Preliminary Draft, Please do not cite or circulate without authors' permission

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

Till Seuring*, Pieter Serneels*, and Marc Suhrcke**

**University of East Anglia*
***University of York*

May 9, 2016

Abstract

Diabetes is increasingly recognized as a major health risk in high income as well as middle income countries. While adverse economic effects of diabetes are highly plausible, the existing empirical evidence is limited. This paper investigates the impact of diabetes on labor market outcomes focusing on Mexico, a country with high and rising diabetes rates. Two challenges present themselves when using survey data to examine this relationship: (1) causality is hard to identify, and (2) diabetes is typically self-reported, potentially causing biased estimates. The paper makes headway on both fronts using rich panel and biomarker data. Applying fixed effects (FE) estimation, we account for time-invariant omitted variables, providing an improved identification strategy. The results indicate a negative effect of self-reported diabetes on the probability of employment of around 5 percentage points (p.p.), which becomes more negative over time. By contrast we find no consistent evidence for an impact on wages or hours worked. Relying on cross section biomarker data to identify people with diabetes also suggests a negative, albeit smaller, relationship with the probability of employment. Accounting for self-reported and undiagnosed diabetes separately, an adverse association is only found for the former. Hence, estimates based on self-reported diabetes may overstate the employment effect of diabetes. Nonetheless, the effects found for self-reported diabetes account for an overall substantive economic burden, given the size of the population that has self-reported the condition.

1 Introduction

Diabetes, and particularly the most common Type II Diabetes mellitus (T2DM), has increased worldwide and is expected to continue to rise over the next decades (NCD Risk Factor Collaboration 2016). It has become a problem for middle-income countries (MICs) and high-income countries (HICs) 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 living in the United States of America (USA) (Schneiderman, Llabre, et al. 2014). In Mexico itself diabetes prevalence has risen from 6.7% in 1994 to 14.4% in 2006, including both diagnosed and undiagnosed cases (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.

This increase in prevalence stems both from 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 also seems to play a role (Williams et al. 2014). Recent evidence indicates that the onset of diabetes increasingly happens at an 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 onset will increase the likelihood of complications during the productive lifespan.

Diabetes is a term used to describe various diseases characterized by high blood glucose values, with the predominant disease being T2DM, which we will henceforth refer to when using the term diabetes. The elevated blood glucose levelsthat 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, the individual has the opportunity to prevent most of not all complications. If not or inadequately treated, diabetes has been documented to lead to conditions such as heart disease and stroke, blindness, kidney problems, and nerve problems that together with impaired wound healing can lead to the loss of limbs (Reynoso-Noverón et al. 2011). These conditions can be seriously debilitating and may therefore reduce an individual's economic activity, including his productivity and labor market participation. Those who are able to "reverse" their diabetes — meaning they manage to return to healthy blood glucose levels as a result of lifestyle changes and medication that re-establish normal insulin sensitivity — are unlikely to suffer from any diabetes related health problems (Gregg et al. 2012; Lim et al. 2011).

The effect of diabetes on labor market outcomes has been studied predominantly in high-income country (HIC) —with the exception of one study on Mexico (Seuring, Goryakin, et al. 2015) and China (Liu et al. 2014) each— and have found diabetes to be associated with reductions in employment probabilities as well as wages and labor supply (H. S. Brown, Pagan, et al. 2005; H. S. Brown, Perez, et al. 2011; T. T. Brown 2014; Latif 2009; Minor 2011, 2013; Minor and MacEwan 2016).[1]

However, while these studies have provided good evidence of the potential labor market effects of diabetes, many of the complexities of the relationship have not been comprehensively addressed in a single study. First of all, unobserved heterogeneity presents a challenge to estimate the relationship between diabetes and labor outcomes. Especially time-invariant unobserved individual characteristics, e.g. health endowments — often re-

[1] A recent systematic overview also underlines the potential effects on labor market outcomes, as well as on individual healthcare expenditures (Seuring, Archangelidi, et al. 2015).

lated to health during uteru, infant and child years and low household income or adverse health shocks during these early years — as well as risk preferences have been shown to adversely affect health in general and the propensity to develop type 2 diabetes more specifically (Ewijk 2011; Li et al. 2010; Sotomayor 2013). These and other unobserved personal characteristics (like ability) may also affect employment probabilities, wages or working hours directly through their effects on today's productivity (Currie et al. 2013) and indirectly by limiting educational attainment and human capital accumulation (Ayyagari et al. 2011). Further, only focusing on the overall effect of a self-reported diabetes diagnosis does not reveal when potential labor market penalties appear, given the dynamic aspect of diabetes and the differences in its potential health effects over time. Additionally, apart from its health effects diabetes might also affect labor market outcomes through other channels. For instance, people aware of their condition may be less inclined to continue working if this interferes with their disease management; they may also use the diagnosis as a justification for decreasing their labor supply, leading to a potential justification bias (Kapteyn et al. 2009) in the estimated effect of diabetes. Importantly, for these reasons the labor market effects may also be distinct for people with self-reported versus those with undiagnosed diabetes, causing ambiguously biased estimates if the analysis is solely based on self-reported diabetes as measurement error is non-classical (Cawley et al. 2015).

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes, while improving upon previous work by paying close attention to the above challenges. We use three waves of panel data from Mexico provided by the Mexican Family Life Survey (MxFLS) covering the period 2002–2012. The MxFLS is particularly useful for the analysis of diabetes as it allows us to account for the above complexities in a more refined way than has been the case so far. Using individual level FE analysis for the first time in this literature, we take account of time-invariant heterogeneity when assessing the impact of self-reported diabetes and self-reported diabetes duration on labor market outcomes.[2] Further, we add to the current literature in exploring the role of undiagnosed diabetes, using novel and rich biomarker data - an issue of considerable importance in light of the large prevalence of undiagnosed diabetes (see Beagley et al. (2014)) that remained unaccounted for in earlier studies which rely on self-reported information. Doing so sheds light on the issue of measurement error and the potentially differential effects of self-reported and undiagnosed diabetes.

Our results using self-reported diabetes suggest an economically important decrease in employment probabilities for people aware of their disease. Wages and working hours, however, do not appear to be associated with self-reported diabetes. We further find that employment probabilities are reduced 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 measurement error of self-reported diabetes leads to an upward biased estimate of the employment penalty, compared to biometrically measured diabetes. Overall, undiagnosed diabetes does not appear to affect any labor market outcome, suggesting that adverse effects mainly occur to those with a diagnosis. We argue that ,nonetheless, the effects found for self-reported diabetes in this study are largely

[2]We are not aware of any other evidence on the effect on wages and working hours in a middle-income country (MIC).

unbiased as long as inference is not extended to the unobserved undiagnosed population, and are economically important given the considerable size of the diagnosed population in Mexico.

2 Diabetes and labor outcomes – existing evidence

Several studies have investigated the effects of diabetes on labor market outcomes.

For the USA, H. S. Brown, Pagan, et al. (2005) estimate the impact on employment in 1996–1997 in an elderly population of Mexican Americans, living close to the Mexican border, using a bivariate probit model. They find diabetes to be endogenous for women but not for men. For the latter, the estimates show a significant adverse effect of 7 p.p.. For women, the negative effect becomes insignificant when using instrumental variable (IV) estimation. In another study, again for a Mexican-American sample, H. S. Brown, Perez, et al. (2011) look at how diabetes management, inferred from measured glycated hemoglobin (HbA1c) levels, is associated with employment chances and wages using cross-sectional USA data. They find a linear negative association between HbA1c levels and both employment chances and wages for men. Two further studies also examine the impact of diabetes on employment and productivity for the USA: Minor (2011) focuses on the effect of diabetes on female employment, earnings, working hours and lost work days in 2006, finding diabetes to be endogenous and its effect underestimated if exogeneity is assumed. In the IV estimates, diabetes has a significant negative effect on female employment as well as annual earnings but not on working hours. Both studies use a Heckman selection model to adjust for a possible selection bias, though, neither discusses the validity of the underlying exclusion restriction. In a later study Minor (2013) investigates the relationship of diabetes duration and labor market outcomes using a cross-sectional analysis, providing evidence of a non-linear relationship, with employment probabilities declining shortly after diagnosis for men and after about 10 years for women; wages are not affected by duration. Finally, a recent study by (Minor and MacEwan 2016) investigates the association of self-reported diabetes and undiagnosed diabetes with employment probabilities and working hours in an adult USA population, using cross-sectional data. They find a reduction in the coefficient size of diabetes if undiagnosed diabetes cases are included in the diabetes indicator instead of only self-reported diabetes. Further, they find that there is no association of undiagnosed diabetes with employment probabilities itself. However, the results of the study, particularly those for undiagnosed diabetes, are based on a very small number of cases, warranting further investigation.

