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

Till Seuring[†1,2], Pieter Serneels[2], and Marc Suhrcke[3]

[1] *Leibniz Institute for Prevention Research and Epidemiology - BIPS*
[2] *University of East Anglia*
[3] *University of York*

March 17, 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. Making use of 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, using fixed effects estimation. Self-reported diabetes reduces the probability of employment with 5 percentage points for men and close to 6 percentage points for women. 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. Using unique biomarker data for a large subsample, we find that two thirds of those tested positive do not self-report diabetes, while 19% of those who self-report have levels below the clinical threshold. We contrast the impact of self-reported and tested diabetes, and find that it is similar for women but not for men. Combined results suggest lower employment for those self-reporting, especially for men, while there is no employment effect for undiagnosed men or women, who tested positive but did not self-report. This lack of employment effect for undiagnosed men seems to stem from 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.

DRAFT

[*]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.

[†]Corresponding author

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, 2014). 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 expected to increase further over the next decades (Meza et al., 2015). Already now, diabetes is the number one cause of death in Mexico (Barquera, Campos-Nonato, et al., 2013).

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 (Villalpando et al., 2010). With treatment as ineffective as it currently is—only a minority achieves adequate blood glucose control (Barquera, Campos-Nonato, et al., 2013)—the earlier start will increase the likelihood of complications during the productive lifespan.

Diabetes is a term used to describe 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, food 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 et al., 2016; Seuring, Archangelidi, et al., 2015).[1] 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, like health endowments, as well as risk preferences may adversely affect health in general and the propensity to develop type 2 diabetes more specifically (Ewijk, 2011; Li et al., 2010; Sotomayor, 2013), and may also affect employment probabilities, wages or working hours—either directly through their effects on contemporaneous productivity (Currie et al., 2013), or indirectly by limiting educational attainment and human capital accumulation (Ayyagari et al., 2011).

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

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 uncovered important differences between those with diagnosed and undiagnosed diabetes. For instance, a diabetes diagnosis may have an effect in itself. People aware of their condition could be less inclined to continue working if this interferes with their disease management; they may also 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). As a consequence labor market effects could be distinct for people with self-reported diabetes versus those unaware of their condition, if results based on self-reports are used to make conclusions about the entire population with diabetes.

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes, and improve upon previous work by paying close attention to the challenges of unobserved heterogeneity, the chronic nature of diabetes and the 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.[2] After that, rich and novel biomarker data from the most recent wave of the MxFLS are used to also explore the role of undiagnosed diabetes—an issue that has remained underexplored 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 people aware of their disease. Wages and working hours, however, remain unaffected. A dynamic analysis indicates that employment probabilities are reduced gradually with each additional year since diagnosis, with some evidence for an even larger effect per year after the initial 10 years.

The biomarker analysis indicates that clinically diagnosed diabetes entails a significant employment penalty for women but not for men. Assessing the effects of both clinical and self-reported diabetes at the same time provides an insight into the labor market impact for those with undiagnosed diabetes—people who are tested positive but did not self-report. The results still indicate that, 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 12 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 developing countries. Liu et al. (2014) exploit a natural experiment in China and find a significant reduction

[2] 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).

in income for those with a recent diagnosis of diabetes. A study for Mexico using cross-section 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, find a negative relationship between the level of glycated hemoglobin (HbA1c) and the probability of employment as well as the male wages. 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. The study finds self-reported diabetes to be endogenous and estimates upward biased compared to IV estimates. In a later study, Minor (2013) shows that employment probabilities decline shortly after diagnosis for men and after about 10 years for women, while wages are not affected by the duration of the condition.

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 estimates (Latif, 2009). For Australia, (Zhang et al., 2009) find reduced labor market participation for men and women as a result of diabetes, with the effects appearing overstated if the endogeneity of diabetes is unaccounted for.

As far as we know, only one paper considers undiagnosed diabetes. Minor et al. (2016) find a negative, but not statistically significant relationship between undiagnosed diabetes and the probability of, in particular, female employment. They further find a negative and statistically significant relationship of self-reported diabetes with employment for both men and women. When combining both the undiagnosed and self-reporting, the relationship with employment remains statistically significant but becomes smaller. However, these biomarker results are based on a very small sample size.

While these studies suggest substantial economic losses for individuals and households due to diabetes, most studies tend to suffer from methodological limitations. Using cross-sectional data, they cannot easily account for unobserved characteristics. A number of papers try to address this bias, typically using the family history of diabetes as the identifying instrument, relying on the genetic and heritable component of type 2 diabetes. However, it remains unclear whether the instrument 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.[3]

Furthermore, most—but not all—studies use self-reported diabetes as a proxy for di-

[3]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).

abetes. Self-reported health data can suffer from several short comings, and introduce non-classical measurement error due to systematic misreporting. This has been shown to cause estimates of economic impacts to be potentially biased and overstated (Cawley et al., 2015; O'Neill et al., 2013; Perks, 2015). In the context of diabetes, the concern is especially related to possible false negatives. False positives might be less of a problem since there seems 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 affecting the effect of diabetes on their labor outcomes.

