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

Till Seuring[*], Pieter Serneels[†], Marc Suhrcke[‡]

November 24, 2015

Abstract

Diabetes is increasingly recognized as a major health risk worldwide, ~~including in developing countries . Although~~ not only in high income but also in low and middle income countries (LMICs). While adverse economic effects ~~may potentially be large, there is so far limitedhard evidence~~are highly plausible, the existing empirical evidence is very limited. This paper investigates the effects of diabetes on labor market outcomes focusing on Mexico, a country with high and persistent diabetes rates. Two challenges present themselves when studying those consequences with survey data: ~~first,~~ (1)causality is hard to identify, and ~~second~~ (2) measurement of diabetes is typically ~~through~~ based on self-reports, ~~and it is unknown whether this introduces a bias~~potentially causing biased estimates. This paper makes headway on both fronts. To study the relationship between self-reported diabetes and labor outcomes we use rich panel data. Making use of fixed effects estimation, the analysis accounts for time-invariant omitted variables, providing an improved identification strategy compared to existing work on the labor consequences of diabetes even in high income countries. The results indicate a strong negative relationship between self-reported diabetes and the probability of employment, which is reduced by 6.5 percentage points for those who self-report to be ~~suffering from diabetes~~diabetic. We find no evidence for an adverse relationship with wages or working hours. Further analysis indicates that each additional year further reduces employment chances, with the adverse relationship strongest after the first ten years since diagnosis.

We then use biomarker data for ~~a~~ the most recent cross section. The results show that the negative relationship remains when using this objective measure, but is smaller in size and is driven by those with diagnosed diabetes. Results are small and insignificant for ~~these~~ those with undiagnosed diabetes. This indicates that estimates based on self-reported diabetes may overstate the employment effect of diabetes. It also raises the possibility that diagnosis itself ~~has an effect~~plays a role, or that people with certain characteristics self-select into diagnosis.

[*]University of East Anglia

[†]University of East Anglia

[‡]University of York

1 Introduction

Diabetes has been increasing worldwide and is expected to continue to do so over the next decades. 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 et al., 2014). In Mexico itself diabetes prevalence has risen from 6.7 percent in 1994 to 14.4 percent in 2006, including both diagnosed and undiagnosed cases (Barquera et al., 2013) and is expected to increase further over the next decades (Meza et al., 2015). Thereby diabetes is already the number one cause of death in Mexico (Barquera et al., 2013). This increase in prevalence stems both from a deterioration in diet and a reduction in physical activity (Barquera et al., 2008; Basu et al., 2013), as well as from a distinct genetic predisposition of many Mexicans with pre-hispanic ancestry (Williams et al., 2014). There is also evidence that the onset of diabetes is increasingly happening at an earlier age in Mexico (Villalpando et al., 2010) which will increase the likelihood of experiencing complications relatively early in the productive lifespan (Barquera et al., 2013). This will especially be the case if the treatment of the disease after diagnosis remains as ineffective as it currently is, with only a minority achieving adequate blood glucose control (Barquera et al., 2013). Due to its well known health effects, diabetes is causing increasing healthcare expenditures and potentially also affects labor market outcomes (Seuring et al., 2015a). Studies for the general USA population as well as those focusing on Mexican Americans have found reductions in employment chances as well as wages and labor supply (Brown et al., 2005; Brown, 2014; Brown I I I et al., 2011; Minor, 2011, 2013). An earlier cross section analysis for Mexico found a reduction in employment chances of 10 percentage points for Mexican men (Seuring et al., 2015b)

However, while these studies have provided good evidence of the potential labor market effects of diabetes , many of the complexities of the relationship have remained unaddressed. Diabetes is a term used to describe various diseases characterized by high blood glucose values, with the predominant disease being type II diabetes mellitus, which we will henceforth refer to when using the term diabetes. It is characterized by elevated blood glucose levels due to the body not being able to use insulin properly to maintain blood glucose at normal levels. Elevated blood glucose levels over a prolonged period of time can lead to irreversible health 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). All these conditions can be significantly debilitating and therefore may reduce an individual's economic activity, including his productivity and labor ~~supply~~market participation. It is important to bear in mind, however, that diabetes is not a homogeneous disease and can have different effects on people's health depending on the success of the long term disease management. On one hand people who are able to "reverse" their diabetes - meaning they manage to get back to healthy blood glucose levels as a result of lifestyle changes and medication that successfully re-establish normal insulin sensitivity - are unlikely to suffer from any diabetes related health problems (Lim et al., 2011; Gregg et al., 2012). On the other hand, if diabetes is either completely untreated or unsuccessfully treated by medicine or lifestyle changes, and blood glucose levels and insulin resistance remain elevated, then people are likely to develop the ~~already mentioned~~ above-mentioned adverse health conditions. The temporal aspect of diabetes and the differences in its potential health effects have to be taken into account when investigating its labor market impact. Further, apart from its health effects diabetes might also affect labor market outcomes through other channels. ~~Employers~~ For instance, employers may discriminate against people with diabetes due to their health problems and greater need for medical treatment. Moreover, people aware of their condition may also be less inclined to continue working if this interferes with their ~~management of the disease~~ disease management; they may also use the diagnosis as a justification for withdrawing from work, i.e. the so called justification bias (Kapteyn et al., 2009). For these reasons the labor market effects may also be distinct for people with diagnosed versus those with undiagnosed diabetes.

Another challenge with estimating the relationship between diabetes and labor outcomes is unobserved heterogeneity. A major source of potential unobserved heterogeneity is related to ~~time-fixed~~ time-invariant individual unobservables. Personal characteristics~~such as unobserved abilities~~, including ability, health during uteru, infant and child years – often related to low household income or adverse health shocks during these early years, as well as risk preferences have been shown to adversely affect health ~~and more specifically~~ in general and the propensity to develop type 2 diabetes ~~(van Ewijk, 2011; Sotomayor, 2013; Li e~~ ~~and may~~ more specifically (van Ewijk, 2011; Sotomayor, 2013; Li et al., 2010) . They may also affect employment chances, wages or working hours indirectly through their adverse effects on educational attainment (Ayyagari et al., 2011) as well as directly through their effects on today's productivity (Currie and Vogl, 2013).

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes~~in~~ , while improving upon existing estimation 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 years 2002, 2005-2006 and ~~2009-2013~~2009-2012. The MxFLS is particularly useful for the analysis of diabetes and it allows us to address the mentioned complexities~~. First of all, we are able to investigate how self-reported diabetes and self-reported diabetes duration, capturing the temporal dimension of diabetes, are associated with labor market outcomes~~ : using individual level fixed effects analysis for the first time in this ~~area, in order to account for any time constant heterogeneity and to provide first evidence of the effects on wages and working hours in a~~ 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 . [1]. Further, we add to the current literature in exploring the role of undiagnosed diabetes, using novel and rich biomarker data ~~, a likely very important issue given the generally~~ - an issue of considerable importance in light of the large numbers of undiagnosed people (see Beagley et al. (2014)) that remained ~~unobserved~~ unaccounted for in all earlier studies relying on self-reported information ~~. This allows us to say something about~~ in assessing the labor market impact. Doing so sheds some light on the issue of measurement error ~~and about~~

[1] This is the first such evidence of the effect on wages and working hours in a middle-income country (MIC)

~~potential differences~~ as well as on potentially differential effects between diagnosed and undiagnosed diabetes. Finally, we assess and try to minimize the issue of inconsistent self-reporting of diabetes over time that could result in measurement error and again ~~has potentially affected~~ may potentially have biased the results presented in the earlier literature.

Our results using self-reported diabetes suggest a significant and economically important decrease in employment chances for people aware of their disease~~of~~ , by over 6 percentage points for men and women alike. We do not find strong evidence for any further effects of diabetes on wages or working hours, but show that if unobserved heterogeneity is unaccounted for, then the estimates are biased and suggest a spurious positive association of diabetes with wages. We further find that employment chances are reduced with each additional year since diagnosis by over one percentage point~~and~~ , with some evidence for an even larger effect per year after the initial 10 years.

Turning to the biomarker analysis we find that measurement error of self-reported diabetes leads to an upward biased estimate of the employment penalty compared to objectively measured diabetes including those diagnosed and undiagnosed. We further find no effect of undiagnosed diabetes on any labor market outcome, suggesting that adverse effects only occur to those with a diagnosis. We conclude that people aware of their diabetes differ from those unaware of the disease leading to the different outcomes for both groups and argue that other unobserved characteristics, such as those related to the added health information of a diagnosis may play an important role and need further exploration. Of course differences in other adverse health events related to diabetes not captured by the data likely also explain much of the difference between diagnosed and undiagnosed diabetes. Finally, we do not find that current diabetes severity as proxied by glycated hemoglobin (HbA1c) measurements is related to labor market outcomes.

In general, our results are consistent with the temporal way in which diabetes is affecting the health of people~~. They indicate that~~ : those with a diagnosis experience an important employment penalty once they are diagnosed, potentially due to the psychological effects of the diagnosis but also because diabetes is not diagnosed at the initial stage of the disease but after several years of asymptomatic but elevated blood glucose levels. A possible explanation would be that a diagnosis frequently happens after some time of living with elevated blood glucose levels shortly before or once symptoms start appearing~~which~~ . This then, potentially in conjunction with psychological effects of the diagnosis ~~, and possibly~~ and with possible discriminatory employer behavior ~~,~~ could reduce employment chances shortly after. In line with that we find that the adverse effects are even stronger several years after diagnosis likely due to the development of additional and more severe complications.

2 ~~Labor Outcomes and~~ Diabetes ~~literature~~and labor outcomes – previous evidence

Several studies have investigated the effects of diabetes on labor market outcomes. For the USA, Brown et al. (2005) estimated the impact ~~of the disease~~ on employment in 1996–1997 in an older population of Mexican Americans~~in the~~ , living in the USA close to the

Mexican border, using a bivariate probit model. They found diabetes to be endogenous for women but not for men. The results of the instrumental variable (IV) estimation suggested no significant effect on women which, compared to the adverse effect found in the probit model, indicated an overestimation of the effect for women when endogeneity was not accounted for. For men, the probit estimates showed a significant adverse effect of about 7 percentage points. For a similar population and using biomarker data, Brown I I I et al. (2011) looked at how diabetes management, inferred from measured HbA1c levels, affected employment chances and wages using cross-sectional data ~~in a mainly Mexican-American population in the US~~USA. They found a linear ~~adverse association of increasing~~ negative association between HbA1c levels ~~with~~ and both employment chances and ~~with~~ wages for men. ~~They~~ This particular study did, however, not investigate the effects of undiagnosed diabetes. Two other studies also ~~investigated~~ examined the effect of diabetes on employment and productivity for the USA ~~. Minor (2011) investigated~~ : Minor (2011) focused on the effect of diabetes on female employment, earnings, working hours and lost work days in the USA in ~~2006. The study found~~ 2006, finding diabetes to be endogenous and its effect underestimated if exogeneity was assumed. In the IV estimates, type 2 diabetes had a significant negative effect on female employment chances as well as yearly earnings but not on working hours. Both of these studies used a Heckman selection model to adjust for a possible selection bias in their estimates of productivity. However, ~~both studies do not discuss~~ neither of the studies discusses the choice of exclusion restrictions which are crucial ~~to identify~~ for identifying the selection equation and for the validity of the results~~, at least casting some doubt on the presented results~~. In a later study Minor (2013) investigated the relationship of diabetes duration and labor market outcomes, providing some evidence for a non-linear relationship of diabetes duration and employment chances shortly after diagnosis for men and after about ten years for women. . For ~~For~~ wages, no strong evidence ~~for~~ of any relationship was found. For Canada, Latif (2009) estimated the effect of the disease on employment probabilities again using and IV strategy similar to Brown et al. (2005). He found diabetes to be exogenous for females and 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) ~~investigated~~ analyzed the effects of diabetes on labor force participation using a multivariate endogeneous probit model. They found reduced labor market participation for males and females of 7.1 and 9 percentage points, respectively. They also found that if the endogeneity of diabetes is unaccounted for, then the effects are overestimated.