For Canada, Latif (2009) estimate the effect of the disease on employment probabilities using an IV strategy similar to H. S. Brown, Pagan, et al. (2005). He finds diabetes to be exogenous for females, and both endogenous and overestimated for males in the univariate model, with the estimates of the bivariate model indicating a significant negative impact on the employment probabilities for women, but not for men. For Australia, Zhang et al. 2009 analyze the effects of diabetes on labor force participation using a multivariate endogeneous probit model. They find reduced labor market participation for males and females, and these effects are overstated if the endogeneity of diabetes is unaccounted for.

Only two studies exist for MICs by our knowledge. Liu et al. (2014) investigate the effect of diabetes diagnosis on labor income in China, exploiting a natural experiment to identify causality and find a significant reduction in income for those with a recent diagnosis. An earlier study for Mexico investigated the effect of self-reported diabetes on the probability of employment using only cross-sectional data from the 2005 wave of the MxFLS, and found a significant ($p<0.01$) reduction in employment chances for males by about 10 p.p. and for females by about 4.5 ($p<0.1$) p.p., using parental diabetes as an IV (Seuring, Goryakin, et al. 2015). The scarcity of evidence for low- and middle-income countries (LMICs) is also underlined in a recent systematic review of the economic cost of diabetes (Seuring, Archangelidi, et al. 2015).

Overall, the majority of existing studies, including those on high income countries, tend to suffer from at least three key limitations:

1. They rely exclusively on cross-sectional data, unable to account for unobserved characteristics.

2. The use of the family history of diabetes, which has been the sole instrumental variable employed so far, relies on the genetic and heritable component of T2DM that could theoretically provide valid identification of the true effect of diabetes. However, it remains unclear whether the variable fully satisfies the exclusion restriction, as it may also proxy for other genetically transferred traits, including those that impact labor outcomes directly, as for instance unobserved abilities. This traditional identification strategy also abstracts from intrahousehold or intergenerational labor supply effects (Seuring, Goryakin, et al. 2015).[3]

3. The use of self-reported diabetes can introduce non-classical measurement error due to systematic misreporting which 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).

As mentioned before, to overcome some of these limitations, this paper applies an individual level FE panel estimation strategy and biomarker data. We also dive more deeply into estimating more detailed models according to the type of employment, i.e. non-agricultural employment, agricultural employment and self-employment, as ill health may have distinct effects across these activities.

3 Data

We use the Mexican Family Life Survey (MxFLS), a nationally representative, longitudinal household survey, which has 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 of the individuals and their families (Rubalcava et al. 2013). Apart from self-reported diabetes information that is available in all rounds, we also use information on the self-reported year of diagnosis as

[3]It is conceivable that diabetes might deteriorate parental health in such a way that the offspring either has to give up their employment to provide care, or has to increase labor supply to compensate for lost income.

well as biomarker data including HbA1c levels for a subsample of respondents. Our main analysis uses all three waves taking advantage of the large amount of observations and the panel structure of the data. Our variable of interest is self-reported diabetes, which is based on the survey question: "Have you ever been diagnosed with diabetes?".

Because the response to this question may well suffer from measurement error due to recall bias, we investigate and try to increase the consistency of the self-reported diabetes variable, using disease information from earlier and ensuing waves to infer on the current, missing or inconsistent, diabetes status (see Appendix A for further details on our correction procedures.) A further, and no less important, source of measurement error is undiagnosed diabetes. In order to investigate how this may affect estimates of the labor market impact of diabetes we use information from a subsample containing over 6000 respondents (everybody aged 45+ and a random subsample of those aged 15–44 (Crimmins et al. 2015)) of the 2009-2012 wave, and use biometrically measured diabetes to identify those with undiagnosed diabetes.

For the entire paper the samples we use are restricted to the working age population (15–64). To prevent pregnant women from biasing our results due to the increased diabetes risk during pregnancy and its effects on female employment status, we have dropped all observations of women reporting to be pregnant at the time of the survey (N=764). We further exclude everybody currently in school.

The detailed information in the MxFLS allows us to construct the following outcome variables of interest: employment[4], hourly wage and weekly working hours[5] For the pooled data of all three waves (Table 1), diabetes was self-reported by 5% of men and 6% of women. Taking into account the different age groups, this is consistent with other prevalence estimates of self-reported diabetes for this time period in Mexico.[6] Most of the respondents in the sample either live in rural or in large urbanized areas. Looking at our outcome variables, 87% 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 employment. Women

[4]Employment status is defined as having worked or carried out an activity that helped with the household expenses the last week and usually working for at least four hours per week. This explicitly includes those employed informally, for instance people working in a family business.

[5]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. This gave us the average earnings per week which was then divided by the weekly working hours to arrive at an hourly wage estimate. Labor income was either reported as the total amount for the whole month or more detailed containing information on the monthly wage, income from piecework, tips, extra hours, meals, housing, transport, medical benefits and other earnings. Over 80% of respondents reported the total amount instead of a detailed amount. Respondents were also asked for their annual income and we used that information to arrive at an hourly wage if information for monthly labor income was missing. Finally, we adjusted the calculated wage for inflation from the year of the interview up to 2013 and took the log of those values. 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 were calculated summing up the self-reported usual working hours of the first and —if applicable— the second job.

[6]Barquera, Campos-Nonato, et al. (2013) show that the prevalence of diagnosed diabetes in Mexico was 7.5% in 2006, only somewhat above our results, which may be the result of the slightly different age groups considered.

Table 1: Descriptive statistics for panel sample and biomarker sample

	Panel		Biomarker	
	Males	Females	Males	Females
Dependent variables				
Employed	0.87	0.37	0.86	0.34
	(0.34)	(0.48)	(0.35)	(0.47)
Hourly wage (Mexican Peso)	42.47	40.45	36.28	35.31
	(485.56)	(141.95)	(53.66)	(43.68)
Usual weekly workinghours	46.82	39.00	45.98	38.15
	(16.79)	(18.90)	(16.88)	(19.64)
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 (years)	0.34	0.42	0.67	0.88
	(2.01)	(2.39)	(2.81)	(3.42)
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	36.28	42.78	42.80
	(13.62)	(13.17)	(14.28)	(13.94)
Rural village of <2,500	0.44	0.43	0.51	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 (<6) in household	1.48	1.57	1.19	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	21413	27373	2790	3626

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

also have lower educational attainment on average.

Turning to the biomarker subsample of the third wave (2009-2012), respondents are somewhat older on average than in the pooled sample, as it includes everybody above the age of 44 but only a random subsample of those aged 44 or below (Crimmins et al. 2015). Also, self-reported diabetes is higher than in the pooled sample as well as in the full sample of wave three. Regarding the other control and outcome variables, the sample is fairly similar to the pooled sample. Remarkably a large share of people have an HbA1c indicative of diabetes, defined by the World Health Organization (WHO) as levels above or equal 6.5% (World Health Organization 2011)[7]: 18% of males and females have undiagnosed diabetes. This suggests that relying on self-reported diabetes as a measure for diabetes in Mexico might considerably underestimate the true extent of diabetes, potentially leading to biased estimates of its economic impact.

4 Estimation strategy

Strauss et al. (1998) provide a useful framework to think about the relationship between health and labor market outcomes:

$$L = L(H, pc, w(H; S, A, B, I, \alpha, e_w), S, A, B, V, \xi) \tag{1}$$

where L is labor supply or labor market participation, pc is a vector of prices for consumer goods, w is the real wage; H is an array of measured health status ; S is education; A is a vector of demographic characteristics; B is the family background of the individual; Icaptures the local community infrastructure; α is an array of unobservables (e.g. ability), e_w represents the measurement error, V is non-labor income and ξ is the taste parameter.

The equation showcases the joint effect of health on both wages and labor supply or labor market participation. Health affects labor supply and participation directly by impacting the ability to work and indirectly by changing wages.

There are several ways diabetes may affect H. First of all, diabetes can deteriorate health if it remains untreated, with the adverse effects potentially increasing over time. Second, a diagnosis of diabetes and ensuing treatment may lead to better health compared to the undiagnosed state. However, compared to healthy people even those receiving treatment for their diabetes may still have worse health outcomes. Third, there is also evidence that the diagnosis itself may affect one's own health perception and could lead to worse self-perceived health (Thoolen et al. 2006). We therefore expect diabetes to adversely affect health and consequently labor market outcomes.