This paper makes headway to overcome these key limitations in two ways. First, it applies fixed effects (FE) to three waves of panel data, allowing to control for unobserved time invariant characteristics. Second, it makes use of biomarker data for a large subsample of the population. This allows us 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 dynamic effects in the years after diagnosis.

3 Context and Data

Mexico is a middle-income country with among the highest prevalences of obesity and diabetes in the world (International Diabetes Federation, 2015; World Health Organization, n.d.). A recent study showed that diabetes accounted for one-third of all deaths in a large sample of Mexicans 35 to 74 years of age 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 coexist with high levels of infectious diseases in Mexico, exposing the health system to a 'double-disease burden' that increases the pressure to identify treatment priorities and to efficiently use the existing resources, with the treatment of non-communicable diseases not being well integrated in the health care system (Gutiérrez-delgado et al., 2009). Earlier studies point to an underwhelming 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; Barquera, Campos-Nonato, et al., 2013; Flores-Hernández et al., 2015). Additionally, 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 et al., 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 biomarker sample is somewhat older on average, as it includes everybody above the age of 44 but only a random subsample of those aged 44 or below (Crimmins et al., 2015). This also leads to self-reported diabetes being higher for this subsample. The biomarker analysis will therefore be compared for this subsample specifically.

The analysis focuses on the labor outcomes employment, hourly wage and weekly working hours, as well as occupation.[4] About half of the respondents in the sample live in rural areas. Looking at our outcome variables, 86% of men report some form of employment compared to 37% of women. Interestingly, men do not report considerably higher hourly wages than women but work more hours per week. Also, men are working 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 (see Table 1).

In the first part of the analysis we focus on the relationship of labor outcomes with self-reported diabetes[5]. 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 elderly above the age of 64. 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 biomarker data including HbA1c levels for a subsample of respondents.

[4]Employment status is defined as having worked or 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, and this only leads to marginal changes in the coefficients and standard errors, not affecting the interpretation of the results. 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 consider 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).

[5]Self-reported diabetes is based on the survey question: “Have you ever been diagnosed with diabetes?”

Because self-reported diabetes reporting exhibited some inconsistency over time for some of the respondents, we tried to increase the consistency by using disease information from earlier and ensuing waves to infer on the current, missing or inconsistent, diabetes status. Appendix 1 provides details on the applied correction procedures. 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. In order to investigate how this may affect estimates of the labor market impact of diabetes, we use information from a subsample of the 2009-2012 wave, containing over 6000 respondents (everybody aged 45+ and a random subsample of those aged 15–44 (Crimmins et al., 2015)) that have biometrically measured blood glucose values, allowing for the identification of those with undiagnosed diabetes. Throughout, our analysis focuses on the working age population (15–64), and excludes pregnant women and those in school.[6] Apart from self-reported diabetes information available in all waves, we also use information on the self-reported year of diagnosis—available in the third wave—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.

The biomarker data suggest that a relatively 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). As argued by others, these rates of elevated HbA1c levels in Mexico are high when compared to HbA1c data from similar surveys in the USA and China (Frankenberg et al., 2015). A second striking observation is that 68% of males and females who test positive do not self-report and hence are unaware of their condition. This may mean that estimates based on self-reported diabetes as a measure for diabetes are biased, and underlines the importance of comparing them to estimates based biomarker analysis. Of the respondents who self-report 19% have levels below the clinical threshold, possibly because they have managed their disease, although we cannot exclude misreporting of diagnosis.

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.[7]

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

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

[6]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.

[7]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.

Table 1: Descriptive statistics for panel and biomarker sample.

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

Notes Mean values, standard deviations in parenthesis. Results for the other variables, i.e. the Mexican states and wealth, are omitted to save space.

1 at time t if the respondent reports having ever received a diagnosis of diabetes[8], X_{it} is a vector of control variables, c_i represents an individual fixed effect, γ_t represents year dummies, and u_{it} is the error term.

For the relationship of self-reported diabetes with wages and working hours, our empirical models 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 the impact of marriage and children on labor outcomes and the effect of childbearing and related gestational diabetes on the probability of developing type 2 diabetes (Bellamy et al., 2009). 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 et al., 2001)[9]. 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.

Note that 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 labor with health. 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 reverse causality, as job loss does not seem to affect the probability to develop diabetes (Bergemann et al., 2011; Schaller et al., 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 likely time-invariant (Schneiderman, Ironson, et al., 2005). While we cannot exclude a role of time-variant unobserved factors or simultaneity, time-invariant variables, including 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.

