C. (2014). “Will knowing diabetes affect labor income? Evidence from a natural experiment.” *Economics Letters* 124 (1), 74–78. DOI: 10.1016/j.econlet.2014.04.019.
- Minor, T. (2011). “The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey.” *Health Economics*. 15th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research, ISPOR 2010 Atlanta, GA United States. Conference Start: 20100515 Conference End: 20100519 20 (12), 1468–1486. DOI: 10.1002/hec.1685.
- (2013). “An investigation into the effect of type I and type II diabetes duration on employment and wages.” *Economics & Human Biology* 11 (4), 534–544. DOI: 10.1016/j.ehb.2013.04.004.
- Minor, T. and MacEwan, J. P. (2016). “A comparison of diagnosed and undiagnosed diabetes patients and labor supply.” *Economics & Human Biology* 20, 14–25. DOI: 10.1016/j.ehb.2015.10.003.
- Mundlak, Y. (1978). “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46 (1), 69–85. DOI: 10.2307/1913646.
- NCD Risk Factor Collaboration (2016). “Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants.” *The Lancet* 387, 1513–1530. DOI: 10.1016/S0140-6736(16)00618-8.

- Nouwen, A., Nefs, G., Caramlau, I., Connock, M., Winkley, K., Lloyd, C. E., Peyrot, M., and Pouwer, F. (2011). "Prevalence of depression in individuals with impaired glucose metabolism or undiagnosed diabetes: A systematic review and meta-analysis of the European Depression in Diabetes (EDID) research consortium." *Diabetes Care* 34 (3), 752–762. DOI: 10.2337/dc10-1414.
- O'Neill, D. and Sweetman, O. (2013). "The consequences of measurement error when estimating the impact of obesity on income." *IZA Journal of Labor Economics* 2 (1), 3. DOI: 10.1186/2193-8997-2-3.
- Paddison, C. a. M., Eborall, H. C., French, D. P., Kinmonth, a. L., Prevost, a. T., Griffin, S. J., and Sutton, S. (2011). "Predictors of anxiety and depression among people attending diabetes screening: A prospective cohort study embedded in the ADDITION (Cambridge) randomized control trial." *British Journal of Health Psychology* 16 (1), 213–226. DOI: 10.1348/135910710X495366.
- Perks, T. A. (2015). "Obesity and its relation to employment income: Does the bias in self-reported BMI matter?" *Canadian Studies in Population* 42 (3-4), 1–10.
- Reynoso-Noverón, N., Mehta, R., Almeda-Valdes, P., Rojas-Martinez, R., Villalpando, S., Hernández-Ávila, M., and Aguilar-Salinas, C. a. (2011). "Estimated incidence of cardiovascular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey." *Cardiovascular Diabetology* 10 (1), 1. DOI: 10.1186/1475-2840-10-1.
- Rubalcava, L. and Teruel, G. (2013). "User's Guide for the Mexican Family Life Survey Third Round."
- Schaller, J. and Stevens, A. H. (2015). "Short-run effects of job loss on health conditions, health insurance, and health care utilization." *Journal of Health Economics* 43, 190–203. DOI: 10.1016/j.jhealeco.2015.07.003.

- Schneiderman, N., Ironson, G., and Siegel, S. D. (2005). "Stress and Health: Psychological, Behavioral, and Biological Determinants." *Annual Review of Clinical Psychology* 1 (1), 607–628. DOI: 10.1146/annurev.clinpsy.1.102803.144141.
- Schneiderman, N., Llabre, M., Cowie, C. C., Barnhart, J., Carnethon, M., Gallo, L. C., Giachello, A. L., Heiss, G., Kaplan, R. C., LaVange, L. M., Teng, Y., Villa-Caballero, L., and Avilés-Santa, M. L. (2014). "Prevalence of Diabetes Among Hispanics/Latinos From Diverse Backgrounds: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)." *Diabetes Care* 37 (8), 2233–2239. DOI: 10.2337/dc13-2939.
- Seuring, T., Archangelidi, O., and Suhrcke, M. (2015). "The Economic Costs of Type 2 Diabetes: A Global Systematic Review." *Pharmacoeconomics* 33 (8), 811–831. DOI: 10.1007/s40273-015-0268-9.
- Seuring, T., Goryakin, Y., and Suhrcke, M. (2015). "The impact of diabetes on employment in Mexico." *Economics & Human Biology* 18, 85–100. DOI: 10.1016/j.ehb.2015.04.002.
- Sicree, B. R., Shaw, J., and Zimmet, P. (2011). *The Global Burden: Diabetes and Impaired Glucose Tolerance*. Brussels, Belgium: International Diabetes Federation.
- Slade, A. N. (2012). "Health Investment Decisions in Response to Diabetes Information in Older Americans." *Journal of Health Economics* 31 (3), 502–520.
- Sotomayor, O. (2013). "Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes." *Economics & Human Biology* 11 (3), 281–293. DOI: 10.1016/j.ehb.2012.02.009.
- Thoolen, B. J., De Ridder, D. T., Bensing, J. M., Gorter, K. J., and Rutten, G. E. (2006). "Psychological outcomes of patients with screen-detected type 2 diabetes: The influence of time since diagnosis and treatment intensity." *Diabetes Care* 29 (10), 2257–2262. DOI: 10.2337/dc06-0617.
- Thornton, R. L. (2008). "The Demand for, and Impact of, Learning HIV Status." *American Economic Review* 98 (5), 1829–1863. DOI: 10.1257/aer.98.5.1829.



- Villalpando, S., Cruz, V. de la, Rojas, R., Shamah-Levy, T., Avila, M. A., Gaona, B., Rebollar, R., and Hernández, L. (2010). “Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey.” *Salud Publica de Mexico* 52 Suppl 1 (1), S19–26.
- Williams, A. L. et al. (2013). “Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico.” *Nature* 506 (7486), 97–101. DOI: 10.1038/nature12828.
- World Health Organization (2011). “Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation.”
- Yuan, X., Liu, T., Wu, L., Zou, Z.-Y., and Li, C. (2015). “Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study.” *British Medical Journal Open* 5 (4), e006633–e006633. DOI: 10.1136/bmjopen-2014-006633.
- Zajacova, A., Dowd, J., Schoeni, R. F., and Wallace, R. B. (2010). “Consistency and precision of cancer reporting in a multiwave national panel survey.” *Population Health Metrics* 8 (1), 20. DOI: 10.1186/1478-7954-8-20.
- Zhang, X., Zhao, X., and Harris, A. (2009). “Chronic Diseases and Labour Force Participation in Australia.” *Journal of Health Economics* 28 (1), 91–108. DOI: 10.1016/j.jhealeco.2008.08.001.
- Zhao, M., Konishi, Y., and Glewwe, P. (2013). “Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption.” *Journal of Health Economics* 32 (2), 367–385. DOI: 10.1016/j.jhealeco.2012.11.007.