Only two studies exist for MIC. Liu and Zhu (2014) investigate the effect of a recent diabetes diagnosis on labor income in China exploiting a natural experiment for identification 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 cross-sectional data from the 2005 wave of the MxFLS, and found a significant ($p<0.01$) reduction in employment chances for males of a about 10 percentage points and for females of about 4.5 ($p<0.1$) percentage points, ~~albeit at a much lower statistical significance~~, using parental diabetes as an IV (Seuring et al., 2015b).

A recent systematic review of the economic cost of diabetes confirms that ~~the evidence from~~ specifically labor market impact evidence from low- and middle-income countries

(LMICs) remains scarce (Seuring et al., 2015a), while most studies, including those on high income countries, suffer from at least three key limitations. ~~First, many~~ (1) Many studies rely exclusively on cross-sectional data, unable to fully account for unobserved characteristics.

~~Second, many~~ (2) Many studies apply an IV estimation strategy, typically using parental diabetes as the instrument. The use of this instrument relies on the finding that type 2 diabetes has a genetic and heritable component 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 other genetically transferred traits, including those that impact labor outcomes directly, ~~like~~ as for instance unobserved abilities.

~~The identification~~ This traditional identification strategy also abstracts from intra-household or intergenerational labor supply effects. 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 ~~inversely,~~ – to the contrary – may have to increase labor supply to compensate for lost income (Seuring et al., 2015b). Instrumentation can further be problematic if the instruments are weak as their use may result in biased estimates in finite samples and generally produce large standard errors (Bound et al., 1995).

~~And third, the~~ (3) The results are likely biased due to reporting error as most studies have relied on self-reported diabetes measures. Self-reported data can suffer from non-classical measurement error due to systematic misreporting which has been shown to cause estimates of economic ~~outcomes~~ impacts to be biased and potentially overstated (Cawley et al., 2015; O'Neill and Sweetman, 2013; Perks, 2015).

To address these limitations, this paper applies an alternative panel estimation strategy using individual level fixed effects (FE), potentially better accounting for unobserved heterogeneity without having to rely on the strong assumptions of an IV strategy. Using the MxFLS we are able to investigate how self-reported diabetes and diabetes duration are associated with labor market outcomes using FE and investigate the heterogeneity of the effects across different employment types, i.e. non-agricultural employment, agricultural employment and self-employment, as ill health may have distinct effects than across these activities. Finally, to investigate the issue of measurment error in self-reported data and to explore how labor outcomes are related with undiagnosed diabetes, we estimate models using diabetes biomarker data.

3 Data

The dataset used for the empirical analysis is the Mexican Family Life Survey (MxFLS), a nationally representative, longitudinal household survey with three waves conducted in 2002, 2005–2006 and ~~2009–2011, respectively~~2009–2012. All household members aged 15 and above were interviewed, ~~and~~ covering information on a wide range of social, demographic, economic characteristics and health behaviours of the individuals and their families ~~was collected~~ (Rubalcava and Teruel, 2013). Apart from self-reported diabetes information ~~throughout~~ that is available in all rounds of the survey, we ~~more~~ also specifically use information provided exclusively in the most recent wave, i.e. information on the self-reported year of diagnosis as well as biomarker data including HbA1c levels for a large proportion of respondents. For our main analysis using self-reported diabetes we exploit all three waves in order to take advantage of the large amount of observations (N=49,323) and the panel structure of the data. Our variable of interest for this analysis is self-reported diabetes based on the survey question: "Have you ever been diagnosed with diabetes?". The response to this question likely suffers from measurement error depending on how long ago a diagnosis might have been made and, most importantly, if the respondent is aware that he has the disease. We try to correct the self-reported diabetes variable for inconsistencies in reporting over time given that we have repeated measures for the same individual. This should add more consistency as this strategy should correct for errors due to coding errors or the misinterpretation of the diabetes question. The exact way we dealt with this issue is detailed in the appendix. Of course, this strategy will not deal with the likely more important problem of undiagnosed diabetes. In order to investigate how such measurement error may affect estimates of the labor market impact of diabetes we use the information from a subsample containing over 6000 respondents (everybody $\geq$ 45 and a random subsample of age 15–44 (Crimmins et al., 2015)) of the ~~2009-2011~~ 2009-2012 wave, allowing us to use objectively measured diabetes by identifying those with undiagnosed diabetes. For the entire paper the used samples are restricted to the working age population of Mexicans ~~age~~ aged 15 to ~~64.~~ 64 years. To prevent that pregnant women might bias 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).

As can be observed in Figure 1, unweighted self-reported diabetes prevalence in the MxFLS has increased from about 6 percent in 2002 to 7.1 percent in 2009 for females and from about 4.3 to 5.7 percent for males. This is still well below the diabetes prevalence estimates published by other institutions, including the International Diabetes Federation (IDF), whose most recent estimates for 2014 indicate a prevalence of about 12 percent (which amounts to about 9 million Mexicans) for those aged between 20 and 79 (International Diabetes Federation, 2014). However, this difference in self-reported and undiagnosed diabetes should be explained, in addition to the somewhat different age group considered, by the large share of undiagnosed people in the sample. Barquera et al. (2013) show that while overall prevalence in Mexico increased from 6.7 percent in 1994 to 7.5 percent in 2000 and 14.4 percent in 2006, only 4.6, 5.8 and 7.5 percent, respectively, had been previously diagnosed. Accordingly, the prevalence based on diabetes self-reports in our sample is more or less in line with other existing data from Mexico. Further ~~, we will~~ ~~,~~ below we will specifically investigate the actual extent of undiagnosed diabetes~~further below~~, using the MxFLS data.

For the pooled data of all three waves (Table 1), diabetes was self-reported by 5 percent of men and 6.2 percent of women. Most of the respondents in the sample either live in rural or in large urbanized areas. Looking at our outcome variables, 86 percent of men report some form of employment compared to 36 percent of women. Interestingly, men do not report higher hourly wages compared to 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 also have lower educational ~~attainments~~ attainment on average. ~~Looking at diabetes, women have a higher~~

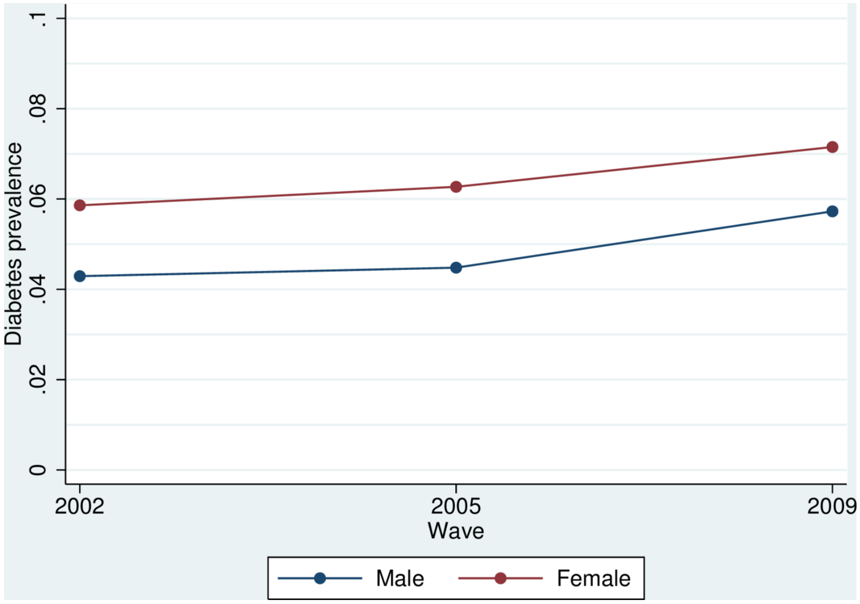


Figure 1: Self-reported diabetes prevalence in MxFLS

~~self-reported diabetes prevalence.~~

~~When looking at~~ Table 2 presents the descriptive statistics from the subsample of the third wave (2009-2012), which is the one containing biomarker data~~in Table 2, it first has to be noted~~. It is important to note that in this subsample respondents are somewhat older on average than in the pooled sample, which can be explained by the fact that everybody above the age of 44 was included ~~and~~ in the biomarker measurement, together with only a random subsample of those ~~younger~~ aged 44 or below Crimmins et al. (2015). Also, self-reported diabetes is ~~much~~ considerably higher than in the pooled sample as well as in the full sample of wave 3. Regarding the other control and outcome variables, the sample is ~~pretty~~ fairly similar to the pooled sample. The added value of this subsample is obviously the biomarker information regarding diabetes. Apparently a large share of the subsample has an HbA1c indicative of diabetes.[2] In ~~the sample used in this study~~ our Mexican sample over 18 percent of males and females ~~are undiagnosed~~ result as having undiagnosed diabetes. These first descriptive results suggest that using self-reported diabetes as a measure for diabetes in Mexico can lead to a strong underestimate of the true diabetes population which potentially has consequences for the interpretation of empirical results of economic studies using self-reported diabetes data. In the ~~ensuing~~ following sections we will therefore try to shed some light on how taking the large undiagnosed population into account could affect the impact estimates of diabetes ~~in relation to~~ on labor outcomes.

[2] In one of the first analyzes of these new biomarker data Frankenberg et al. (2015) show that the rates in Mexico of elevated ~~acHbA1c~~HbA1c levels are very high when compared to HbA1c data from similar surveys in the USA and China~~and note~~. This study also notes that "The extremely high levels of elevated ~~HbA1c~~ HbA1c among Mexicans adults (...) is profoundly troubling." (Frankenberg et al., 2015, p.18) .