[8] We are not able to distinguish between type 1 diabetes and type 2 diabetes using this data. Other studies that tried to assess the effect of type 1 diabetes on labor outcomes have found no association (Minor, 2011; Minor et al., 2016). Including type 1 diabetes therefore likely attenuates any adverse relationship we may find.

[9] 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.

4.2 Labor outcomes and time since diagnosis

Given the chronic nature and irreversibility of diabetes, it is of interest to look at the dynamic effects after diagnosis. We estimate the following model with years since self-reported diagnosis as the right hand side variable:

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

where $\beta_1 Dyears_{it}$ is continuous indicating years since first reported diabetes diagnosis. To capture possible non-linearities in the relationship we then use a spline function that allows for the effect of an additional year with diabetes to vary over time.

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

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 η_n being the place of the n-th node for $n = 1, 2, \ldots, N$. We choose three nodes that—based on visual inspection—best captured any possible non-linearity in the relationship between diabetes duration and labor outcomes, see Figure 1 on page 18. These are located at 4, 11 and 20 years after diagnosis. The first four years should capture any immediate effects of the diagnosis, the years five to eleven should capture any effects of adaptation to the disease. After 11 years it is conceivable that many of the debilitating complications of diabetes would appear that could deteriorate health and lead to adverse effects on labor outcomes. The coefficient δ_n captures the effect of diabetes for the n-th interval. The effects are linear if $\delta_1 = \delta_2 =, \ldots, = \delta_n$.[10][11]

4.3 Labor outcomes and biometrically measured diabetes

Self-reported diabetes only captures part of the diabetes population as many individuals remain undiagnosed; it may also contain cases of people who misreport having diabetes. Estimations based on self-reports may therefore be biased due to one of the following three reasons: First respondent may systematically overreport, 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. Respondents may be concerned about negative stigma associated with the condition or, more importantly, diabetes has remained undiagnosed, leaving people unaware of their condition. Third, a

[10]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'.

[11]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.

diagnosis is more likely to happen for those who are more likely to have visited a doctor, for instance because they are 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 of the impact if they have other attributes that negatively affect labor outcomes, like 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 than those with self-reported diabetes. 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 may, 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. Given that people with undiagnosed diabetes are not more prone to exhibit psychological conditions, this may point towards a causal relationship with the diagnosis and subsequent treatment (Nouwen et al., 2011). Health information may also lead to a change in behaviour. Slade (2012) shows how patients, when learning about their diagnosis of diabetes, change their consumption of alcohol and smoking and start to loose 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 shows that receiving a diabetes diagnosis in itself reduces labor income, possibly through psychological effects of the diagnosis (Liu et al., 2014).[12]

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:

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 X_i + c_i + u_i \tag{4}$$

where c_i are community fixed effects which reflect community characteristics such as 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.[13]

[12]In a very different context Dillon et al. (2014), 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.

[13]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.

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

$$Y_i = \beta_0 + \beta_1 Dbio_i + \beta_2 X_i + c_i + u_i, \tag{5}$$

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:

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i + \beta_3 X_i + c_i + u_i. \tag{6}$$

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. In relative terms—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 by the type of work, as those with diabetes working in an agricultural job that requires strenuous, physical efforts may see their productivity more adversely affected than those engaged in more sedentary work, we include interaction terms between self-reported diabetes and agricultural employment, and between self-reported 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 3 show that while male agricultural workers have lower wages in general, the relationship with diabetes does not depend on the type of work, as none of the interaction terms shows up as significant. In the working hours regression, one interaction term is significant, suggesting that those with self-reported diabetes working in agriculture supply 5 hours less relative to non-agricultural workers and employees. However, because we have more than two work types we cannot draw conclusions solely on the basis of the t-statistic. Using a Wald test to assess the overall significance of the interaction term, we cannot reject the null of no interaction effects ($p = .15$).

In summary, we find no evidence for an association between self-reported diabetes and wages or working hours. One possible explanation is selection bias, as those with 'mild' or asymptomatic diabetes are more likely to remain in their job and continue to earn the same wage. 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 proceed by estimating a FE model applied to the individual probability of being in non-agricultural wage employment, agricultural employment or self-employment, respectively, using a dummy variable indicating the respective type of work as the left hand side variable. The results in Table 4 indicate a negative association with self-employment, though the estimates have quite large standard errors. Women with diabetes are less likely to work in agriculture and potentially self-employment, indicating that having diabetes drives women out of self-employment and agricultural jobs, possibly because these jobs are physically more demanding or because they provide less protection in terms of health and income insurance.[14]

[14]We prefer this model to a multinomial logit because it allows full control of fixed effects, and 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 et al. (2015) and Mundlak (1978)) (Bell et al., 2015), based on the work of Mundlak (1978). The results indicate a very similar pattern both in size and significance.