When estimating equation 1 empirically with observational data, unobserved heterogeneity may bias the results. As mentioned in section 1 unobserved factors captured in α such as early childhood investments, innate ability and time preference could affect wages as well as the probability to develop diabetes. Further, changes in lifestyle due to changes in wages or employment status may also affect the probability to develop diabetes through changes in diet and physical activity. Finally, measurement error e_w may be an

[7]In one of the first analyzes of these new biomarker data Frankenberg et al. (2015) show that the rates in Mexico of elevated HbA1c levels are very high when compared to HbA1c data from similar surveys in the USA and China. (Frankenberg et al. 2015).

important issue due to the large undiagnosed population with diabetes, particularly if being diagnosed is related to employment or wages via better access to healthcare through employment benefits and higher income.

The following section describes our estimation strategy for the different parts of the data.

4.1 Panel data on self-reported diabetes

We investigate the relationship of self-reported diabetes and three labor market outcomes: employment, wages and labour supply, respectively, using a FE model. While using individual level FE does not allow to fully identify a causal relationship, this strategy does improve on the degree of causal inference, compared to a simple cross-sectional analysis.[8] In particular it does allow controlling for unobserved personal characteristics that could bias the estimates, without the drawbacks of a less than convincing IV strategy that has been widely applied in this literature. We have also estimated random effects (RE) models but do not present them here as the Hausman test suggested the use of the FE model throughout.[9]

We estimate the following model:

$$Y_{it} = \beta_0 + \beta_1 Diabetes_{it} + \beta_2 X_{it} + c_i + \gamma_t + u_{it}. \tag{2}$$

where Y_{it} is a binary variable taking a value of 1 if respondent i reports being in employment at time t and 0 otherwise, $Diabetes_{it}$ is a binary variable taking a value of 1 at time t if the respondent reports having ever received a diagnosis of diabetes[10], X_{it} is a vector of control variables, c_i represents an individual fixed effect, γ_t represents a 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 having positive wages and being employed, respectively. In these models Y_{it} represents the log hourly wage of respondent i at time t or the usual weekly working hours over the last year.

The control variables in both FE specifications include dummy variables for the effects of any changes in the living environment, of living in a small, medium or large city with rural as the reference category, and state dummies capturing the effects of moving to a different state. We also include a marital status dummy to control for the impact of marriage on the probability of being employed. A variable capturing the number of children residing in the household below the age of 6 is used to control for the impact of children on labor market 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

[8] Other forms of unobserved heterogeneity could also affect our estimates — for instance time-variant unobserved heterogeneity or omitted variables simultaneously driving labor outcomes and health

[9] Results are available on request.

[10] We are not able to distinguish between type 1 diabetes and type 2 diabetes using this data, but given information from other studies that have tried to do this and found no association with labor market outcomes (Minor 2011; Minor and MacEwan 2016), we think including those with type 1 diabetes likely attenuates any adverse relationship we may find.

and housing conditions to create an indicator for household wealth (Filmer et al. 2001).[11]. Finally, a quadratic age term and calendar year dummies are included to capture the non-linear effect of age and any trends over time, respectively.

Before moving on, it bears emphasizing that despite our efforts to reduce any bias in our estimates, the estimated coefficients do not reflect true causal effects since time-variant unobserved heterogeneity may still bias the estimates. With respect to employment status, one example would be that job loss affects lifestyle choices that increase the probability to develop diabetes, which could then in turn negatively affect labor market outcomes. So far, no strong adverse effects of being laid off on diabetes self-reports have been found in the literature (Bergemann et al. 2011; Schaller et al. 2015), although this has only been researched in a high-income country context. Another example relates to stress at work, which has been linked to the development of 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 the role of these time variant unobserved factors, it seems that the role of time-invariant variables, e.g. genetic predisposition and relatively stable personality traits, is predominant. The applied FE approach should then limit the bias due to these time-invariant confounding factors.

4.2 Self-reported diabetes duration

To explore the role of the duration of diabetes for labor outcomes, we estimate the following model using a self-reported measure of the years since diagnosis:

$$Y_{it} = \beta_0 + \beta_1 Dyears_{it} + \beta_2 X_{it} + c_i + u_{it}, \tag{3}$$

where $\beta_1 Dyears_{it}$ is a continuous variable indicating years since first diabetes diagnosis.

In an effort to capture possible non-linearities in the relationship of interest we then use a spline function that allows for the effect of an additional year with diabetes to vary over time.

$$Y_{it} = \delta_0 + g(Dyears_{it}) + \delta_2 X_{it} + c_i + u_{it}. \tag{4}$$

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 $n = 1, 2, \dots, N$. We choose three nodes that - based on visual inspection (see Figures 1, 2 and 3 in the result section) - best captured any possible non-linearity in the relationship between diabetes duration and labor outcomes. 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

[11] Our composite wealth index consists of owning a vehicle, owning a house or other real estate, owning another house, owning a washing machine, dryer, stove, refrigerator or furniture, owning any electric appliances, owning any domestic appliances, owning a bicycle and owning 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.

effects on labor market outcomes. The coefficient δ_n captures the effect of diabetes for the n-th interval. The effects are linear if $\delta_1 = \delta_2 =, \ldots, = \delta_n$.

Because the year of diagnosis was only reported in the third wave, duration of diabetes (or time since diagnosis) for the earlier waves was only calculated for those that had also responded to the third wave. To arrive at the time passed since diagnosis, the year of diagnosis was subtracted from the year of the interview.

One caveat of using FE is that, when year dummies are included, any variable that varies by one unit in each time period, is not separately identified (Wooldridge 2012). Because this is also the case for diabetes duration, in equations (3) and (4), identification of this variable relies on the presence of people without diabetes in the sample, for which diabetes duration does not increase at the same rate as time.[12] 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.[13]

4.3 Cross-section: biomarker and self-reported data

Self-reported diabetes only captures part of the diabetes population as many individuals remain undiagnosed; it may also contain cases of people who misreport suffering from diabetes. Estimations based on self-reports may therefore suffer from selection bias in at least tow ways:

1. Systematic overreporting of diabetes: people without diabetes may report a diabetes diagnosis, unintentionally — for instance due to misdiagnosis, either from a health professional or because of self-diagnosis, or intentionally — for instance with a view to justifying some other adverse event or status in their life (e.g. being unemployed).

2. Systematic underreporting of diabetes: people with diabetes may also underreport because they are concerned about negative stigma associated with the condition. Furthermore, diabetes often remains undiagnosed leaving people unaware of their condition. [14]

Overreporting may attenuate the effect of diabetes if those falsely reporting a diabetes diagnosis are in fact in good health; it may also lead to overestimation of the impact if some of those misreports reflect other factors that negatively affect labor outcomes (e.g. other illnesses or general ill health), or if they are used to justify other adverse events that may negatively affect labor outcomes. Similarly, underreporting may lead to

[12]Consequently, those that reported a diagnosis in the year of the interview were counted as 'one year since diagnosis'. From this follows that if the respondent reported to having been diagnosed in the year before the interview he or she was counted as 'two years since diagnosis' and so on.

[13]We also estimate all models excluding the calendar year dummies with similar results.

[14]More formally, assume that the true model of the effect of diabetes on labor market outcomes is y=$X^*\beta + \epsilon$. Because we do not observe the true values of X^* we have to use self-reported measures that contain errors: $X = X^* + u$. Since u may be correlated with ϵ - in contrast to classic measurement error which is randomly distributed, we cannot sign the bias of β.

overestimation if those with undiagnosed diabetes are generally healthier, hence more likely to have positive labor market outcomes than those with diagnosed diabetes. However, if the undiagnosed and the diagnosed groups are similar in terms of health, then this would lead to an underestimation of the effect of diabetes.