Table 1: Pooled sample characteristics (2002, 2005-2006, ~~2009-2011~~2009-2012)

	Males				Females			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Dependent variables								
Employed	0.861	0.346	0.000	1.000	0.359	0.480	0.000	1.000
Hourly wage	42.165	483.712	1.010	55384.590	41.260	168.405	1.007	8803.946
Log hourly wage	3.165	0.882	0.010	10.922	3.143	0.956	0.007	9.083
Usual weekly working hours	46.819	16.758	4.000	112.000	38.957	18.908	4.000	112.000
Agricultural worker	0.222	0.416	0.000	1.000	0.043	0.202	0.000	1.000
Self-employed	0.189	0.392	0.000	1.000	0.277	0.448	0.000	1.000
Non-agricultural worker or employee	0.589	0.492	0.000	1.000	0.680	0.467	0.000	1.000
Diabetes variables								
Diagnosed diabetes	0.050	0.219	0.000	1.000	0.063	0.243	0.000	1.000
Diabetes duration	0.359	2.044	0.000	38.000	0.416	2.362	0.000	65.000
Other selected control variables								
Demographics								
Age of respondent	37.151	13.382	15.000	64.000	37.036	13.063	15.000	64.000
Living in rural area	0.441	0.497	0.000	1.000	0.433	0.495	0.000	1.000
Married	0.549	0.498	0.000	1.000	0.536	0.499	0.000	1.000
Number of children (<6) in household	1.482	1.446	0.000	11.000	1.577	1.478	0.000	13.000
Indigenous group	0.189	0.391	0.000	1.000	0.183	0.387	0.000	1.000
Education								
Secondary	0.302	0.459	0.000	1.000	0.300	0.458	0.000	1.000
High school	0.154	0.361	0.000	1.000	0.132	0.338	0.000	1.000
Higher education	0.113	0.317	0.000	1.000	0.091	0.287	0.000	1.000
Observations	21739				28174			

Table 2: Biomarker sample characteristics(2009-2012)

	Males				Females			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Dependent variables								
Employed	0.860	0.347	0.000	1.000	0.339	0.474	0.000	1.000
Hourly wage	36.271	53.612	1.128	1038.462	35.470	44.000	1.274	461.538
Log hourly wage	3.166	0.859	0.121	6.945	3.143	0.915	0.242	6.135
Usual weekly working hours	46.011	16.860	4.000	112.000	38.136	19.673	4.000	97.000
Agricultural worker	0.252	0.434	0.000	1.000	0.035	0.185	0.000	1.000
Self-employed	0.213	0.410	0.000	1.000	0.321	0.467	0.000	1.000
Non-agricultural worker or employee	0.535	0.499	0.000	1.000	0.644	0.479	0.000	1.000
Diabetes variables								
Glycated hemoglobin (HbA1c)	6.459	1.888	4.000	14.000	6.565	2.020	4.000	14.000
HbA1c $\geq$ 6.5%	0.262	0.440	0.000	1.000	0.273	0.446	0.000	1.000
Undiagnosed diabetes	0.184	0.387	0.000	1.000	0.182	0.386	0.000	1.000
Diagnosed diabetes	0.094	0.291	0.000	1.000	0.115	0.319	0.000	1.000
Diabetes duration	0.689	2.850	0.000	38.000	0.863	3.395	0.000	40.000
Other selected control variables								
Demographics								
Age of respondent	42.749	14.287	15.000	64.000	42.405	14.084	15.000	64.000
Rural village of <2,500	0.505	0.500	0.000	1.000	0.463	0.499	0.000	1.000
Married	0.603	0.489	0.000	1.000	0.559	0.497	0.000	1.000
Number of children (<6) in household	1.185	1.288	0.000	8.000	1.229	1.320	0.000	11.000
Indigenous group	0.188	0.391	0.000	1.000	0.180	0.384	0.000	1.000
Education								
Secondary	0.264	0.441	0.000	1.000	0.262	0.440	0.000	1.000
High school	0.138	0.344	0.000	1.000	0.123	0.329	0.000	1.000
Higher education	0.117	0.322	0.000	1.000	0.088	0.283	0.000	1.000
Observations	2792				3695			

4 Estimation ~~Strategy~~strategy

The conceptual framework of our study is based on the work of Strauss and Thomas (1998), who specify the following model ~~to~~ of the relationship ~~of~~ between health and labor market outcomes, i.e. labor force participation and labor supply conditional on health and wages.

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

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 human capital; S is education; A is a vector of demographic characteristics; B is the family background of the individual; I is local community infrastructure; α is an array of unobservables such as ability and e_w is 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 indirectly by changing wages and directly by impacting the ability to work.

There are several ways diabetes may affect H. First of all, diabetes can deteriorate health if it remains untreated with the adverse effects increasing over time. Second, a diagnosis of diabetes and ensuing treatment may lead to better health compared to the undiagnosed state~~, however~~. However, compared to healthy people even those with diabetes receiving treatment may still have worse health outcomes. Third, there is also evidence that the diagnosis itself may affect one's own health perception and can 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 ~~though~~ through changes in diet and physical activity patterns. Finally, measurement error e_w may be an 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.

In the following section we ~~will show how we will~~ describe how we deal with these issues in our estimation strategy.

4.1 Panel data estimation strategy: Self-reported Diabetes

We investigate the relationship of self-reported diabetes and three labor market outcomes, i.e. employment chances, log hourly wages and normal weekly working hours using a fixed effects model. While using individual level FE does not fully identify a causal relationship, as other forms of heterogeneity could affect our estimates (i.e. via time-variant unobserved heterogeneity or simultaneity) it does improve on the degree of causal inference, compared to a simple cross-sectional analysis~~, and in~~. In particular it does allow to control 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. For comparison purposes we also present the results of the respective random effects models that assume no unobserved heterogeneity.

We estimate a two-part model where the model for employment chances takes the following form:

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

where Y_{it} is a binary variable taking a value of 1 if respondent i reports being employed 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, X_{it} is a vector of control variables, c_i represents an individual fixed effect, γ_t represents a year fixed effect, and u_{it} is the error term.

For the relationship of self-reported diabetes with log hourly wages and weekly 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 log hourly wage was calculated by adding up the reported monthly income from the first and potential second job and dividing it by 4.33 which corresponds to the average number of weeks per month in a year with 52 weeks. This gave us the average earnings per week which was then divided by the usual weekly working hours to arrive at an hourly wage estimate[3] Finally, we adjusted the calculated wage for inflation from the year of the interview up to 2013. Working hours were calculated summing up the self-reported usual working hours of the first and potential second job. We dropped 39 observations where the sum of working hours exceeded 112 hours per week, i.e. more than 16 hours of work per day on each day of the week as we deemed these reports to be unrealistic. Due to a considerable number of missing values the sample used for the wage estimation is considerably smaller than the model for working hours.

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 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 employment chances. A variable capturing the number of children residing in the household below the age of 18 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 that changes in household wealth might have on diabetes and employment chances, we use standard principal component analysis of multiple indicators of household assets and housing conditions to create an indicator for household wealth (Filmer and Pritchett, 2001). The components used for the construction of the wealth index are detailed in Seuring et al. (2015b). Finally, calender year fixed effects are included to capture the effects of increasing age and any macroeconomic shocks over time.

[3] ~~labor~~ Labor income was either reported as the total amount for the whole month or if possible more detailed containing information on the monthly wage, income from piecework, tips, extra hours, meals, housing, transport, medical benefits and other earnings. Over 80 percent of respondents reported the total amount instead of a detailed amount. Respondents were also asked for their yearly income and we used that information to arrive at an hourly wage if information for monthly labor income was missing.

Additionally, we estimate models using a self-reported measure of the years since diagnosis ~~instead~~ in order to explore the ~~temporal nature of diabetes . First, using~~ role of the duration of diabetes in affecting labor outcomes. We start by following a linear specification indicating the association of labor market outcomes with each additional year since diagnosis, ~~using the following specification~~as follows:

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

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

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

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

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 $n = 1, 2, \ldots, N$. We choose three nodes that - based on visual inspection (see Figures 2, 3 and 4 in the result section) - best captured any possible non-linearity in the relationship ~~of~~ between diabetes duration and ~~our dependent variables~~labor outcomes. These are located at four, eleven and twenty 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 eleven years it is conceivable that many of the debilitating complications of diabetes would appear that could deteriorate health and lead to adverse effects on ~~our~~ labor market outcomes. The coefficientδ_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. In order to have zero years representing people without a self-reported diagnosis, those that reported a diagnosis in the year of the interview were counted as 'one year since diagnosis'. Accordingly, 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.

4.2 Cross-section data estimation strategy: biomarker and self-reported data

Self-reported diabetes only captures part of the diabetes population as many individuals remain undiagnosed. Estimations based on self-reports may therefore suffer from a selection bias.

The issue of measurement error can be depicted in a simple theoretical model. Assuming that the true model of the effect of diabetes on labor market outcomes is $y^* = X^*\beta + \epsilon$ where y^* and ϵ are scalars and X^* and β are vectors. Because we do not observe the true values of y^* and X^* we have to use reported measures that contain errors: $X = X^* + u$ and $y = y^* + v$. In the case of diabetes as a right-hand side variable, measurement error cannot be classical as it is negatively correlated with the true diabetes status. Hence, the direction of the bias is unknown and cannot be assumed to be attenuating~~making inference based on self-reported measures without some information about the direction of a potential bias difficult~~.

Self-report of diabetes can lead to errors due to two main factors:

1. **Systematic overreporting of diabetes:** people without diabetes could intentionally report a diabetes diagnosis with a view to justifying some other adverse event or status in their life (e.g. being unemployed).

2. **Systematic underreporting of diabetes:** it is conceivable that people with diabetes ~~are ashamed of their disease~~ feel bad about suffering from the condition and therefore do not report it. Further, diabetes often remains undiagnosed for long periods of time or is not diagnosed at all, potentially leaving many people unaware of the disease which hence results in the non-report of the disease.

Overreporting could attenuate the coefficient of diabetes as those falsely reporting a diabetes diagnosis experience no adverse health effects of diabetes that could affect labor outcomes. However, if some of those misreports were made in order to justify some other adverse event, e.g. current unemployment or other health problems, then this could potentially lead to an overestimation of the true effect of diabetes on employment. Underreporing due to a non-diagnosis could cause either an overestimation or attenuation bias: it would lead to an overestimation if people with undiagnosed diabetes were generally healthier and therefore more likely to have positive labor market outcomes than people with diagnosed diabetes and would therefore have attenuated the effect of diabetes if they were observed in the data. Yet, because they are unobserved, the effect of diabetes using self-reports may be overstated. However, if those with undiagnosed diabetes were very similar in terms of health to those with diagnosed diabetes and consequently experienced similar health problems that adversely affected their labor market outcomes, then this would lead to an attenuation of the diabetes coefficient as the control group without observed diabetes, i.e. including people with undiagnosed diabetes, were on average less healthy and have worse labor market outcomes than it would have had, if all diabetes cases had been observed.