Table 2: Self-reported diabetes and labor outcomes.

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Self-reported diabetes	$-.054^{**}$	$-.059^{**}$	0.054	0.081	$-.524$	-1.955
	(.025)	(.024)	(.067)	(.158)	(1.499)	(2.517)
Hausman test	255.260	388.822	1084.317	91.096	967.007	106.455
p-value	0.000	0.000	0.000	0.000	0.000	0.000
N	21388	27341	13828	7068	17616	9112

Notes Individual fixed effects regression. Robust standard errors in parentheses. Reference category: dependent non-agricultural worker or employee. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number of children < 6, wealth, health insurance status, age squared and calender year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect of self-reported diabetes on wages and working hours, by type of work.

	Log hourly wage		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females
Agricultural worker	$-.078^{*}$	$-.280$	-3.577^{***}	-4.473^{*}
	(.044)	(.186)	(.800)	(2.702)
Self-employed	0.028	$-.144^{*}$	-1.452^{**}	-4.713^{***}
	(.043)	(.087)	(.704)	(1.388)
Self-reported diabetes	0.105	0.064	0.617	$-.524$
	(.076)	(.169)	(1.606)	(2.252)
Self-reported diabetes x agricultural worker	$-.242$	$-.409$	-5.495^{*}	-3.535
	(.188)	(.373)	(2.833)	(22.300)
Self-reported diabetes x self-employed	$-.105$	0.125	0.306	-4.149
	(.192)	(.326)	(2.503)	(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: non-agricultural worker or employee. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number of children < 6, wealth, health insurance status, age squared and calender year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Relationship between self-reported diabetes and selection into types of work.

	Males			Females		
	(1) Non-agric.	(2) Agric.	(3) Self-employed	(4) Non-agric.	(5) Agric.	(6) Self-employed
Self-reported diabetes	−.006	−.008	−.043	−.001	$-.022^{**}$	−.029
	(.029)	(.022)	(.026)	(.018)	(.009)	(.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. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number of children < 6, wealth, health insurance status, age squared and calender year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Labor outcomes and time since diagnosis

Because diabetes is a chronic and generally life-long disease, we investigate how soon after the first diagnosis diabetes affects labor outcomes. Given that complications of diabetes develop over time, the effect may increase linearly as the years go by. Non-linear relationships are also plausible: health problems that have led to the diagnosis, as well as psychological effects after the diagnosis may affect labor outcomes immediately after having been diagnosed with diabetes. Similarly, management of the disease may be successful only after some initial period. It is also possible that complications only appear after some time, reducing health, labor supply and productivity years after the initial diagnosis.

To obtain an initial idea of the relationship between our outcome variables and diabetes duration we use a non-parametric kernel-weighted local polynomial regression. As Figure 1 shows, 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.[15] A similar analysis for wages (Figure 1) shows a somewhat more erratic relationships, with a possibly long term negative trend for women but not for men. A similar trend is observed for working hours (Figure 1).

The results in Table 5 panel A for estimating Eq. 2, show that male employment probabilities fall every year in all models, 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 of Table 5 shows the estimates when using a spline function as 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 after which it becomes stronger again. However, standard errors are quite large.