The health information received at a diabetes diagnosis may also have an effect in itself. It may for instance affect an individual's psychology which in turn may influence economic behavior. Two studies found a diabetes diagnosis and subsequent treatment to increase the odds of psychological problems, including depression and anxiety (Paddison et al. 2011; Thoolen et al. 2006), while similar results have not been found for people with undiagnosed diabetes (Nouwen et al. 2011). Looking at behavioral change, health information has been shown to affect behavior after the diagnosis of not only diabetes (Slade 2012) but also of other chronic diseases (see Baird et al. (2014), Gong (2015), Thornton (2008), and Zhao et al. (2013)). However, little is known about the effects of health information on labor market outcomes. For diabetes, only Liu et al. (2014) investigate the effect of receiving a diabetes diagnosis on labor income in Chinese employees. This study finds a reduction in labor, income which was attributed to the psychological effects of the diagnosis.[15]

The use of biomarker data allows to explore the relationship of measured diabetes, as well as of undiagnosed diabetes, with labor outcomes. The estimated relationship from those can then be compared to the estimated impact of self-reported diabetes. The biomarker data also enables us to look at diabetes severity, as measured by HbA1c values. Since this data is only available for one wave —the most recent— our analysis here is limited to cross-section data. Moreover, as mentioned above, biomarkers were only taken from about one-third of the initial representative sample, which is still a sizable number, but no longer directly comparable to the panel-based results in this paper. Nonetheless, it allows for a first exploration of the relationships of undiagnosed diabetes and disease severity with labor market outcomes.

Our analysis of the biomarker sample consists of three main steps. As a comparison, we first estimate Eq. 5 including only self-reported diabetes as our diabetes indicator.

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

. In a second step we investigate the association of biometrically measured diabetes (HbA1c $\geq$ 6.5%) with labor market outcomes:

$$Y_i = \beta_0 + \beta_1 Dbiom_i + \beta_2 X_i + c_i + u_i \tag{6}$$

where $\beta_1 Dbiom_i$ is equal to 1 if HbA1c $\geq$ 6.5%.[16] In a third step we estimate the following equation

$$Y_i = \beta_0 + \beta_1 Dud_i + \beta_2 Dsr_i + \beta_2 X_i + v_i + u_i. \tag{7}$$

[15] In a very different context Dillon et al. (2014), using a randomised intervention, find that the news stemming from diagnosis of malaria affects productivity and income, but not labor supply among sugar cane cutters in Nigeria.

[16] We include in the indicator for biometrically measured diabetes also those that self-reported a diabetes diagnosis but had HbA1c levels below the diabetes threshold. These people may have received a diabetes diagnosis but are well controlled and have HbA1c levels below the threshold. We therefore do not be assume that they falsely reported a diagnosis.

to investigate how the associations differ between people with self-reported and undiagnosed diabetes. $\beta_1 Dud_i$ identifies the effect of those undiagnosed and is equal to 1 if the person did not self-report a diabetes diagnosis but has an HbA1c $\geq 6.5\%$. $\beta_2 Dsr_i$ identifies the effect of those self-reported and is equal to 1 if the person self-reported a diabetes diagnosis. We rely on within household variation v_i for identification in these models using the biomarker data to account for unobserved community characteristics, such as the access to healthcare and the quality of healthcare in the community, poverty and unemployment levels in the community or the amount of public green space and recreational possibilities available. These factors potentially affect both the propensity to develop diabetes and to receive a diagnosis; they may also be related to labor market outcomes.[17]

5 Results

5.1 Incidence of self-reported diabetes

Table 2 presents the estimation results of the FE model using equation 2, which indicate significant and substantial reductions in the probability of employment for men and women with self-reported diabetes. The effects are surprisingly similar across both sexes showing a reduction in employment probabilities of over 5 p.p..

Table 2: Self-reported diabetes and labor market outcomes

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Diagnosed diabetes	$-.054^{**}$	$-.059^{**}$	0.054	0.081	$-.524$	-1.955
	(.025)	(.024)	(.067)	(.158)	(1.499)	(2.517)
N	21388	27341	13828	7068	17616	9112

Fixed effects estimation; robust standard errors in parentheses; reference category: non-agricultural 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.
The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural employment as the base) and for health insurance status.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The results in Columns 3–6 indicate no significant relationship between self-reported diabetes and wages or working hours. One may expect this relationship to differ 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 therefore estimate a model including interaction terms between self-reported diabetes and agricultural employment and between self-reported diabetes

[17] We did not account for fixed household characteristics as the average number of observations per household was close to one, i.e. for most households only one member provided biomarker information in our subsample, significantly limiting the variation within households that would be needed for identification.

and self-employment, respectively, using non-agricultural employment as the comparison group, and restricting our sample to those employed only.

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)
Diagnosed diabetes	0.105	0.064	0.617	$-.524$
	(.076)	(.169)	(1.606)	(2.252)
Diagnosed diabetes x agricultural worker	$-.242$	$-.409$	-5.495^{*}	-3.535
	(.188)	(.373)	(2.833)	(22.300)
Diagnosed diabetes x self-employed	$-.105$	0.125	0.306	-4.149
	(.192)	(.326)	(2.503)	(4.739)
N	13828	7068	17616	9112

Fixed effects estimation; robust standard errors in parentheses; reference category: non-agricultural 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 show up 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. We therefore perform a Wald test for the overall significance of the interaction term which does not reject the null of no interaction effects ($p = .15$), indicating that the effect of diabetes on working hours does not vary significantly by type of work.

In summary, we find no evidence for an association between self-reported diabetes and wages or working hours. The lack of effects at the intensive margin may be explained by selection: potentially, only those with "mild" or asymptomatic diabetes are still in the same job continuing to earn similar wages. 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 explore whether diabetes affects the selection into certain types of work we estimate FE models of the probability of being in non-agricultural employment, agricultural employment or self-employment using three dummy variables indicating the respective type of work as the left hand side variables. The results in Table 4 indicate a negative association with self-employment, though the estimates are quite imprecise. For women, those who self report diabetes are less likely to work in agriculture and potentially self-employment. This may suggest that having diabetes drives people out of self-employment and agricultural jobs, for instance because these jobs are physically more demanding and possibly also because they provide less protection in terms of insurance and employment duration. We also estimated a pooled multinomial logit model augmented with the within-between approach (Bell et al. 2015), based on the work of Mundlak (1978), which allows interpreting the coefficients of all time-varying variables as within-effects by including individual means of all time-varying covariates[18]. The results indicate a very similar pattern both in size and significance (results available on request).[19]

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
Diagnosed diabetes	−.006	−.008	−.043	−.001	−.022**	−.029
	(.029)	(.022)	(.026)	(.018)	(.009)	(.018)
N	20719	20719	20719	26577	26577	26577

Standard errors in parentheses

Robust standard errors in parentheses

Other control variables: state dummies, urbanisation dummies, education dummies, married dummy, number children < 6, wealth, age squared and calendar year dummies

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

5.2 Duration of self-reported diabetes

Because diabetes is a chronic and generally life-long disease, we investigate how soon after the first diagnosis diabetes may affect labor market 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 market 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 after some time complications start to appear, again reducing health and leading to reductions in labor supply and productivity.

To obtain a first 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 relationship between diabetes duration and the probability of employment

[18] Several other studies in economics have used this approach recently, e.g., Boll et al. (2016), Geishecker et al. (2011), and Wunder et al. (2014)

[19] Using the same methods, we also investigated the impact of diabetes on changes in the type of work restricted to those in employment, finding no evidence that diabetes leads to changes in the type of work. These results are also available on request.

for men shows a more or less steady decline which becomes more pronounced as time progresses. For women, a first drop-off occurs right after diagnosis; thereafter no consistent pattern is observed.[20]. A similar analysis for wages shows somewhat more erratic relationships, although there seems to be a long term negative trend for women but not for men (see figures 2 and 3). A similar negative trend can be observed for working hours for women, but not for men.