In this context, however, an important additional issue arises with the potential effect that being informed via a diabetes diagnosis might have on labor market outcomes. In assessing the labor market effects of diabetes with the help of self-reported diabetes as our diabetes indicator, we need to be aware that we could be ~~measuring something else then~~ capturing something other than the pure health effects of diabetes. A diabetes diagnosis is likely to also affect an individual's psychology and health behaviour which in turn could have its own effects on economic outcomes. One study found a diabetes diagnosis and subsequent treatment to increase the odds of psychological problems, including depression and anxiety (Thoolen et al., 2006). Other research has also shown that anxiety and depression increase with the number of diabetes symptoms people with diabetes self-report after a diagnosis Paddison et al. (2011). Interestingly, similar results have not been found for people with undiagnosed diabetes (Nouwen et al., 2011). When looking at economic studies of the effect of a diabetes diagnosis, Liu and Zhu (2014) found that receiving a diabetes diagnosis considerably reduced labor income in Chinese employees shortly after their diagnosis. Similarly, others have shown that a hypertension diagnosis can affect health behaviours, ~~i.e.a reduce~~ e.g. in the form of a reduction in fat intake (Zhao et al., 2013). Similar effects have also been found for the ~~US~~USA, in that people receiving a diabetes diagnosis changed their health behaviours favourably, albeit only over the short term (Slade, 2012). It is not difficult to imagine that such changes in health behaviours resulting from a diabetes diagnosis might also translate into changes in employment chances or productivity, on top of a potential effects resulting from the diabetes-related health changes. Hence, the labor market effects we measure with self-reported diabetes are likely different from those that we would measure based on a purely medical assessment of diabetes obtained from blood tests, not only due to measurement error but also due to the additional information effects of the diagnosis itself.

We use the biomarker data in wave three to explore the relationship of measured diabetes, including undiagnosed diabetes with labor outcomes and compare it to estimates using self-reported diabetes. The biomarker data also allows us to look at diabetes severity, as measured by HbA1c values. Since this data is only available for one wave - the last wave ~~,~~ - we focus on a cross-section analysis. Moreover, as mentioned above, biomarkers were only taken from about one-third of the initial representative sample, which leaves us with information on 6994 survey participants. Therefore the analysis cannot be directly compared to the panel-based results in this paper. Nonetheless, it allows for a first exploration of the relationships of undiagnosed diabetes as well as disease severity with labor market outcomes. It further allows us to explore some of the heterogeneity that may exist within the diabetes population in terms of how well different people manage their diabetes condition. We first estimate a model to investigate the association of objectively measured diabetes (HbA1c $\geq 6.5\%$) with labor market outcomes taking the following form:

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

where $\beta_1 Dobj_i$ is equal to 1 if HbA1c $\geq 6.5\%$. In a further step we estimate the following equation

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

to investigate how the associations differ between people with diagnosed and undiagnosed diabetes. $\beta_1 Dsr_i$ identifies those diagnosed and is equal to 1 if the person self-reported a diabetes diagnosis. $\beta_1 Dud_i$ identifies those undiagnosed and is equal to 1 if the person did not self-report a diabetes diagnosis but has an HbA1c $\geq 6.5\%$.

Before moving on to the results ~~we want to stress,~~ it bears emphasising that despite our efforts to reduce any bias in our estimates we do not make any causal claims as the used methods cannot prove causation. While the fixed effects model accounts for time invariant unobserved confounders, it is still possible that the estimates are biased due to important time variant heterogeneity we did not account for, or due to labor market outcomes simultaneously affecting the chances to develop or being diagnosed with diabetes. For example, changes in lifestyle due to employment or unemployment could affect the probabilities of developing diabetes which then in turn could affect labor market outcomes. Another pathway could be high stress levels at work that have been shown to have a positive association with diabetes incidence for obese women and a negative association for non-obese men (Heraclides et al., 2012; Eriksson et al., 2013). However, part of the possible effects of work stress on diabetes should be accounted for by our fixed effects model. Genetics play an important role for the development of type 2 diabetes and it is likely that people with a genetic predisposition are more likely to develop diabetes due to stress. Further, coping mechanism related to stress are also likely to differ between persons depending on their genetics and coping methods learned early in life (Schneiderman et al., 2005), so that our fixed effect technique should account for these time invariant factors~~and reduce~~ , reducing the possible bias. Also the literature on the effects of employment status on diabetes so far has not found strong adverse effects of being laid off on diabetes self-reports (Bergemann et al., 2011; Schaller and Stevens, 2015), ~~albeit only for~~ although this has only been researchers in a high-income ~~countries~~country context so far.

5 Results

5.1 Panel data analysis

Self-reported diabetes

Table 3 presents the estimation results of the FE model 2. The results indicate significant and substantial reductions in employment chances for people with self-reported diabetes. The effects are surprisingly similar between males and females showing a reduction in employment chances of around 6.5 percentage points. The adverse effects on employment confirm findings of an earlier study on Mexico by Seuring et al. (2015b) albeit ~~we now find a~~ with a now stronger indication of an effect for females. This earlier study used an IV approach and found diabetes to be exogenous in the determination of employment chances, a finding which is supported by the small difference between our random effects (RE) and FE estimates for employment chances.

Table 3: Self-reported diabetes and labor market outcomes

	Employment			Log hourly wages			Weekly working hours		
	(1) Pooled	(2) Males	(3) Females	(4) Pooled	(5) Males	(6) Females	(7) Pooled	(8) Males	(9) Females
Fixed Effects									
Diagnosed diabetes	−.066***	−.065**	−.065***	0.033	0.009	0.100	−.853	−.225	−2.346
	(.018)	(.025)	(.024)	(.064)	(.066)	(.158)	(1.269)	(1.458)	(2.518)
Random Effects									
Diagnosed diabetes	−.068***	−.071***	−.057***	0.067**	0.110***	−.017	−.841	−.575	−1.742
	(.009)	(.013)	(.012)	(.031)	(.037)	(.058)	(.604)	(.704)	(1.134)
N	49323	21801	27522	20974	13925	7049	26882	17801	9081

Robust standard errors in parentheses
Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6
wealth, age and calender year fixed effects
The random effects model additionally controls for initial age when entering the survey, being indigenous and gender
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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

While the adverse relationship between self-reported diabetes and employment chances appears to be quite strong, we find no evidence for any relationship with wages or working hours using the FE specification. Further, comparing the results with the RE model, we see that there are no stark differences for employment chances but that the FE model appears to correct for a positive bias in the effect of diabetes on wages. The RE model suggests – counter-intuitively – that men with diabetes earn more than men without diabetes~~which is an unintuitive finding~~. Controlling for fixed unobservables, however, the positive effect disappears and becomes insignificant suggesting that assuming exogeneity of diabetes leads to biased estimates of wages. A potential explanation is that men with a comparatively higher innate ability are able to earn higher wages and at the same time to better manage their diabetes preventing them from dropping out of the labor market due to health reasons. They may also be more valuable to their employers preventing them from being dismissed even if the employer would normally discriminate against people with diabetes.

To investigate whether there are differences in wages and working hours between different types of work, we estimate a model including interaction terms between self-reported diabetes and agricultural employment and between self-reported diabetes and self-employment, respectively, using non-agricultural employment as the ~~base~~benchmark. There would be reason to expect the relationship between diabetes and labor market outcomes to differ with the type of work, as the health consequences of diabetes could affect the productivity of workers by reducing their physical abilities. Accordingly, people with diabetes working in an agricultural job that requires strenuous, physical efforts to work the land could see their productivity ~~to be~~ being more adversely affected by diabetes than those working in an office, hence being less physically active.

Table 4: Relationship of self-reported diabetes by type of work and wages and working hours

	Log monthly wages (FE)			Monthly work hours (FE)		
	(1) Pooled	(2) Males	(3) Females	(4) Pooled	(5) Males	(6) Females
Fixed Effects						
Reference category: No diabetes and non-agricultural employee						
~~(.043) (.044) (.190) (.764) (.802) (2.713)~~						
No diabetes and agricultural worker	$-.111^{***}$	$-.092^{**}$	$-.294$	-3.934^{***}	-3.697^{***}	-4.176
	(.043)	(.044)	(.190)	(.764)	(.802)	(2.713)
No diabetes and self-employed	$-.029$	0.015	$-.149^{*}$	-2.503^{***}	-1.817^{**}	-4.318^{***}
	(.041)	(.045)	(.087)	(.643)	(.706)	(1.419)
Diagnosed diabetes and non-agricultural employee	0.065	0.060	0.089	$-.006$	0.486	-1.077
	(.071)	(.074)	(.169)	(1.315)	(1.585)	(2.262)
Diagnosed diabetes and agricultural worker	$-.209$	$-.200$	$-.394$	-5.251^{*}	-4.950^{*}	-5.911
	(.191)	(.198)	(.373)	(2.797)	(2.879)	(15.395)
Diagnosed diabetes and self-employed	$-.036$	$-.099$	0.128	$-.371$	1.078	-3.963
	(.161)	(.188)	(.322)	(2.227)	(2.484)	(4.740)
R2 within	0.022	0.022	0.031	0.010	0.012	0.018
Random Effects						
Reference category: No diabetes and non-agricultural employee						
No diabetes and agricultural worker	$-.226^{***}$	$-.236^{***}$	$-.220^{***}$	-3.536^{***}	-3.429^{***}	-2.512^{**}
	(.020)	(.021)	(.063)	(.352)	(.377)	(1.002)
No diabetes and self-employed	$-.038^{*}$	0.039	$-.154^{***}$	-2.911^{***}	-1.187^{***}	-4.577^{***}
	(.022)	(.026)	(.039)	(.344)	(.408)	(.608)
Diagnosed diabetes and non-agricultural employee	0.090^{***}	0.134^{***}	0.029	-1.049	0.040	-3.246^{***}
	(.034)	(.041)	(.063)	(.710)	(.875)	(1.218)
Diagnosed diabetes and agricultural worker	$-.132$	$-.178$	0.386	-2.941^{*}	-4.287^{***}	1.749
	(.112)	(.116)	(.514)	(1.588)	(1.660)	(5.955)
Diagnosed diabetes and self-employed	$-.119$	$-.043$	$-.204$	1.397	0.009	3.391
	(.081)	(.100)	(.133)	(1.324)	(1.596)	(2.290)
R2	0.213	0.201	0.246	0.067	0.026	0.049
N	20974	13925	7049	26882	17801	9081

Robust standard errors in parentheses
Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6, wealth, health insurance status, age and calender year fixed effects
The random effects model additionally controls for initial age when entering the survey, being indigenous and gender
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As Table 4 shows, male agricultural workers have lower wages generally, but the relationship with diabetes remains unaffected by the type of work. The only statistically more relevant relationship that we find relates to working hours of agricultural workers with diabetes, who appear to work about ~~6~~ five hours less than non-agricultural workers without self-reported diabetes. 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, suggesting that the effect of diabetes on working hours does not vary significantly by type of work. When stratified by gender, standard errors become larger, which should be explained by the small number of observations of females with diabetes reporting working hours in agricultural employment (n=10) compared to men (n=139). Overall, we find no evidence for an association of self-reported diabetes and wages or working hours.