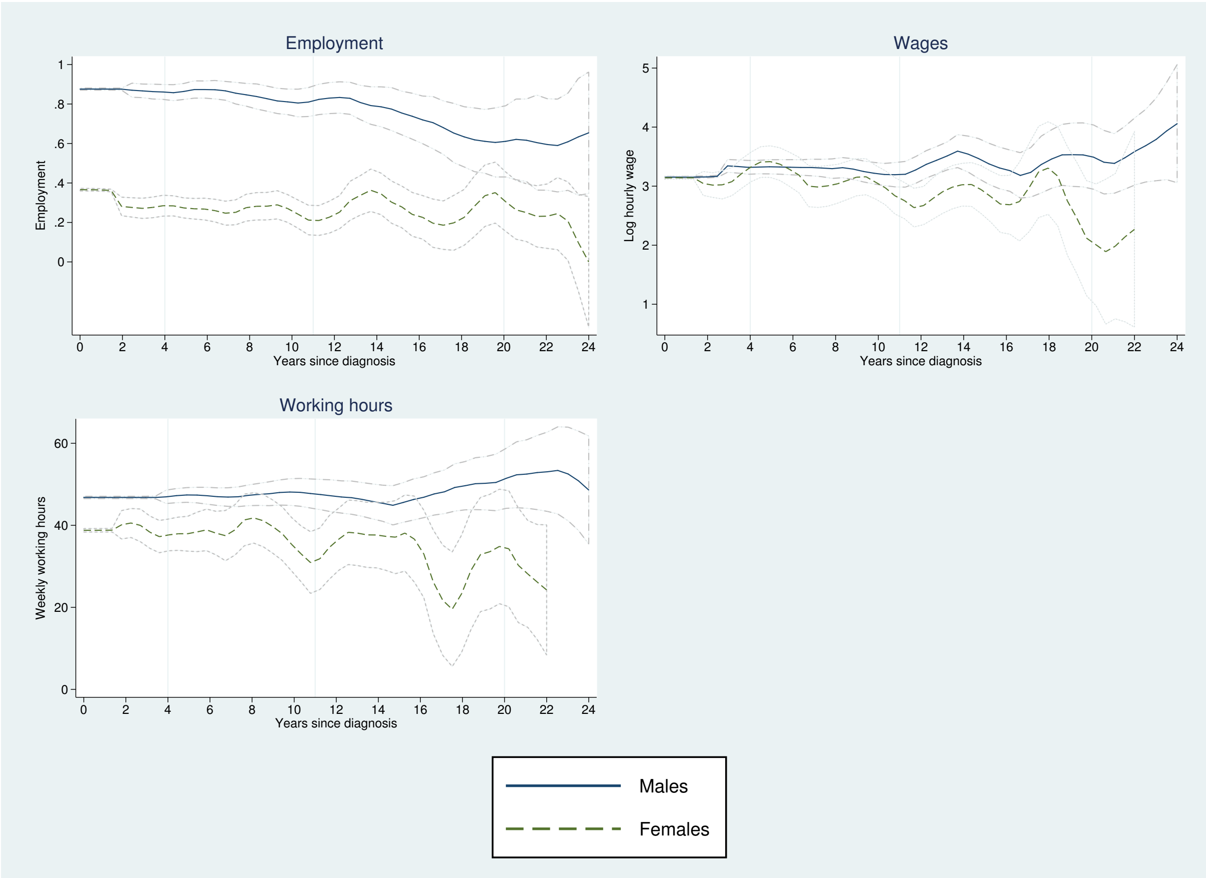
The results in Table 6, indicate a reduction in female wages of about 7% per year after diagnosis in the FE model. For men we find no consistent effect. The results of the non-linear specification in panel B of Table 6 indicate that there may be a reduction in wages 5–11 years after the initial diagnosis 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.[16] Interestingly, the reductions in wages found in the non-linear specification appear exactly at the time where employment probabilities are less affected. This could suggest that at this time reductions in productivity affect wages but are not so severe that they would cause job loss.

The estimation results in Table 7 indicate that there appears to be no consistent relationship between working hours and time since being diagnosed with diabetes, neither for men nor for women.

[15] 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.

[16] There are only 9 and 3 observations for male and female wages with more than 20 years since diagnosis in wave 3, respectively, and 17 and 7 in the pooled sample, respectively. For male and female working hours there are 12 and 7 observations with more than 20 years since diagnosis in wave 3, respectively, and 20 and 12 for the pooled sample, respectively.

Figure 1: Kernel-weighted local polynomial regression of employment, wages and working hours on diabetes duration.



Notes The dashed lines show 95% confidence interval.

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 particular in comparison to our results for Mexico. However, the non-linear results we presented are not directly comparable to these estimates as Minor used pooled cross-sectional data, constructed dummy variables to indicate time since diagnosis instead of splines and used different duration groups.[17]

[17]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 5: 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
Panel A: linear						
Diabetes duration	$-.008^{***}$	$-.007^{***}$	$-.017^{***}$	$-.005^{***}$	$-.004^{***}$	$-.009^{*}$
	(.002)	(.002)	(.006)	(.002)	(.001)	(.005)
Hausman test			153.024			200.073
p-value			0.000			0.000
Panel B: splines						
Diabetes duration						
0–4	$-.007$	$-.007$	$-.026^{*}$	$-.010$	$-.015^{**}$	$-.017$
	(.007)	(.006)	(.014)	(.007)	(.006)	(.016)
5–11	0.000	$-.003$	$-.003$	$-.004$	0.004	$-.003$
	(.009)	(.006)	(.009)	(.008)	(.006)	(.008)
12–20	$-.030^{**}$	$-.017^{*}$	$-.029^{*}$	0.005	$-.004$	$-.014$
	(.012)	(.010)	(.016)	(.008)	(.006)	(.011)
> 20	0.011	0.007	$-.046^{*}$	$-.010^{*}$	$-.003$	$-.015$
	(.016)	(.014)	(.028)	(.006)	(.003)	(.018)
Hausman test			161.953			198.692
p-value			0.000			0.000
N	8217	16292	16292	10467	22407	22407