Figure 1: Kernel-weighted local polynomial regression of employment status on diabetes duration

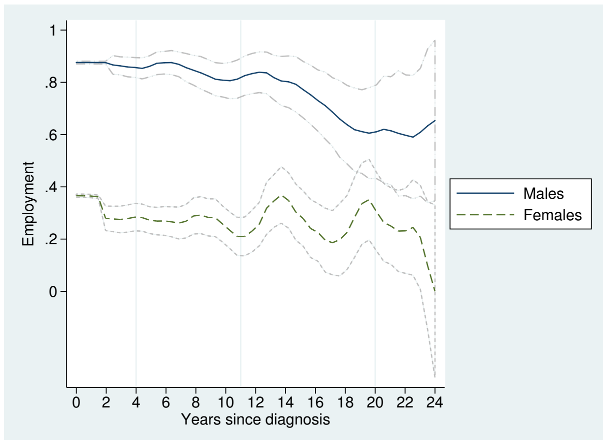
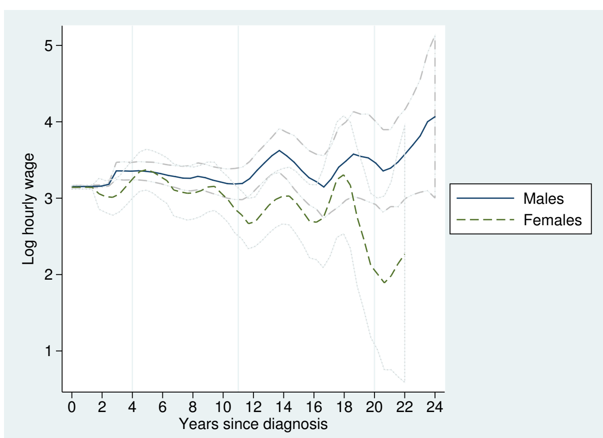


Figure 2: Kernel-weighted local polynomial regression of log hourly wages on diabetes duration



[20]Since long run estimations suffer from large standard errors — as sample size is strongly reduced — this limits its interpretation and we therefore truncate the graphs at a disease duration of 24 years

Figure 3: Kernel-weighted local polynomial regression of working hours on diabetes duration

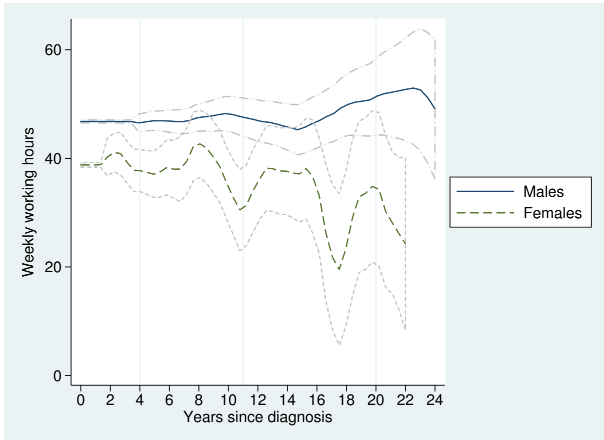


Table 5 presents the results of the linear and non-linear duration models (for which we created the following splines to capture the immediate, intermediate and long-term relationships:0–4, 5–11, 12–19 and 20+), starting with the results of the cross-sectional LPM, followed by the pooled LPM and then the FE model as specified in equations (3) and (4).

For employment probabilities the results indicate a yearly reduction in male employment probability throughout. For women the coefficient shows a reduction of up to almost 1 p.p. per year, though the association is not as strong. The coefficients in the spline models provide some evidence for an immediate effect of diabetes, which then levels off for some time after which it becomes stronger again. Nonetheless, for males and particularly females, the coefficients are quite imprecisely measured.

Turning to wages, the FE model indicates a reduction in female wages of about 7% per year with diabetes. This finding is, however, not supported by the other estimated models. For men we find no consistent effect. The results of the non-linear specification indicate that there may be a reduction in wages 5–11 years after the initial diagnosis. 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.[21]. There appears to be no consistent relationship between working hours and time since being diagnosed with diabetes.

Overall these results suggest a fairly constant decrease in the probability of employment for both men and women and in earnings for women. They are in contrast to estimates for the USA (Minor 2013), where no such linear relationship is observed. Our non-linear results are not directly comparable to those of Minor (2013), who used pooled cross-sectional data, constructed dummy variables instead of splines and also created different

[21]There are only 9 and 3 observations for male and female wages with more than 20 years since diagnosis in wave 3, respectively, and similarly 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.

Table 5: Relationship between self-reported years since diagnosis and labor market outcomes using continuous duration and duration splines

	Males			Females		
	(1) OLS (wave 3)	(2) Pooled OLS	(3) FE	(4) OLS (wave 3)	(5) Pooled OLS	(6) FE
			Employment probabilities			
Panel A:						
Diabetes duration (linear)	$-.008^{***}$	$-.007^{***}$	$-.017^{***}$	$-.005^{***}$	$-.004^{***}$	$-.009^{*}$
	(.002)	(.002)	(.006)	(.002)	(.001)	(.005)
Panel B:						
Diabetes duration (splines)						
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)
N	8217	16292	16292	10467	22407	22407
			Log hourly wage			
Panel A:						
Diabetes duration (linear)	0.001	0.010^{**}	$-.019$	$-.014^{*}$	$-.009$	$-.073^{**}$
	(.006)	(.005)	(.018)	(.008)	(.008)	(.029)
Panel B:						
Diabetes duration (splines)						
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)
N	5509	10767	10767	2874	5741	5741
			Weekly working hours			
Panel A:						
Diabetes duration (linear)	0.069	0.048	0.181	$-.020$	$-.124$	0.208
	(.124)	(.102)	(.330)	(.187)	(.127)	(.652)
Panel B:						
Diabetes duration (splines)						
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)
N	6807	13579	13579	3591	7383	7383

The table presents the results of three estimation methods for the three dependent variables: employment probabilities, log hourly wages and 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, urbanisation dummies, education dummies, married dummy, number children < 6, wealth, age squared and calendar year dummies. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural employment as the base) and for health insurance status. The OLS and pooled OLS models additionally control for age.

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

duration groups. He finds a reduction in employment probabilities of 82 p.p. for females after 11 to 15 years and a reduction of 60 p.p. for males after 2-5 years, indicating very large employment penalties, in particular in comparison to our results for Mexico.[22]

5.3 Cross-sectional biomarker analysis

In this section we gain additional insights from using the biomarker data collected in the third wave of the MxFLS. As noted in section 3, these data enable us to identify respondents with HbA1c levels equal to or above the internationally recognized diabetes threshold of 6.5%. This will allow the investigation of the direction of bias introduced when relying on self-reported diabetes only, in cases where it is not possible to identify the undiagnosed as well.

5.3.1 Biomarkers versus self-reported diabetes

The results of the complete biomarker analysis are presented in Table 6. First, we estimate the model solely accounting for self-reported diabetes using only the data of the biomarker sample (columns (1) and (2)). The results still indicate reduced employment probabilities for people with self-reported diabetes, nonetheless, the estimates are less precise, potentially due to the smaller size and somewhat different age composition of the biomarker sample.[23] In the next step, we estimate Eq. 6 using biometrically measured diabetes and find evidence for an adverse association of biometrically measured diabetes and the probability of employment for women (columns (3) and (4)), though the coefficient is smaller compared with self-reported diabetes. This suggests that relying on self-reported diabetes exclusively may lead to an upward bias due to those undiagnosed experiencing less pronounced adverse labor market effects. However, only using the threshold to identify diabetes does not take into account the potential effect of receiving a diabetes diagnosis, as discussed in Section 4.3, nor does it allow for a distinct investigation of undiagnosed diabetes. Including both self-reported and undiagnosed diabetes and estimating Eq. 7, we find no association between undiagnosed diabetes and any labor outcome (columns (5) and (6)). For the probability of employment, the self-reported diabetes coefficients are now almost the same for men and women, indicating a reduction similar to when we did not account for undiagnosed diabetes. Consistent with our earlier findings there is no indication of an association between any form of diabetes and wages or working hours.

[22] We estimated a comparable model to that of Minor (2013) using dummy variables and find a significant reduction in employment chances throughout, regardless of whether we use our duration groups to construct the dummies or the duration groups used by Minor (2013). For men, we find a significant reduction of about 6 to 12 p.p., depending on the used specification, in the first 2 and 4 years after diagnosis, respectively. In the following years the effect size tends to increase somewhat. For women, we find less evidence for an immediate effect of diagnosis, but effects do emerge after about 2 years of living with the disease and also increase somewhat over time. These results are available on request.

[23] To test how the different age composition of and potential non-random selection into the biomarker sample may affect the estimates, we estimated the same model but applying it to the complete sample of wave 3. The coefficients indicate a larger association and are more precise, suggesting that both samples are not directly comparable (results available on request). However, because we are more generally interested in the effect of accounting for the measurement error of self-reported diabetes, the potential selection into the biomarker sample should be of minor importance.