Finally, we explore if diabetes affects the selection into employment according to the different types of work. We refrain from using a fixed effects multinomial logit model due to the problems surrounding the calculation of interpretable marginal effects (Pforr, 2014). We instead estimate fixed effects 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 dependent variable.[4] The results in table 5 indicate that for men diabetes mainly appears to be associated with reduced employment chances for those self-employed~~and~~ , while for women an adverse association is found for those in agricultural and self-employment. This could suggest that having diabetes drives people out of self-employment and agricultural jobs, potentially because these jobs are more physically demanding than other forms of work and likely also because they provide less protection in terms of insurance and formal employment. We interpret the fact that the coefficient signs are negative for all types of work as an indication that people with diabetes predominantly become unemployed rather than ~~to sort~~ sorting into other types of work that would be better suited for them given their health status[DOES THIS INTERPRETATION MAKE SENSE?].

Table 5: Relationship of self-reported diabetes with selection into types of work

	Pooled			Males			Females		
	(1) Non-agric.	(2) Agric.	(3) Self-employed	(4) Non-agric.	(5) Agric.	(6) Self-employed	(7) Non-agric.	(8) Agric.	(9) Self-employed
Fixed Effects									
Diagnosed diabetes=1	−.015	−.017*	−.039***	−.021	−.009	−.050*	−.009	−.022***	−.033*
	(.016)	(.010)	(.015)	(.029)	(.021)	(.026)	(.018)	(.009)	(.018)
Random Effects									
Diagnosed diabetes=1	−.043***	−.044***	0.003	−.053***	−.066***	0.033**	−.042***	−.012***	−.016*
	(.009)	(.005)	(.008)	(.017)	(.011)	(.015)	(.010)	(.002)	(.009)
N	47330	47330	47330	20730	20730	20730	26600	26600	26600

Robust standard errors in parentheses
Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6, wealth, age and calender year fixed effects
The random effects model additionally controls for initial age when entering the survey, being indigenous and gender
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Self-reported diabetes duration

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 it could be that the effect increases linearly as the years go by. ~~However~~On the other hand, non-linear relationships are also ~~imaginable~~ plausible as psychological effects of the immediate diagnosis as well as the likely health problems that led to the diagnosis could affect labor market outcomes immediately after a person has been diagnosed. ~~They~~ The effects could then level off later as people with diabetes develop their skills in managing and living with the disease. ~~Contrary, if~~ If, however, diabetes management is unsuccessful, a longer disease duration likely leads to significant diabetes complications so that the adverse labor market effects of diabetes may appear at a later disease stage.

In order to obtain a first idea of the non parametric relationship we use kernel-weighted local polynomial regression to graph the relationship between our outcome variables and

[4]For example, "non-agricultural employment" is coded as 1 if the respondent is engaged in agricultural work and is coded as 0 for those unemployed, in non-agricultural employment or self-employed. This is repeated for the dummy variables "agricultural employment" and "self-employed".

diabetes duration. As Figure 2 shows, the relationship of diabetes duration and employment chances seems to be more or less linear for the pooled sample, showing a steady decline, though less so when separated by gender. We find a first decline in employment probabilities for men after about seven years which continues to about twenty years post-diagnosis. For women, a first drop off occurs right after diagnosis and thereafter no consistent pattern can be observed. However, the displayed relationships after twenty years suffer from large standard errors (not shown here) as sample size is reduced, considerably limiting the interpretation of the relationship for those with a very long disease duration. A similar analysis for wages and working hours shows somewhat more erratic relationships, especially after the first twelve years (see figures 3 and 4). In an effort to best capture any non-linearities for the pooled and gender stratifies samples, we created the following splines to capture the immediate, intermediate and long-term relationships (0–4, 4–11, 11–20 and 20+).

Figure 2: Kernel-weighted local polynomial regression of employment status on diabetes duration

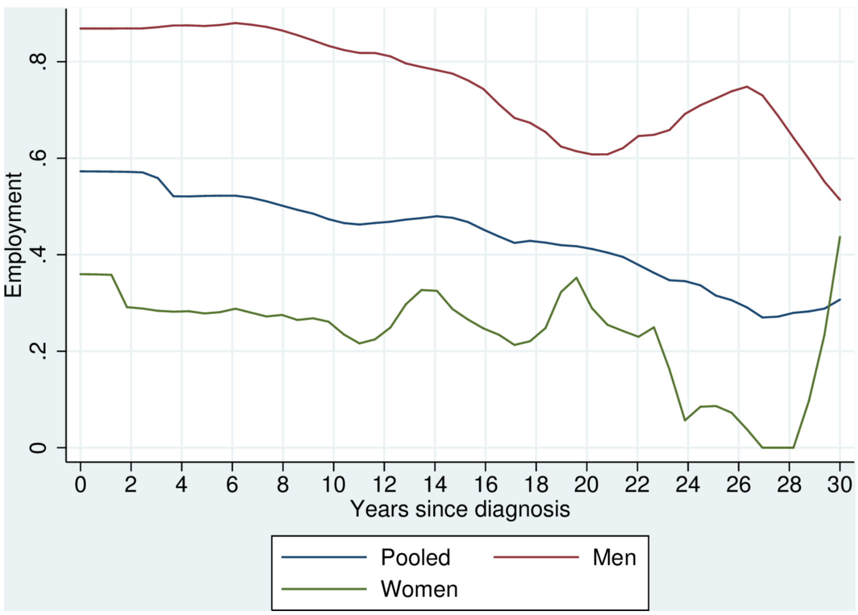


Figure 3: Kernel-weighted local polynomial regression of log hourly wages on diabetes duration

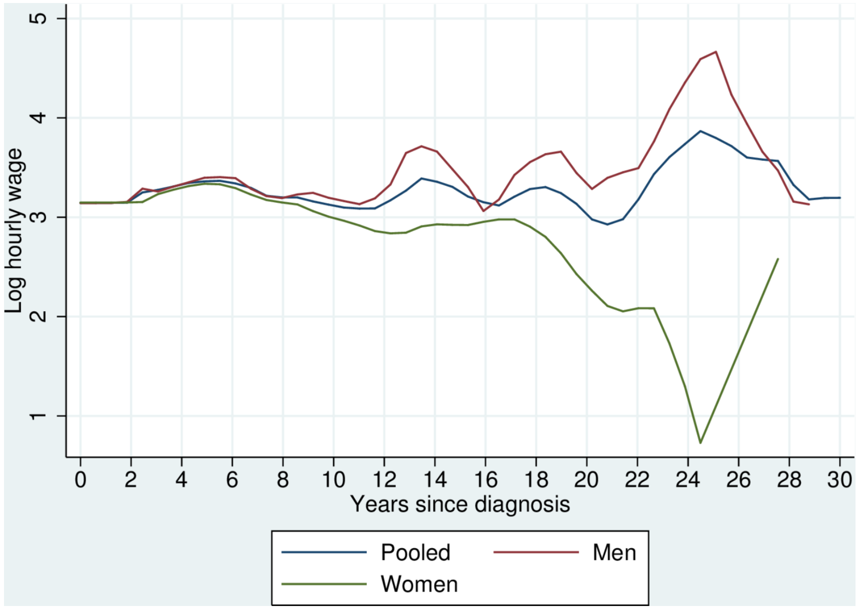
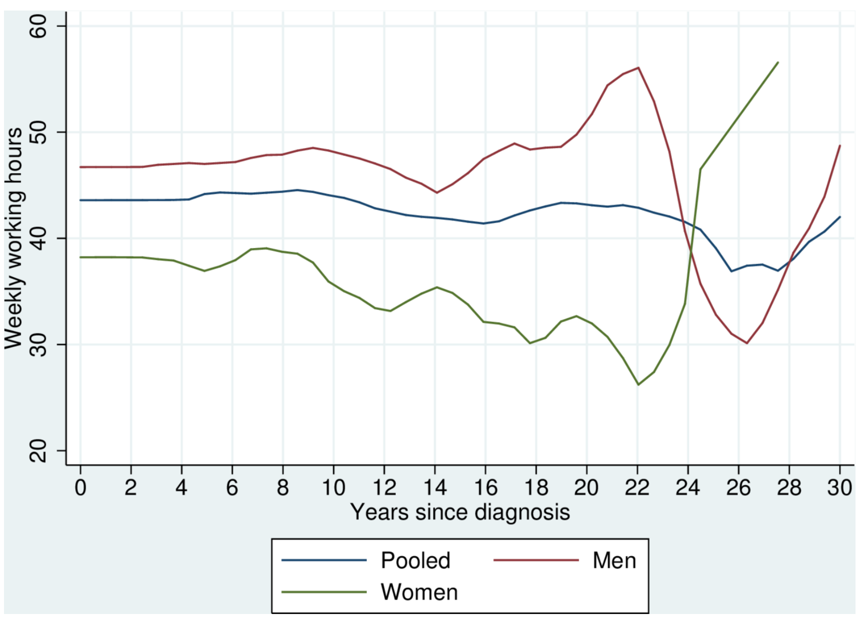


Figure 4: Kernel-weighted local polynomial regression of working hours on diabetes duration



First, models with a linear diabetes duration specification are presented in table 6. The results indicate a reduction of about 1.8 percentage points for men with each additional year since diagnosis. For women the coefficient shows a reduction of about one percentage point per year, though the association is not as strong. Interestingly, we find a reduction in female wages of about seven percentage points per year with diabetes but no effects for males and for the working hour models.

Table 6: Relationship of self-reported years since diagnosis and labor market outcomes

	Employment			Log monthly wages			Monthly work hours		
	(1) Pooled	(2) Males	(3) Females	(4) Pooled	(5) Males	(6) Females	(7) Pooled	(8) Males	(9) Females
Fixed Effects									
Diabetes duration	$-.013^{***}$	$-.018^{***}$	$-.011^{**}$	$-.028^{*}$	$-.018$	$-.069^{**}$	0.125	0.167	0.094
	(.004)	(.006)	(.005)	(.016)	(.018)	(.029)	(.316)	(.353)	(.679)
Random Effects									
Diabetes duration	$-.013^{***}$	$-.011^{***}$	$-.008^{***}$	$-.002$	0.007	$-.013$	$-.086$	$-.064$	$-.136$
	(.004)	(.002)	(.001)	(.004)	(.005)	(.009)	(.083)	(.103)	(.129)
N	38398	16073	22325	16288	10604	5684	20691	13375	7316

Robust standard errors in parentheses
Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6
wealth, age and calender year fixed effects
The random effects model additionally controls for initial age when entering the survey, being indigenous and gender
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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Using the non-linear specification with splines, we find that for the pooled sample employment chances are mainly reduced for those reporting a diagnosis more than eleven years ago, with each additional year reducing employment chances by about two percentage points (Table 7). When stratified by gender, the negative association persists for males and females, albeit much ~~weaker~~ more weakly for the latter. Interestingly, using splines we find a relatively strong reduction of wages 5-11 years after diagnosis for women. We also find some significant associations for those with more than 20 years of diabetes~~, however~~. However, these estimates are likely spurious due to the considerably reduced number of observations, particularly for wages and working hours.