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. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6, wealth, age squared and calendar year dummies. The OLS and pooled OLS models additionally control for age. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Relationship between self-reported years since diagnosis and log hourly wage 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
Panel A: linear						
Diabetes duration	0.001	0.010**	−.019	−.014*	−.009	−.073**
	(.006)	(.005)	(.018)	(.008)	(.008)	(.029)
Hausman test			838.213			93.232
p-value			0.000			0.000
Panel B: splines						
Diabetes duration						
0–4	0.034*	0.046***	0.033	0.027	0.030	0.015
	(.017)	(.016)	(.055)	(.031)	(.026)	(.138)
5–11	−.041*	−.037**	−.055*	−.039	−.034	−.101*
	(.021)	(.018)	(.033)	(.030)	(.024)	(.056)
12–20	0.015	0.044	0.062	−.032	−.071*	−.051
	(.033)	(.029)	(.056)	(.042)	(.039)	(.047)
> 20	0.053	0.014	−.111	−.007	0.041***	−.204***
	(.054)	(.040)	(.104)	(.028)	(.015)	(.053)
Hausman test			1037.290			96.266
p-value			0.000			0.000
N	5509	10767	10767	2874	5741	5741

Notes The table presents the results of three estimation methods for log hourly wages. Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications. Robust standard errors in parentheses. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6, wealth, age squared, calendar year dummies, type of work (agricultural and self employed with dependent non-agricultural wage employment as the base) and health insurance status. The OLS and pooled OLS models additionally control for age. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Relationship between self-reported years since diagnosis and weekly working hours 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
Panel A: linear						
Diabetes duration	0.069	0.048	0.181	−.020	−.124	0.208
	(.124)	(.102)	(.330)	(.187)	(.127)	(.652)
Hausman test			704.904			107.709
p-value			0.000			0.000
Panel B: splines						
Diabetes duration						
0–4	−.033	−.233	0.709	0.739	0.470	2.014
	(.421)	(.325)	(.938)	(.645)	(.586)	(2.947)
5–11	0.269	0.338	−.218	−.410	−.479	−.508
	(.539)	(.399)	(.568)	(.728)	(.553)	(1.020)
12–20	0.209	0.137	0.698	−.164	−.051	−.402
	(.730)	(.538)	(.945)	(.995)	(.700)	(1.207)
> 20	−1.300	−.768	0.039	−.499	−.418	8.117***
	(.944)	(.930)	(2.184)	(.930)	(.305)	(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 for weekly working hours. Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications. Robust standard errors in parentheses. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6, wealth, age squared, calendar year dummies, type of work (agricultural and self employed with dependent non-agricultural wage employment as the base) and health insurance status. 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 8 presents a cross tabulation of self-reported diabetes and biomarker results. The table shows that 27% of the sample have HbA1c levels indicative of diabetes and 81% of those self-reporting a diabetes diagnosis also have HbA1c levels equal to or above the diabetes threshold. The cell percentages in the table underline the large proportion of false negatives (18%). The 3% with self-reported diabetes but biomarker levels below the diabetes threshold could be interpreted as false positives. However, many of these likely have received an actual diabetes diagnosis but achieved blood glucose levels below the diabetes threshold due to successfully managing the disease. Of the respondents who test positive for diabetes according to the biomarker analysis, 32% self-report a diagnosis, while 68% do not. There are no considerable differences to the presented proportions in Table 8 if we look at men and women separately and we therefore do not show these results here.

Table 8: 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 %	71%	18%	z
Self-reported diabetes (N)	129	554	683
Row %	19%	81%	100%
Column %	3%	32%	11%
Cell %	2%	9%	z
Total (N)	4673	1735	6408
Row %	73%	27%	100%
Column %	100%	100%	100%
Cell %	x	y	z

Table 9 presents the results from estimating Eq. 4, Eq. 5 and Eq. 6. The results in columns 1 and 2 of Table 9 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.[18] Results in columns 5 and 6, obtained from estimating

[18] 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 marginally different to those presented in columns 3 and 4 of Table 8, which is why we do not present them here.