Table 6: Biomarker results (community level FE)

	Only self-reported		HbA1c $\geq 6.5\%$		Self-reported and undiagnosed	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Dependent variable: Employment						
Diagnosed diabetes	$-.049^{*}$	$-.043^{*}$			$-.048^{*}$	$-.048^{**}$
	(.026)	(.023)			(.026)	(.023)
Undiagnosed diabetes					0.005	$-.017$
					(.019)	(.018)
HbA1c $\geq 6.5\%$			$-.013$	$-.029^{*}$		
			(.015)	(.017)		
R2	0.079	0.108	0.078	0.107	0.079	0.108
N	2785	3623	2790	3626	2785	3623
Dependent variable: Log hourly wages						
Diagnosed diabetes	$-.008$	$-.038$			$-.004$	$-.050$
	(.066)	(.113)			(.068)	(.117)
Undiagnosed diabetes					0.018	$-.052$
					(.048)	(.079)
HbA1c $\geq 6.5\%$			0.011	$-.054$		
			(.044)	(.069)		
R2	0.146	0.236	0.146	0.239	0.146	0.237
N	1803	884	1805	885	1803	884
Dependent variable: Usual weekly working hours						
Diagnosed diabetes	$-.338$	$-.780$			$-.301$	$-.176$
	(1.286)	(2.164)			(1.325)	(2.216)
Undiagnosed diabetes					0.160	2.635
					(.955)	(1.838)
HbA1c $\geq 6.5\%$			0.000	1.699		
			(.835)	(1.424)		
R2	0.031	0.035	0.031	0.037	0.031	0.038
N	2302	1144	2306	1145	2302	1144

Community FE estimation, robust standard errors in parentheses. Other control variables: age, age squared, state dummies, urbanisation dummies, education dummies, married dummy, number of 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 as the base) and for health insurance status. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Diagnosed, undiagnosed diabetes, diabetes severity and self-reported health and their association with labor market outcomes (community level FE)

	Employment		Log hourly wages		Weekly working hours	
	(1)	(2)	(3)	(4)	(5)	(6)
	Males	Females	Males	Females	Males	Females
Panel A (self-reported health)						
Diagnosed diabetes	−.032	−.040*	0.004	0.016	0.111	−.668
	(.025)	(.024)	(.069)	(.109)	(1.374)	(2.065)
Undiagnosed diabetes	0.005	−.019	0.020	−.049	0.119	2.753
	(.019)	(.018)	(.049)	(.079)	(.960)	(1.811)
Self-reported health status						
good	0.024	0.059*	0.066	−.115	−1.169	3.423*
	(.027)	(.035)	(.070)	(.136)	(1.342)	(2.054)
fair	−.005	0.006	0.031	−.152	−1.646	4.540*
	(.026)	(.033)	(.070)	(.141)	(1.404)	(2.405)
bad	−.125**	−.025	−.018	−.376*	−6.124**	6.882*
	(.056)	(.046)	(.174)	(.200)	(2.678)	(3.972)
very bad	−.167	0.119	−.347*	0.306	−1.427	−17.360***
	(.127)	(.153)	(.182)	(.433)	(3.274)	(4.914)
R2	0.089	0.111	0.148	0.244	0.035	0.048
N	2785	3621	1803	882	2302	1142
Panel B (HbA1c levels)						
Self-reported diabetes						
$6.5 \leq$ HbA1c < 8	−.130**	−.068	−.211*	0.003	1.878	−5.682
	(.056)	(.049)	(.122)	(.245)	(2.727)	(4.797)
$8 \leq$ HbA1c < 12	−.051	−.104***	0.040	0.026	−2.193	−1.369
	(.033)	(.032)	(.091)	(.167)	(1.861)	(3.341)
HbA1c ≥ 12	0.015	−.027	−.103	−.291	−.361	−.676
	(.046)	(.049)	(.143)	(.230)	(3.071)	(3.603)
Undiagnosed diabetes						
$6.5 \leq$ HbA1c < 8	0.018	0.002	0.020	−.037	0.928	3.710
	(.021)	(.025)	(.058)	(.094)	(1.175)	(2.318)
$8 \leq$ HbA1c < 12	0.011	−.018	0.011	−.146	−1.016	−.437
	(.037)	(.031)	(.078)	(.134)	(1.511)	(2.716)
HbA1c ≥ 12	0.008	−.062	−.018	0.183	−1.740	1.591
	(.042)	(.045)	(.087)	(.185)	(2.100)	(4.066)
R2	0.026	0.079	0.146	0.232	0.029	0.040
N	2785	3623	1804			

Community FE estimation, robust standard errors in parentheses. Other control variables: age, age squared, state dummies, urbanisation dummies, education dummies, married dummy, number of 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 as the base) and for health insurance status. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Overall, the finding that undiagnosed diabetes is not associated with any adverse labor outcomes suggests that the two populations with diabetes, i.e. those self-reporting the condition and those undiagnosed, are different in ways that are not yet captured by our models. Potential reasons include, first, the differences in subjective health between self-reporting and undiagnosed people with diabetes, and second, another, potentially related issue, i.e. the differences in diabetes severity. To investigate the former, we add a measure of self-reported health and find that the coefficient of self-reported diabetes is reduced in size, no longer significant for males and only borderline significant for females (Table 7, Panel A). Particularly for males reporting bad health is associated with a reduction in employment probabilities, suggesting that it is bad health in self-reporting males that is driving the reduction in employment probabilities, while it does not affect the coefficient of undiagnosed diabetes. For women additional factors may play a role.However, accounting for measures of overweight and obesity, self-reported hypertension, heart disease, depression and insurance status does not further affect the interpretation of the diabetes coefficient (results available on request).

To investigate the second issue, we use the HbA1c information in wave three as a proxy for current disease severity and diabetes management. If current disease severity would be related to labor market outcomes, one would expect an adverse association with increasing HbA1c levels, for both self-reporting and undiagnosed. To investigate this, we construct three dummy variables for HbA1c groups above the diabetes threshold, i.e. 6.5–7.9, 8–11.9 and 12–14 (Table 7, Panel B). For employment probabilities we again find adverse associations for those with self-reported diabetes, particularly for HbA1c levels of 6.5% to 7.9% and 8% to 11.9%, but no indication that the association would increase with higher HbA1c levels. For undiagnosed diabetes, no association with increasing HbA1c levels is observed. We therefore conclude that current diabetes management or severity of diabetes are likely not associated with any labor market outcome.

To the best of our knowledge only one study has previously used biomarker data to analyze the relationship with labor market outcomes in a comparable population. H. S. Brown, Perez, et al. (2011) use data for a Mexican American population in a broadly comparable way to this paper, though stopping short of investigating the labor market impact of undiagnosed diabetes. Similar to us, the study finds that once diabetes is diagnosed, current management plays a minor role in determining labor market 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.

6 Conclusion

Diabetes has become 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 the people it affects. In light of this, the demand for thus far extremely scarce evidence on the associated economic consequences has been large.

We contribute to filling this gap by providing important additional evidence on the globally much under-researched labor market effects of diabetes, taking into account several methodological challenges that arise due to the nature of the disease. To do this we use detailed longitudinal panel data from Mexico, a MIC for which the biomarker data used in this paper indicates that diabetes, and particularly undiagnosed diabetes, has reached alarming levels.

In terms of the methodology this is —to the best of our knowledge— the first paper that comprehensively addresses several issues related to the estimation of labor market effects of diabetes. We account for the endogeneity of self-reported diabetes by estimating individual fixed effects, which allows us to account for unobserved individual traits that likely affect both labor market outcomes and the propensity to be self-report diabetes. Within this framework, we further use information on the year of diagnosis to investigate the potential heterogeneity in the effect of self-reported diabetes on labor market outcomes over time. Finally, we are able to investigate the potential bias in estimates relying on self-reported diabetes, taking advantage of biomarker data allowing for the identification of people with undiagnosed diabetes.

The first part of our results confirms a considerable gap in employment for both men and women reporting a diabetes diagnosis. We also find some evidence that diabetes is more likely to reduce the probability to be employed in the agricultural and self-employment sector, where informality is quite common and work may be more strenuous compared to non-agricultural employment. However, those who remain employed do not suffer any wage or labor supply effects, possibly because they are still relatively healthy or are able to resort to a type of work where their diabetes does not inhibit their performance. Future research will be needed to confirm and further investigate this finding and interpretation. Regarding the heterogeneity in the effects of diabetes over time our results indicate a continuous adverse relationship of self-reported diabetes with employment chances, which appears to increase especially after the first 10 years from diagnosis. This is consistent with the observation that many complications of diabetes increasingly appear after some time of living with the disease (Adler et al. 2003). Our analysis on wages also uncovers some adverse effects for females, indicating a quite important reduction with time since diagnosis. These findings may bode ill for countries, including diabetes, were diabetes has started appearing at an increasingly younger age, 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).