Overall, the duration results are somewhat different to those found by Minor (2013) for the USA, who does not find evidence for a linear relationship of diabetes duration with employment chances. Our results from the non-linear analysis are not directly comparable as Minor (2013) uses dummy variables to identify different diabetes durations. Nonetheless, they are similar in that he does not find an immediate effect of a diabetes diagnosis on employment chances. However, the effects still appear at an earlier point than in our analysis and are in general much larger. However, apart from not using splines he does not use individual fixed effects so that his results might suffer from bias leading to the larger effect sizes~~or~~. It may of course also be that the effects of diabetes in the USA are simply ~~much larger~~ larger than in Mexico.

As a robustness check and due to the reduced number of observations for twenty or more years with diabetes we estimated the linear model again excluding those above nineteen years of diabetes to investigate if the effects were driven by these rare observations. These results are available on request and indicate marginally smaller effect sizes, but do not change the qualitative interpretation of the effects.

Generally, the results so far have revealed an important adverse association of self-

Table 7: Relationship of self-reported years since diagnosis and labor market outcomes using linear splines

	Employment			Log monthly wages			Monthly work hours		
	(1) Pooled	(2) Males	(3) Females	(4) Pooled	(5) Males	(6) Females	(7) Pooled	(8) Males	(9) Females
	Fixed Effects								
Years since diagnosis									
0–4	$-.018^{*}$	$-.017$	$-.018$	$-.009$	$-.022$	0.037	0.592	0.430	1.499
	(.010)	(.013)	(.015)	(.044)	(.047)	(.116)	(.841)	(.880)	(2.306)
5–11	$-.005$	$-.008$	$-.005$	$-.052^{*}$	$-.038$	$-.112^{**}$	$-.150$	0.005	$-.536$
	(.006)	(.009)	(.008)	(.031)	(.036)	(.050)	(.524)	(.602)	(1.016)
12–20	$-.024^{***}$	$-.036^{**}$	$-.017^{*}$	0.018	0.052	$-.045$	$-.011$	0.158	$-.738$
	(.009)	(.017)	(.010)	(.049)	(.063)	(.055)	(.811)	(.977)	(1.501)
> 20	$-.021$	$-.024$	$-.019$	$-.064$	$-.006$	$-.212^{***}$	1.986	0.770	8.291^{***}
	(.017)	(.040)	(.018)	(.123)	(.124)	(.058)	(3.363)	(3.439)	(1.832)
	Random Effects								
Years since diagnosis									
0–4	$-.026^{***}$	$-.012^{*}$	$-.021^{***}$	0.024^{*}	0.027^{*}	0.027	0.102	$-.127$	0.493
	(.005)	(.006)	(.006)	(.014)	(.016)	(.027)	(.296)	(.341)	(.574)
5–11	$-.002$	$-.005$	$-.000$	$-.031^{**}$	$-.023$	$-.048^{**}$	$-.107$	0.007	$-.356$
	(.005)	(.007)	(.006)	(.015)	(.018)	(.024)	(.321)	(.394)	(.544)
12–20	$-.018^{***}$	$-.023^{**}$	$-.009$	$-.002$	0.027	$-.065^{*}$	$-.116$	0.022	$-.368$
	(.006)	(.010)	(.007)	(.025)	(.031)	(.039)	(.428)	(.572)	(.726)
> 20	$-.006$	$-.009$	$-.002$	0.027^{**}	0.033	0.041^{***}	$-.470$	$-.569$	$-.330$
	(.004)	(.008)	(.004)	(.012)	(.044)	(.015)	(.310)	(1.191)	(.321)
N	38398	16073	22325	16288	10604	5684	20691	13375	7316

Robust standard errors in parentheses
Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number children < 6
wealth, age and calender year fixed effects
The random effects model additionally controls for initial age when entering the survey, being indigenous and gender
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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

reported diabetes with employment chances for men~~and also has~~; the findings have also provided additional evidence of adverse effects for females, supporting earlier findings for Mexico ~~.~~(Seuring et al., 2015b) .[5] We only find very limited evidence for any relationship of self reported diabetes with wages or working hours ~~which is~~ - in line with the results of Minor (2013), who does not find strong evidence for any effects beyond those on employment, though he does not look at working hours. Reasons for this difference in effects at the extensive and intensive margin might be that people who are diagnosed while employed are able to continue to work if they do not suffer from any severe complications yet. They also may resort to doing different activities within the workplace that are less

[5]As a robustness check of our results of the linear probability model for employment status, we also estimated a population-averaged panel-data model with the probit link function. To account for fixed unobserved heterogeneity in this type of model we used the 'within-between' formulation proposed by Bell and Jones (2015) based on the work of Mundlak (1978), which allows to account for ~~time-constant~~ time-invariant unobserved heterogeneity by adding the individual group means as regressors. The 'within-between' formulation also adds group means but uses the demeaned values instead of the complete observation to capture any within variation. This has some advantages as it allows to differentiate clearly between the within and the between effects of the used variables and is less prone to issues of collinearity and multicollinearity. Using this strategy has yielded qualitatively similar results to those of the linear fixed effects specifications~~and the~~. The results are available on request from the authors.

physically demanding and allow them to continue working and earning similar wages as before their diagnosis. Only once complications become increasingly severe would they drop out of the labor market, without going through a previous phase of reduced productivity and labor supply. Some support for this reasoning comes from the RE results in tables 3 and 4~~that show~~ , showing that people with diabetes still employed actually earn more if we disregard any unobserved selection into employment, i.e., those with diabetes that are able to remain employed have some unobserved characteristics that also allow them to earn higher wages. These people might have abilities that make them better at managing their diabetes and also lets them earn higher wages, or they may be so valuable to their working environment that they do not lose their job even if they are ~~limited healthwise~~ constraint health-wise due to diabetes. This ~~appears~~ seems to be particularly the case for those in non-agricultural employment. Overall, it appears that diabetes affects those self-employed or in agricultural employment much more, potentially because they often cannot rely on formal contracts that would protect them against job loss, they are likely less able to access appropriate treatment or are diagnosed much later also due to a lack of access to proper health care, leading to a more rapid development of diabetes complications and also because their jobs are likely more physically demanding, making it harder to continue working once complications appear. However, this reasoning remains somewhat speculative and deserves further research in future studies.

Finally, before moving on to further analysis, we ~~want to note~~ need to emphasise that our fixed effects strategy used to identify the effect of self-reported diabetes relies on those reporting a new diabetes diagnosis between any of the waves, which disregards people that have had diabetes before the start of the survey and so limits the estimation of effects to those more or less recently diagnosed. Accordingly, the obtained results may be different from those with a longer disease history. However, at least the results of the RE analysis, which also includes ~~between person~~ between-person variation, do not indicate that the effects are substantially different even when people with a longer disease history are taken into account. We have to note nonetheless that the analysis of diabetes duration shows somewhat different results ~~with regards to~~ in terms of when the adverse effects appear, as it indicates that the largest effects are found for those with a disease duration over 10 years, which would suggest no immediate effect of diabetes. ~~However, and~~ Yet, as we have mentioned before, the duration analysis relies on information only from the most recent wave to construct a measure of disease duration for all previous waves and hence excludes information from participants dropping out of the sample before the last wave. It further relies on people correctly identifying the year of first diagnosis which is likely subject to recall bias, the longer the diagnosis lies in the past, so that the results might not be directly comparable and generally less reliable than the estimates using self-reported diabetes from all three waves. Also, given that the signs point ~~toward~~ towards a negative effect also immediately after diagnosis (which is also supported by the RE results), but standard errors are ~~to~~ too large to suggest statistical significance~~could be a result of~~ , this hints at a lack of statistical power to identify a relationship for the other duration groups in the analysis using splines. Accordingly, further research in this area is needed to better understand when adverse effects appear. [NOT SURE IF THIS PARA SHOULD BE INCLUDED AS IT PARTICULALRY QUESTIONS THE RESULTS OF THE DURATION ANALYSIS. WHAT DO YOU THINK?]

5.2 Cross-sectional biomarker analysis

The results presented in the previous section did provide important information on the association between self-reported diabetes and ~~with~~ labor market outcomes in Mexico.

While panel data methods have obvious advantages over a cross sectional analysis, there are additional insights to be gained from making use of the biomarker data collected in the third wave of the MxFLS, even if it only allows for a cross-sectional analysis. The biomarker data allows us to identify respondents with HbA1c levels, thereby identifying diabetes cases that have not yet been 'officially' diagnosed. As mentioned before, biomarker data are only available for a subsample of the survey, including everybody aged 45+ as well as a random selection of participants below age 45 (Crimmins et al., 2015). In this sub-sample 694 (10 percent) self-report diabetes and 1,248 (18 percent) have undiagnosed diabetes, i.e. as measured by an HbA1c equal to or greater than 6.5 percent. In an effort to reduce bias in our estimates we additionally estimate every model using community fixed effects 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 could potentially affect the propensity to ~~develop diabetes as well as~~ (1) develop diabetes and (2) to receive a diagnosis of diabetes~~and might also~~ , plus they might affect labor market outcomes.[6]

Ojectively measured diabetes

The results of the complete biomarker analysis are presented in table 8. In order to investigate how the subsample of biomarkers compares to the overall sample of the third wave we first estimate a model showing the association of self-reported diabetes with our outcome variables using data of the entire third wave (columns 1–3 and 13–15 in table 8). We find similar results to our panel data analysis~~albeit~~ , if with somewhat smaller coefficients for the effect on employment chances, particularly for females~~, suggesting~~. This suggests that not accounting for fixed effects might underestimate the effect of self-reported diabetes on women employment chance. Running the same model but using only the data of the biomarker sample (columns 4–6 and 16–18), we find that the coefficient becomes smaller for men and standard ~~increase throughout likely die~~ errors increase throughout, likely due to the reduction in sample size. This ~~suggests~~ indicates that the oversampling of older people in the subsample may ~~introduces~~ introduce a downward bias in the estimates for males, which has to be taken into account when interpreting the results.

In the next step we investigate the association between objectively measured diabetes, irrespective of whether the condition has or has not been diagnosed, and labor market outcomes, by estimating equation 5. [7] We find an adverse association of objectively

[6] We did not use household fixed effects 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.

[7] The diabetes indicator variable still also contains those that self-reported a diabetes diagnosis but had HbA1c levels below the diabetes threshold. This is sensible as it is possible that people with a diabetes diagnosis are able to manage their diabetes in such a way that their HbA1c drop below the threshold and should not be assumed to having falsely reported a diabetes diagnosis ~~, albeit~~ (even though

measured diabetes and employment chances for the pooled (columns 7–9) and to some ~~extend~~ extent for the female sample (columns 19–21), but the coefficient size is reduced throughout, compared to 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. To further investigate this, we include both diagnosed and undiagnosed diabetes jointly as independent explanatory variables, to account for the differences between the two groups as in equation 6.