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 9: Biomarker results

	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Dependent variable: Employment						
Self-reported diabetes	$-.051^{**}$	$-.044^{*}$			$-.053^{**}$	$-.032$
	(.026)	(.023)			(.026)	(.026)
HbA1c $\geq$ 6.5			$-.012$	$-.031^{*}$	0.003	$-.022$
			(.016)	(.018)	(.017)	(.019)
N	2785	3623	2785	3623	2785	3623
Dependent variable: Log hourly wages						
Self-reported diabetes	$-.010$	$-.040$			$-.006$	$-.010$
	(.065)	(.113)			(.078)	(.119)
HbA1c $\geq$ 6.5			$-.007$	$-.057$	$-.006$	$-.055$
			(.044)	(.070)	(.049)	(.075)
N	1803	884	1803	884	1803	884
Dependent variable: Weekly working hours						
Self-reported diabetes	$-.293$	$-.751$			$-.286$	-1.566
	(1.305)	(2.178)			(1.419)	(2.351)
HbA1c $\geq$ 6.5			$-.088$	1.153	$-.012$	1.525
			(.844)	(1.462)	(.925)	(1.565)
N	2302	1144	2302	1144	2302	1144

Notes Community level fixed effects. Robust standard errors in parentheses. Other control variables: age, age squared, state dummies, urbanization dummies, education dummies, married dummy, number children < 6 and wealth. Calender year dummies are included as data collection for the third wave was stretched out over several years. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural wage employment as the base) and for health insurance status. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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.[19] We then extend the model to take into account the severity of diabetes more 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,

[19]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 ($p < 0.05$).

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 10, 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 10, 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.[20]

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 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 et al. (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.

[20]Further analysis indicates that the reductions in the diabetes coefficient for men appear after the inclusion of heart disease as well as hypertension, while ovwerweight and obesity play a minor role.

Table 10: 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
Panel A (HbA1c levels)						
Self-reported diabetes	−.057*	−.027	−.004	−.009	−.101	−1.607
	(.031)	(.025)	(.069)	(.113)	(1.370)	(2.363)
HbA1c (if HbA1c $\geq$ 6.5%)	0.001	−.003	−.001	−.005	−.030	0.151
	(.002)	(.002)	(.005)	(.008)	(.103)	(.172)
N	2785	3623	1803	884	2302	1144
Panel C (self-reported health)						
Self-reported diabetes	−.036	−.023	0.002	0.060	0.123	−2.191
	(.026)	(.027)	(.079)	(.121)	(1.433)	(2.386)
Hba1c $\geq$ 6.5%	0.003	−.023	−.004	−.051	−.066	1.829
	(.017)	(.019)	(.049)	(.075)	(.926)	(1.569)
Self-reported health status						
good	0.023	0.057*	0.061	−.115	−1.131	3.521
	(.025)	(.034)	(.074)	(.124)	(1.376)	(2.499)
fair	−.007	0.006	0.025	−.157	−1.606	4.646*
	(.026)	(.034)	(.076)	(.128)	(1.424)	(2.607)
bad	−.127***	−.024	−.016	−.371*	−6.190**	6.918*
	(.043)	(.046)	(.135)	(.189)	(2.521)	(3.858)
very bad	−.165	0.117	−.331	0.316	−1.869	−17.400*
	(.110)	(.116)	(.300)	(.439)	(6.433)	(9.005)
N	2785	3621	1803	883	2302	1143
Panel B (other chronic diseases)						
Self-reported diabetes=1	−.044*	−.029	0.005	0.053	−.131	−1.411
	(.026)	(.027)	(.079)	(.120)	(1.438)	(2.385)
HbA1c $\geq$ 6.5%=1	0.001	−.021	−.010	−.037	−.077	1.382
	(.017)	(.019)	(.049)	(.076)	(.927)	(1.572)
Overweight	0.021	−.003	0.093**	0.034	−.225	−1.520
	(.016)	(.020)	(.046)	(.072)	(.874)	(1.528)
Obese	0.022	−.030	0.086	−.026	0.895	−.385
	(.019)	(.020)	(.053)	(.075)	(1.003)	(1.578)
Hypertension	−.035	−.020	−.106	−.089	−.447	−1.828
	(.025)	(.022)	(.072)	(.089)	(1.370)	(1.854)
Heart disease	−.165***	−.045	0.060	−.039	−2.640	−5.430
	(.059)	(.050)	(.179)	(.215)	(3.659)	(4.854)
N	2785	3621	1803	883	2302	1143

Notes Community level fixed effects. Robust standard errors in parentheses. Other control variables: age, age squared, state dummies, urbanization dummies, education dummies, married dummy, number children < 6 and wealth. Calender year dummies are included as data collection for the third wave was stretched out over several years. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural wage employment as the base) and for health insurance status. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 et al., 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 heath status and later life incidence of diabetes and other chronic diseases (Hanson et al., 2012; 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 11. 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 11) 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 11). 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 11), given that most diabetes self-reports tend to be correct.

Table 11: 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 11) and 49 with one self-report of diabetes (from lines 1 and 2 in Table 11)). 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 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.