The second part of our results indicates that relying on self-reported diabetes can lead to an overestimation of the association of diabetes with labor market outcomes, in particular in relation to employment probabilities. We find that a negative association only exists for the self-reporting group, while none is found for the undiagnosed group. This difference in results, is at least for men likely mediated by the subjective health status being worse for self-reporting than undiagnosed men. However, actual current disease severity does not appear to be related to any labor outcomes.

These findings have several implications. First, methodologically, when interpreting findings relying on self-reported diabetes it cannot be assumed that the results extend to those with undiagnosed diabetes, or are attenuated due to classical measurement error, nor will the use of an IV strategy deal with this problem (Cawley et al. 2015). However, we do not think that lumping both groups, self-reporting and undiagnosed, together into one diabetes group is the most sensible approach, as this will not account for the heterogeneity between the groups in the amount of health information they possess, the time they have already been exposed to elevated blood glucose levels and consequently their subjective as well as true health status, leading to an important loss of information. In our opinion, the best way to analyze the effects of diabetes is to explicitly include both groups in the model, thereby acknowledging their inherent differences and receiving important information about the distribution of the economic burden across the two groups.

Further, because the results of the biomarker analysis did also show that the coefficient of self-reported diabetes is not strongly affected by the inclusion of undiagnosed diabetes in the model, we think it is reasonable to assume that the estimates using self-reported diabetes in the FE panel data analysis are still largely unbiased. However, they can only be used to draw conclusions about the effect of self-reported diabetes, not diabetes overall. So given that more than 7% of the Mexican population have been diagnosed with diabetes —a figure that is expected to increase further (Meza et al. 2015)— the found reduction in employment probabilities still suggests a significant economic burden caused by diabetes in Mexico.

Hence our results support the need to reduce the incidence and progression of diabetes, with the potential to lead to important economic gains by increasing the productive lifespan of people. This is of particular importance in MICs, where parental health shocks, related job loss and increasing health expenditures can affect the entire family. Other family members, including children, may be forced to increase their labor supply and to reduce other expenditures to maintain their economic status. Both 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). Further, the large portion of undiagnosed people indicates that diagnosis appears to happen too late or not at all, reducing the possibility to prevent complications via appropriate treatment and hence increasing the risk of severe complications appearing early. As a result, much of the burden may be prevented by earlier diagnosis and, given the generally little success in achieving good control in the diagnosed population in Mexico, better treatment. Of course, eventually the aim should be to prevent more diabetes cases and it will be interesting to see if measures, such as taxing beverages high in sugar (), will be successful in this respect.

Appendix

A Strategies to deal with measurement error

Reporting error is likely to pose a considerable challenge in the use of self-reported data. Fortunately, the MxFLS data provides several possibilities to assess the amount of misreporting and to attempt to reduce it before estimating the labor market effects of diabetes. In what follows we describe our approach of dealing with inconsistencies in self-reported diabetes over time.

One of the key advantages of panel data is in its repeated measurement for many of the individuals, thereby allowing to uncover inconsistencies for those with at least two observations. While we are not aware of any literature investigating the issue of inconsistencies in self-reported diabetes over time, a study by Zajacova et al. (2010), on the consistency of a self-reported cancer diagnosis over time in a USA population, found that 30% of those who had reported a cancer diagnosis at an earlier point did report at a later point that they never had received a cancer diagnosis. They also found that a more recent diagnosis was reported with greater consistency possibly due to increasing recall problems and/or reduced salience as time since diagnosis progresses.

We also find inconsistencies in the diabetes self-reports over the three waves of the MxFLS data, with between 10–20% of those reporting diabetes in one wave not doing so in one of the subsequent waves. In order to reduce the amount of inconsistencies we were interested in the validity of diabetes self-reports. While we could not find a study assessing the validity of self-reported diabetes in Mexico, a study from China has shown that specificity of self-reported diabetes —meaning that those who self-report a diabetes diagnosis actually have diabetes— was very high (>98% for China), while sensitivity - a measure of how many people with diabetes, diagnosed or undiagnosed, actually self-report the disease - was low (40% for China) (Yuan et al. 2015). This indicates that people who report a diabetes diagnosis are likely to indeed have the condition while many of those not reporting a diabetes diagnosis are unaware of their diabetes.

We do something similar and assess the validity of self-reported diabetes in our data by using HbA1c levels and the self-reports of diabetes related medicine use from wave three. We find that 90% of those self-reporting a diabetes diagnosis had an HbA1c $\geq 6.5\%$ or did report taking diabetes medication, indicating relatively high specificity in our data as well.

We used this information of relatively high specificity to infer the "true" diabetes status for those with inconsistent reports. For those with two waves, we assumed that if they reported a diabetes diagnosis in a prior wave they also had diabetes in the ensuing wave, even if they did not report a diabetes diagnosis. For people where we had data from all three waves, we used that additional information to make a decision on how to deal with inconsistencies using the rules outlined in Table 8:

This approach should add more consistency to the self-reported diabetes information by using all available information. We tested if the results of the HbA1c tests for the subpopulation in 2009 with biomarker information and inconsistencies in their diabetes reporting (n=96, 48 with two and 48 with one self-report of diabetes) would support

Table 8: Inconsistencies in diabetes self-report in MxFLS

Inconsistency	Assumption	Number of observations replaced
Diabetes self report in 2002, 2005 but not in 2009	Has diabetes in 2009 as well	34
Diabetes self report in 2002, 2009 but not in 2005	Has diabetes in 2005 as well	94
Diabetes self report only in 2002, but not in 2005 and 2009	Has no diabetes in 2002 either	86
Diabetes self report only in 2005, but not in 2002 and 2009	Has no diabetes in 2005 either	71
Diabetes self report in 2002, but not in 2005. Not in survey in 2009	Has diabetes in 2005 as well	43
Diabetes self report in 2005, but not in 2009. Not in survey in 2002	Has diabetes in 2009 as well	32

this decision. Therefore we compared the mean HbA1c values for those who had two self-reports of a diagnosis of diabetes in the full three waves with those with only one self-report. Using a t-test we found a significantly ($p<0.001$) higher mean HbA1c of 9.6% for those with two self-reports compared to 7.0% for those with only one self-report of diabetes. Further, of those with one self-report, only 30% had an HbA1c$\geq$ 6.5% compared to 87% of those with two self-reports. Based on these results we are reassured that the way we have dealt with the inconsistencies in the data minimizes misclassification of people into diabetes or no-diabetes and reduce some of the measurement error in the diabetes data. Unfortunately we cannot use a similar method for dealing with inconsistencies in the self-reported year of diabetes diagnosis as it has only been reported once. Hence, the results from duration analysis should be interpreted with care.

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., R. J. Stevens, S. E. Manley, R. W. Bilous, C. A. Cull, and R. R. Holman (2003). "Development and progression of nephropathy in type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS 64)". *Kidney International* 63 (1), pp. 225–232. DOI: 10.1046/j.1523-1755.2003.00712.x.

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

Baird, S., E. Gong, C. McIntosh, and B. Ozler (2014). "The heterogeneous effects of HIV testing." *Journal of health economics* 37C, pp. 98–112. DOI: 10.1016/j.jhealeco.2014.06.003.

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

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

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

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

Bell, A. and K. Jones (2015). “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data”. *Political Science Research and Methods* 3 (01), pp. 133–153. DOI: 10.1017/psrm.2014.7.

Bellamy, L., J.-P. Casas, A. D. Hingorani, and D. Williams (2009). “Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis.” *Lancet* 373 (9677), pp. 1773–9. DOI: 10.1016/S0140-6736(09)60731-5.

Bergemann, A., E. Grönqvist, and S. Gudbjörnsdottir (2011). “The effects of job displacement on the onset and progression”.

Boll, C., J. S. Leppin, and K. Schömann (2016). “Who is overeducated and why? Probit and dynamic mixed multinomial logit analyses of vertical mismatch in East and West Germany”. *Education Economics* (661), pp. 1–24. DOI: 10.1080/09645292.2016.1158787.

Bratti, M. and M. Mendola (2014). “Parental health and child schooling”. *Journal of Health Economics* 35 (1), pp. 94–108. DOI: 10.1016/j.jhealeco.2014.02.006.

Brown, H. S., J. A. Pagan, and E. Bastida (2005). “The impact of diabetes on employment: Genetic IVs in a bivariate probit”. *Health Economics* 14 (5), pp. 537–544. DOI: 10.1002/hec.942.