As shown in columns 10–12 and 22–24, no association between undiagnosed diabetes and any labor outcome is found. For those with diagnosed diabetes the coefficients for the association with employment chances for the pooled sample and women are similar in size and significance to those of the panel analysis, particularly when we include community level fixed effects. They also do not differ much from the results of the specification only accounting for self-reported diabetes. Consistent with our earlier findings there is no indication of any associations between any form of diabetes and wages or working hours.

this ~~admittedly is a possibility as well~~could potentially also be the case). Robustness checks where we specifically accounted for this subpopulation did not show qualitatively different results.

Table 8: Biomarker results

	SR (complete sample)			SR (biomarker sample)			HbA1c $\geq$ 6.5%			SR and undiagnosed			SR (complete sample) (FE)			SR (biomarker sample) (FE)			HbA1c $\geq$ 6.5% (FE)			SR and undiagnosed (FE)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females
Dependent variable: Employment																								
Diagnosed diabetes	−.062***	−.058***	−.044**	−.057***	−.042	−.041*				−.058***	−.041	−.043*	−.066***	−.068***	−.046***	−.060***	−.050*	−.043*				−.063***	−.049*	−.048**
	(.013)	(.018)	(.017)	(.017)	(.026)	(.022)				(.017)	(.026)	(.023)	(.012)	(.017)	(.017)	(.017)	(.026)	(.023)				(.018)	(.026)	(.023)
HbA1c $\geq$ 6.5%							−.023**	−.009	−.021										−.032***	−.013	−.032*			
							(.012)	(.015)	(.017)										(.012)	(.015)	(.018)			
Undiagnosed diabetes										−.004	0.006	−.008										−.013	0.005	−.018
										(.014)	(.016)	(.020)										(.014)	(.019)	(.018)
R2	0.307	0.095	0.143	0.326	0.086	0.151	0.326	0.085	0.326	0.326	0.086	0.151	0.302	0.093	0.104	0.328	0.082	0.108	0.328	0.080	0.328	0.328	0.082	0.108
N	18748	8234	10514	6408	2785	3623	6416	2790	3623	6408	2785	3623	18748	8234	10514	6408	2785	3623	6416	2790	6416	6408	2785	3623
Dependent variable: Log hourly wages																								
Diagnosed diabetes	−.002	0.034	−.054	0.012	0.021	0.004				0.001	0.007	0.006	0.004	0.043	−.048	0.003	−.008	−.039				0.006	−.004	−.051
	(.039)	(.044)	(.072)	(.053)	(.062)	(.101)				(.054)	(.064)	(.104)	(.038)	(.047)	(.072)	(.053)	(.066)	(.113)				(.056)	(.068)	(.117)
HbA1c $\geq$ 6.5%							−.025	−.027	−.025										0.009	0.011	0.009			
							(.034)	(.040)	(.034)										(.036)	(.044)	(.036)			
Undiagnosed diabetes										−.029	−.035	−.036										0.012	0.018	−.052
										(.039)	(.045)	(.072)										(.039)	(.048)	(.079)
R2	0.221	0.206	0.264	0.257	0.246	0.312	0.258	0.246	0.258	0.266	0.255	0.326	0.141	0.117	0.183	0.170	0.146	0.236	0.171	0.146	0.171	0.170	0.146	0.237
N	8402	5515	2887	2687	1803	884	2690	1805	2690	2687	1803	884	8402	5515	2887	2687	1803	884	2690	1805	2690	2687	1803	884
Dependent variable: Usual weekly working hours																								
Diagnosed diabetes	0.533	0.428	0.283	−1.375	−.723	−2.780				−1.136	−.586	−2.431	0.349	0.195	0.394	−1.122	−.339	−.772				−.903	−.302	−.168
	(.829)	(.967)	(1.533)	(1.129)	(1.321)	(2.110)				(1.146)	(1.342)	(2.151)	(.785)	(1.022)	(1.441)	(1.111)	(1.286)	(2.164)				(1.149)	(1.325)	(2.215)
HbA1c $\geq$ 6.5%							0.589	0.420	0.589										0.337	0.000	0.337			
							(.690)	(.786)	(.690)										(.719)	(.835)	(.719)			
Undiagnosed diabetes										1.111	0.468	2.441										0.939	0.160	2.638
										(.823)	(.915)	(1.623)										(.875)	(.955)	(1.838)
R2	0.074	0.043	0.045	0.076	0.053	0.054	0.076	0.053	0.076	0.080	0.059	0.071	0.066	0.026	0.032	0.065	0.031	0.035	0.065	0.031	0.065	0.065	0.031	0.038
N	10428	6820	3608	3446	2302	1144	3451	2306	3451	3446	2302	1144	10428	6820	3608	3446	2302	1144	3451	2306	3451	3446	2302	1144

Robust standard errors in parentheses; other control variables:age, age squared, state dummies, urbanization dummies, education dummies, married dummy, number children < 6 and wealth; calender year dummies are included as data collection for the third wave was stretched out over several years

The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural employment as the base) and for health insurance status; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the finding that undiagnosed diabetes appears to not be associated with any adverse labor outcomes suggests that the two populations with diabetes, those with diagnosed and those with undiagnosed diabetes, are different in terms that are not yet captured by the ~~used control variables~~ control variables we have included. One possible ~~factor that may explain~~ explanatory factor behind these different outcomes ~~is~~ could be the psychological effect of a diabetes diagnosis leading to a reduction in employment chances due to justification bias or by causing depression or anxiety. Further, people with a diabetes diagnosis could be generally less healthy and have more ~~diabetes related complications or other diseases~~ diabetes-related (or unrelated) complications that affect their employment chances. To get a better idea of the driving factors behind the results we included a measure of self-reported health as one of our regressors, despite the fact that it is clearly endogenous due to its relationship with diabetes, which likely leads to worse self-reported health itself. Nonetheless, including it should account for the differences in health between those aware of their diabetes and those undiagnosed, and subsequent changes in the diabetes coefficients may indicate if the differences in health might be driving the more adverse employment effects of self-reported diabetes. Indeed, the coefficients of self-reported diabetes are reduced in size and no longer significant for males and only borderline significant for females (table 9). In the pooled model, however, the coefficient remains highly significant. Interestingly, particularly for males reporting bad health is associated with a reduction in employment chances while this is not the case for females, suggesting that diagnosed males might actually be in a worse health state that is driving the reduction in employment chances while for women other factors could play a more important role. Accounting for measures of overweight and obesity, self-reported hypertension, self-reported heart disease and insurance status does not further affect the diabetes coefficient and significance to an important degree (results ~~not shown here).~~ available on request). A final worry may be that the HbA1c measurements are upward biased as a result of the used technical device Bio-Rad in2it which has been shown to correlate strongly with laboratory measurements (Health Quality Ontario, 2014) but had also been found by one study to produce somewhat higher HbA1c estimates (Yeo et al., 2009) . This could cause many people just above the diabetes threshold to be falsely qualified as having undiagnosed diabetes potentially attenuating the estimates. To test this, we reestimated above models using an alternative diabetes threshold of HbA1c ≥ 7. The results, however, remained similar not providing evidence that did not provide evidence for an attenuation bias due to misclassification.

5.2.1 HbA1c levels and labor market outcomes

The HbA1c information in wave 3 also enables us to use them as a proxy for disease severity and diabetes management. Using three dummy variables for HbA1c groups above the diabetes threshold, i.e. 6.5–7.9, 8–11.9 and 12–14, we investigate how labor market effects differ when taking HbA1c values into account (table 10. For employment chances we again only find significant adverse associations for those with diagnosed diabetes, particularly for HbA1c levels of 6.5 % to 7.9% and 8% to 11.9% compared to people without diabetes and an HbA1c below the diabetes threshold. Given that neither the coefficient size nor significance increases but rather levels off with higher blood glucose levels, we conclude

Table 9: Diagnosed, undiagnosed diabetes and self-reported health and their association with labor market outcomes

	Employment			Log hourly wages			Weekly working hours		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled	Males	Females	Pooled	Males	Females	Pooled	Males	Females
	No community level fixed effects								
Self-reported health status									
good	0.047**	0.028	0.059*	0.006	0.060	−.094	0.238	−1.065	2.773
	(.021)	(.025)	(.032)	(.058)	(.068)	(.110)	(1.077)	(1.224)	(2.089)
fair	0.008	0.003	0.015	−.039	0.018	−.139	0.202	−1.135	3.265
	(.022)	(.025)	(.033)	(.060)	(.071)	(.114)	(1.138)	(1.290)	(2.254)
bad	−.049	−.127**	−.004	−.060	0.060	−.229	−1.529	−5.850**	6.274*
	(.033)	(.050)	(.043)	(.120)	(.165)	(.169)	(2.067)	(2.444)	(3.559)
very bad	0.014	−.153	0.102	−.045	−.304*	0.272	−7.632**	−.854	−14.544***
	(.093)	(.139)	(.127)	(.170)	(.180)	(.263)	(3.288)	(2.705)	(4.667)
Diagnosed diabetes	−.048***	−.023	−.037	0.020	0.013	0.041	−.972	−.237	−2.837
	(.018)	(.025)	(.023)	(.055)	(.066)	(.103)	(1.161)	(1.360)	(2.174)
Undiagnosed diabetes	−.004	0.005	−.009	−.030	−.034	−.037	1.138	0.436	2.591
	(.014)	(.016)	(.020)	(.039)	(.045)	(.073)	(.824)	(.914)	(1.620)
R2	0.329	0.093	0.154	0.267	0.257	0.330	0.081	0.061	0.078
	Community level fixed effects								
Self-reported health status									
good	0.042*	0.024	0.059*	0.004	0.066	−.115	0.096	−1.169	3.423*
	(.022)	(.027)	(.035)	(.066)	(.070)	(.136)	(1.225)	(1.342)	(2.054)
fair	−.003	−.005	0.006	−.038	0.031	−.152	−.163	−1.646	4.540*
	(.021)	(.026)	(.033)	(.069)	(.070)	(.141)	(1.330)	(1.404)	(2.405)
bad	−.059*	−.125**	−.025	−.095	−.018	−.376*	−2.653	−6.124**	6.882*
	(.034)	(.056)	(.046)	(.134)	(.174)	(.200)	(2.225)	(2.678)	(3.972)
very bad	0.007	−.167	0.119	−.085	−.347*	0.306	−8.931**	−1.427	−17.360***
	(.088)	(.127)	(.153)	(.186)	(.182)	(.433)	(3.605)	(3.274)	(4.914)
Diagnosed diabetes	−.052***	−.032	−.040*	0.030	0.004	0.016	−.606	0.111	−.668
	(.018)	(.025)	(.024)	(.055)	(.069)	(.109)	(1.154)	(1.374)	(2.065)
Undiagnosed diabetes	−.014	0.005	−.019	0.012	0.020	−.049	0.965	0.119	2.753
	(.014)	(.019)	(.018)	(.039)	(.049)	(.079)	(.887)	(.960)	(1.811)
R2	0.331	0.089	0.111	0.171	0.148	0.244	0.067	0.035	0.048
N	6406	2785	3621	2685	1803	882	3444	2302	1142

Robust standard errors in parentheses

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

Calender year dummies are included as data collection for the third wave was stretched out over several years with 2009 as the base

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 base for health status is "very good" health

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

that current diabetes management or severity of diabetes as indicated by HbA1c levels is unrelated to the current employment status or any labor market outcome.