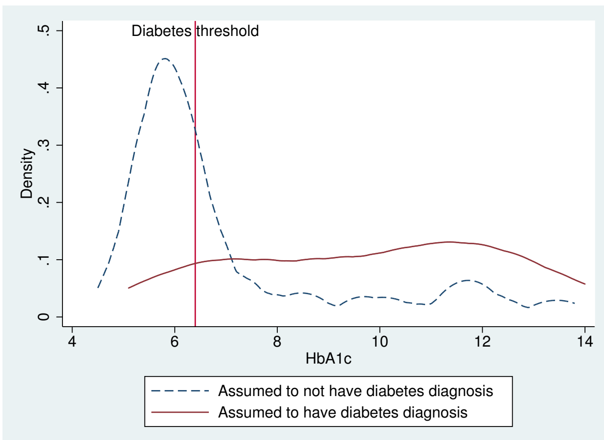


Table 12: Studies estimating the relationship between diabetes and employment

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 et al. (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)
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)

Table 12: 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	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)
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 12: 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 and HbA1c levels for subsample	No statistically significant relationship between undiagnosed diabetes and the probability of employment. Self-reported diabetes significantly related with lower employment probabilities for men (-11 % points) and women (-19 % points). Using only biomarker information (HbA1c >6.4 %), statistically significant reductions in employment probabilities for men (-8.3 % points) and women (-11 % points). No significant effects of undiagnosed diabetes on hours worked. Increase in HbA1c by 1 % point related to 1.3 % points lower employment probabilities for men. No effect for women.	Exogeneity assumed	Probit model for binary outcomes, OLS for continuous outcomes; all applied to pooled data	Minor et al. (2016)
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 12: 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

Adler, A. I., Stevens, R. J., Manley, S. E., Bilous, R. W., Cull, C. A., and Holman, R. R. (2003). "Development and progression of nephropathy in type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS 64)." *Kidney International* 63 (1), 225–232.

Alegre-Díaz, J. et al. (2016). "Diabetes and Cause-Specific Mortality in Mexico City." *New England Journal of Medicine* 375 (20), 1961–1971.

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.

Baird, S., Gong, E., McIntosh, C., and Özler, B. (2014). "The heterogeneous effects of HIV testing." *Journal of Health Economics* 37, 98–112.

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.

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.

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.

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.

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.

Bellamy, L., Casas, J.-P., Hingorani, A. D., and Williams, D. (2009). "Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis." *The Lancet* 373 (9677), 1773–1779.

Bergemann, A., Grönqvist, E., and Gudbjörnsdottir, 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.

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.

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.

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.

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.

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.

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.

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.

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.

Gregg, E. W. et al. (2012). "Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes." *Journal of the American Medical Association* 308 (23), 2489.

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.

Hanson, M. A., Gluckman, P. D., Ma, R. C., Matzen, P., and Biesma, R. G. (2012). "Early life opportunities for prevention of diabetes in low and middle income countries." *BMC Public Health* 12 (1), 1025.

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.

Hu, F. B. (2011). "Globalization of Diabetes: The role of diet, lifestyle, and genes." *Diabetes Care* 34 (6), 1249–1257.

International Diabetes Federation (2014). *Diabetes Atlas 2014 Update*. 6th ed. International Diabetes Federation.

— (2015). *IDF Diabetes Atlas*. 7th ed.

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.

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.

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.

Liu, X. and Zhu, C. (2014). "Will knowing diabetes affect labor income? Evidence from a natural experiment." *Economics Letters* 124 (1), 74–78.

Meza, R., Barrientos-Gutierrez, T., Rojas-Martinez, R., Reynoso-Noverón, N., Palacio-Mejia, L. S., Lazcano-Ponce, E., and Hernández-Ávila, M. (2015). "Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates." *Preventive Medicine*.

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

— (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.

Minor, T. and MacEwan, J. P. (2016). "A comparison of diagnosed and undiagnosed diabetes patients and labor supply." *Economics & Human Biology* 20, 14–25.

Mundlak, Y. (1978). "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46 (1), 69–85.

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.

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.

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.

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.

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.

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.

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.

Schneiderman, N., Llabre, M., et al. (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.

Seuring, T., Archangelidi, O., and Suhrcke, M. (2015). "The Economic Costs of Type 2 Diabetes: A Global Systematic Review." *PharmacoEconomics* 33 (8), 811–831.

Seuring, T., Goryakin, Y., and Suhrcke, M. (2015). "The impact of diabetes on employment in Mexico." *Economics & Human Biology* 18, 85–100.

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.

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.

Thornton, R. L. (2008). "The Demand for, and Impact of, Learning HIV Status." *American Economic Review* 98 (5), 1829–1863.

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.

World Health Organization. *Global database on body mass index*.

— (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.

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.

Zhang, X., Zhao, X., and Harris, A. (2009). “Chronic Diseases and Labour Force Participation in Australia.” *Journal of Health Economics* 28 (1), 91–108.

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.