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

Brown, T. T. (2014). “How effective are public health departments at preventing mortality?” *Economics and human biology* 13, pp. 34–45. DOI: 10.1016/j.ehb.2013.10.001.

Cawley, J., J. C. Maclean, M. Hammer, and N. Wintfeld (2015). “Reporting error in weight and its implications for bias in economic models”. *Economics & Human Biology* 19, pp. 27–44. DOI: 10.1016/j.ehb.2015.07.001.

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

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

Dillon, A., J. Friedman, and P. M. Serneels (2014). “Health information, treatment, and worker productivity: Experimental evidence from malaria testing and treatment among Nigerian sugarcane cutters”.

Eriksson, A.-K., M. van den Donk, A. Hilding, and C.-G. Ostenson (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), pp. 2683–2689. DOI: 10.2337/dc12-1738.

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

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

Frankenberg, E., J. Y. Ho, and D. Thomas (2015). “Biological Health Risks and Economic Development”.

Geishecker, I. and T. Siedler (2011). “Job Loss Fears and (Extreme) Party Identification: First Evidence from Panel Data”.

Gong, E. (2015). “HIV Testing and Risky Sexual Behaviour”. *The Economic Journal* 125 (582), pp. 32–60. DOI: 10.1111/ecoj.12125.

Gregg, E. W., H. Chen, L. E. Wagenknecht, J. M. Clark, L. M. Delahanty, J. Bantle, H. J. Pownall, K. C. Johnson, M. M. Safford, A. E. Kitabchi, F. X. Pi-Sunyer, R. R. Wing, A. G. Bertoni, and for the Look AHEAD Research Group (2012). “Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes”. *JAMA* 308 (23), p. 2489. DOI: 10.1001/jama.2012.67929.

Heraclides, A. M., T. Chandola, D. R. Witte, and E. J. Brunner (2012). “Work Stress, Obesity and the Risk of Type 2 Diabetes: Gender-Specific Bidirectional Effect in the Whitehall II Study”. *Obesity* 20 (2), pp. 428–433. DOI: 10.1038/oby.2011.95.

Hu, F. B. (2011). “Globalization of Diabetes: The role of diet, lifestyle, and genes”. *Diabetes Care* 34 (6), pp. 1249–1257. DOI: 10.2337/dc11-0442.

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

Kapteyn, A., J. P. Smith, and A. Van Soest (2009). “Work disability, work, and justification bias in Europe and the United States”.

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

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

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

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

Meza, R., T. Barrientos-Gutierrez, R. Rojas-Martinez, N. Reynoso-Noverón, L. S. Palacio-Mejia, E. Lazcano-Ponce, and M. Hernández-Ávila (2015). "Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates". *Preventive Medicine*. DOI: 10.1016/j.ypmed.2015.10.015.

Minor, T. (2011). "The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey". *Health Economics*. 15th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research, ISPOR 2010 Atlanta, GA United States. Conference Start: 20100515 Conference End: 20100519 20 (12), pp. 1468–1486. DOI: 10.1002/hec.1685.

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

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

Mundlak, Y. (1978). "On the Pooling of Time Series and Cross Section Data". *Econometrica* 46 (1), pp. 69–85. DOI: 10.2307/1913646.

NCD Risk Factor Collaboration (2016). "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with $4 \cdot 4$ million participants". *The Lancet* 387, pp. 1513–1530. DOI: 10.1016/S0140-6736(16)00618-8.

Nouwen, A., G. Nefs, I. Caramlau, M. Connock, K. Winkley, C. E. Lloyd, M. Peyrot, and F. Pouwer (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), pp. 752–762. DOI: 10.2337/dc10-1414.

O'Neill, D. and O. Sweetman (2013). "The consequences of measurement error when estimating the impact of obesity on income". *IZA Journal of Labor Economics* 2 (1), p. 3. DOI: 10.1186/2193-8997-2-3.

Paddison, C. A. M., H. C. Eborall, D. P. French, A. L. Kinmonth, A. T. Prevost, S. J. Griffin, and S. Sutton (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), pp. 213–226. DOI: 10.1348/135910710X495366.

Perks, T. A. (2015). "Obesity and its relation to employment income: Does the bias in self-reported BMI matter?" *Canadian Studies in Population* 42 (3-4), pp. 1–10.

Reynoso-Noverón, N., R. Mehta, P. Almeda-Valdes, R. Rojas-Martinez, S. Villalpando, M. Hernández-Ávila, and C. a. Aguilar-Salinas (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), p. 1. DOI: 10.1186/1475-2840-10-1.

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

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

Schneiderman, N., G. Ironson, and S. D. Siegel (2005). "Stress and Health: Psychological, Behavioral, and Biological Determinants". *Annual Review of Clinical Psychology* 1 (1), pp. 607–628. DOI: 10.1146/annurev.clinpsy.1.102803.144141.

Schneiderman, N., M. Llabre, C. C. Cowie, J. Barnhart, M. Carnethon, L. C. Gallo, A. L. Giachello, G. Heiss, R. C. Kaplan, L. M. LaVange, Y. Teng, L. Villa-Caballero, and M. L. Avilés-Santa (2014). "Prevalence of Diabetes Among Hispanics/Latinos From Diverse Backgrounds: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)". *Diabetes Care* 37 (8), pp. 2233–2239. DOI: 10.2337/dc13-2939.

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

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

Slade, A. N. (2012). "Health Investment Decisions in Response to Diabetes Information in Older Americans". *Journal of Health Economics* 31 (3), pp. 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), pp. 281–293. DOI: 10.1016/j.ehb.2012.02.009.

Strauss, J. and D. Thomas (1998). "Health, Nutrition, and Economic Development". *Journal of Economic Literature* 36 (2), pp. 766–817. DOI: http://www.aeaweb.org/jel/index.php.

Thoolen, B. J., D. T. De Ridder, J. M. Bensing, K. J. Gorter, and G. E. Rutten (2006). "Psychological outcomes of patients with screen-detected type 2 diabetes: The influence of time since diagnosis and treatment intensity". *Diabetes Care* 29 (10), pp. 2257–2262. DOI: 10.2337/dc06-0617.

Thornton, R. L. (2008). "The Demand for, and Impact of, Learning HIV Status." *The American economic review* 98 (5), pp. 1829–1863. DOI: 10.1257/aer.98.5.1829.

Villalpando, S., V. de la Cruz, R. Rojas, T. Shamah-Levy, M. A. Avila, B. Gaona, R. Rebollar, and L. Hernández (2010). "Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey." *Salud pública de México* 52 Suppl 1 (1), S19–26.

Williams, A. L., S. B. R. Jacobs, H. Moreno-Macías, A. Huerta-Chagoya, C. Churchhouse, C. Márquez-Luna, H. García-Ortíz, M. J. Gómez-Vázquez, N. P. Burtt, C. a. Aguilar-Salinas, C. González-Villalpando, J. C. Florez, L. Orozco, C. a. Haiman, T. Tusié-Luna, and D. Altshuler (2014). "Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico." *Nature* 506 (7486), pp. 97–101. DOI: 10.1038/nature12828.

Wooldridge, J. (2012). *Introductory Econometrics. A Modern Approach.* 5th ed. Cengage Learning.

World Health Organization (2011). "Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation". (WHO/NMH/CHP/CPM/11.1), p. 25. DOI: WHO/NMH/CHP/CPM/11.1.
Wunder, C. and R. T. Riphahn (2014). "The dynamics of welfare entry and exit amongst natives and immigrants". *Oxford Economic Papers* 66 (2), pp. 580–604. DOI: 10.1093/oep/gpt025.
Yuan, X., T. Liu, L. Wu, Z.-Y. Zou, and C. Li (2015). "Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study". *BMJ Open* 5 (4), e006633–e006633. DOI: 10.1136/bmjopen-2014-006633.
Zajacova, A., J. Dowd, R. F. Schoeni, and R. B. Wallace (2010). "Consistency and precision of cancer reporting in a multiwave national panel survey". *Population Health Metrics* 8 (1), p. 20. DOI: 10.1186/1478-7954-8-20.
Zhang, X., X. Zhao, and A. Harris (2009). "Chronic diseases and labour force participation in Australia". *Journal of Health Economics* 28 (1), pp. 91–108. DOI: 10.1016/j.jhealeco.2008.08.001.
Zhao, M., Y. Konishi, and P. Glewwe (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), pp. 367–385. DOI: 10.1016/j.jhealeco.2012.11.007.