Table 10: Diabetes HbA1c levels and labor market outcomes

	Employment			Log monthly wages			Monthly work hours		
	(1) Pooled	(2) Males	(3) Females	(4) Pooled	(5) Males	(6) Females	(7) Pooled	(8) Males	(9) Females
	No community level fixed effects								
Self-reported diabetes									
$6.5 \leq$ HbA1c < 8	$-.083^{**}$	$-.105^{*}$	$-.041$	$-.125$	$-.241^{**}$	0.152	-1.136	1.841	-7.930^{*}
	(.036)	(.056)	(.046)	(.115)	(.117)	(.240)	(2.308)	(2.571)	(4.452)
$8 \leq$ HbA1c < 12	$-.074^{***}$	$-.040$	$-.073^{**}$	0.067	0.085	$-.002$	-2.333	-2.180	-2.923
	(.024)	(.035)	(.030)	(.080)	(.090)	(.175)	(1.704)	(1.925)	(3.591)
HbA1c ≥ 12	$-.007$	0.030	$-.003$	$-.109$	$-.122$	$-.009$	-1.444	$-.038$	-4.804
	(.040)	(.054)	(.053)	(.102)	(.135)	(.161)	(2.685)	(3.338)	(4.048)
Undiagnosed diabetes									
$6.5 \leq$ HbA1c < 8	0.005	0.004	0.010	$-.025$	$-.040$	0.010	1.829^{*}	1.158	3.431
	(.017)	(.021)	(.026)	(.048)	(.058)	(.088)	(1.026)	(1.163)	(2.123)
$8 \leq$ HbA1c < 12	$-.008$	0.008	$-.022$	$-.050$	$-.019$	$-.180$	$-.364$	$-.540$	$-.361$
	(.023)	(.026)	(.034)	(.064)	(.074)	(.128)	(1.300)	(1.517)	(2.487)
HbA1c ≥ 12	$-.022$	0.016	$-.047$	$-.032$	$-.094$	0.080	0.573	$-.580$	2.140
	(.034)	(.042)	(.046)	(.088)	(.096)	(.177)	(2.051)	(2.156)	(4.192)
R2	0.327	0.088	0.152	0.267	0.258	0.329	0.081	0.060	0.076
	Community level fixed effects								
Self-reported diabetes									
$6.5 \leq$ HbA1c < 8	$-.086^{**}$	$-.118^{**}$	$-.040$	$-.138$	$-.212^{*}$	0.010	$-.947$	2.144	-5.711
	(.036)	(.055)	(.048)	(.108)	(.122)	(.250)	(2.348)	(2.747)	(4.765)
$8 \leq$ HbA1c < 12	$-.076^{***}$	$-.048$	$-.080^{**}$	0.046	0.035	0.015	-2.178	-2.224	-1.193
	(.024)	(.033)	(.031)	(.078)	(.091)	(.168)	(1.637)	(1.863)	(3.393)
HbA1c ≥ 12	$-.018$	0.029	$-.032$	$-.083$	$-.111$	$-.061$	$-.851$	0.235	-1.595
	(.036)	(.044)	(.052)	(.104)	(.144)	(.158)	(2.626)	(3.039)	(3.953)
Undiagnosed diabetes									
$6.5 \leq$ HbA1c < 8	$-.005$	0.006	$-.003$	0.012	0.016	$-.024$	1.679	1.002	3.569
	(.018)	(.021)	(.025)	(.046)	(.058)	(.099)	(1.140)	(1.193)	(2.359)
8 <= HbA1c < 12	$-.017$	0.006	$-.029$	$-.012$	0.011	$-.223^{*}$	$-.766$	$-.955$	$-.100$
	(.024)	(.035)	(.031)	(.067)	(.078)	(.132)	(1.281)	(1.495)	(2.634)
HbA1c >= 12	$-.032$	0.010	$-.058$	0.021	$-.016$	0.182	0.597	-1.653	1.726
	(.031)	(.041)	(.046)	(.074)	(.087)	(.179)	(1.996)	(2.094)	(4.055)
R2	0.328	0.083	0.109	0.171	0.148	0.242	0.067	0.033	0.041
N	6408	2785	3623	2685	1803	882	3444	2302	1142

Robust standard errors in parentheses

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

Calender year dummies are included as data collection for the third wave was stretched out over several years

The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural employment as the base) and for health insurance status

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

~~Only~~ To the best of our knowledge only one other study has used biomarker data to analyze the relationship with labor market outcomes. Brown I I I et al. (2011) use data for a Mexican American population in a broadly comparable way to this paper, though it stops short of investigating the labor market impact of undiagnosed diabetes. The study indicates an increasing negative relationship of diabetes and employment chances and wages for men as HbA1c levels increased, interacting HbA1c levels with the diabetes dummy. We estimate a similar model but do not find any indication that increasing

HbA1c levels were related with employment chances (results available on request). We find the same to be true when using indicator variables for different HbA1c groups above the diabetes threshold, where the main effects are found for those with relatively well managed diabetes and no effects for those with the highest HbA1c levels. While we cannot explain ~~theses~~ these differences in the results between our study and the study of Brown I I I et al. (2011) other than by acknowledging the fact that our ~~studies is placed in a~~ study is located in a very different country and uses a different age group, other results ~~of their study also showed~~ in their study were reasonably in line with our findings that once people are diagnosed with diabetes current diabetes 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 month, which may not be very representative of blood glucose levels in the years before and after the diabetes diagnosis and that ultimately determine how soon complications appear and how severe they are.

6 Conclusion

According to the relatively novel panel data used here, self-reported diabetes has increased in Mexico between 2002 and ~~2011~~2012, and overall diabetes prevalence has reached worrying levels, in particular when including the strikingly large share of the undiagnosed diabetes population that we are able to investigate in this analysis. ~~Given~~ In light of the rising importance of diabetes in Mexico and other MICs~~knowledge about the economic costs it entails has also become highly needed and we tend to this need~~, 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 ~~relatively less studied~~ generally much under-researched labor market effects of diabetes, taking into account several complexities that arise due the nature of the disease.

Our first important finding is the confirmation of a considerable employment burden for men and women reporting a diabetes diagnosis, even after accounting for an important source of potential unobserved heterogeneity by using individual FE for the first time in the investigation of diabetes labor market effects~~,~~. We see this as a potentially more valid strategy than relying on the at least questionable IV strategies used so far in the literature. This strategy also allowed us to account for any sample selection into employment based on time-invariant heterogeneity such as innate ability~~and we~~. We find that diabetes very likely is not related to a decrease in wages or working hours, suggesting that people who receive a diabetes ~~diagnosed~~ diagnosis often drop out of the labor market once debilitating diabetes related complications appear ~~and~~ that would limit their work performance~~, in particular~~. This is of particular relevance in jobs where formal job protection and good access to healthcare may be less available and that are also relatively more strenuous. However, those who remain employed do not appear to suffer any wage or labor supply effects, possibly because they are still relatively healthy or ~~work at a job or~~ they are able to resort to a type of work were their diabetes does not inhibit their performance. Future research will be needed to confirm and further investigate this finding and interpretation.

Using the same strategy we also investigate the effect of diabetes duration on labor market outcomes and find that diabetes likely has a continuous adverse relationship with employment chances, which might become increasingly stronger after the first ten years since diagnosis. This is not surprising given that many complications of diabetes only appear after some time of living with the disease. This ~~is likely bad news~~ likely bodes ill for many countries were diabetes, and in particular diabetes start appearing at earlier ages in recent years, causing people to live with the disease for larger parts of their productive lifespan~~likely~~ , possibly exacerbating the economic effects of unemployment due to diabetes.

The other main contribution of this paper is the ability to identify previously unobserved people with diabetes via the use of biomarker data and the finding that undiagnosed diabetes is not associated with labor market outcomes in Mexico. We still mostly find a negative association of diabetes with labor market outcomes but this association is significantly reduced. While this is only a cross-sectional analysis and the results can only be ~~seen~~ interpreted as associations, this is still an indication that studies relying solely on self-reported diabetes may not under- but overestimate the effect of diabetes on labor market outcomes due to non-classical reporting error. A similar finding was ~~made~~ reported by Cawley et al. (2015) regarding reporting error in weight. Therefore, it is important to ~~see~~ consider the population of people with diabetes as consisting of those with self-reported diabetes and those with undiagnosed diabetes, where the former differs from the latter in the amount of health information ~~it~~ they possess and likely also ~~its~~ their health status. The best way then to analyze the effects of diabetes should be to, if possible, explicitly account for both groups~~if possible~~ , as using an IV strategy will not correct for non-classical measurement error (Cawley et al., 2015).

To reiterate, the study finds a considerable negative impact of self-reported diabetes on labor market outcomes, a finding that is likely also of relevance to other MICs facing similar diabetes problems. Further, when estimating any effects of diabetes using self-reported measures, any conclusions drawn should be limited to ~~the~~ those already diagnosed as otherwise any effects are likely to be overstated due to the often large population with undiagnosed diabetes.

Appendix

A Strategies to deal with measurement error

As discussed above, 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 as well as to partly correct it, in the attempt to reduce any resulting bias when 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 individuals 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 ~~US~~ USA population found that 30 percent 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 find similar inconsistencies in the diabetes self-reports over the three waves of the MxFLS data, with between 10 to 20 percent of those reporting diabetes in one wave not reporting diabetes in one of the subsequent waves. In order to correct these inconsistencies we were interested in the validity of diabetes elf-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 percent for China), while sensitivity - a measure of how many people with diabetes, diagnosed or undiagnosed, actually self-report the disease- was low (40 percent 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 could have diabetes as well but are unaware.

Thanks to the biomarker data provided in the third wave of the MxFLS, we are also able to assess the validity of self-reported diabetes by using HbA1c levels and the self-reports of diabetes related medicine use. The World Health Organization (WHO) recommends a cut-off value of an HbA1c $\geq$ 6.5 percent to diagnose a person with diabetes (World Health Organization, 2011). Of the subsample selected for biomarker measurements and answering the diabetes question (n=6895), 705 reported a diabetes diagnosis and of those 632 (90 percent) had an HbA1c $\geq$ 6.5 percent or did report taking diabetes medication, indicating relatively high specificity in our data as well.

We therefore assumed for people with information from only two waves, 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 11:

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

Table 11: 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

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 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 percent for those with two self-reports compared to 7.0 percent for those with only one self-report of diabetes. Further, of those with one self-report, only 30 percent had an HbA1c$\geq$ 6.5 percent compared to 87 percent of those with two self-reports. Bases on these results we were reassured that the way we have dealt with the inconsistencies in the data should minimize misclassification of people into diabetes or no-diabetes and should 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 models using duration of diabetes instead of self-reported diabetes must be interpreted particularly carefully.

Acknowledgement

We are grateful to the participants at the European Health ~~Economes~~ 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

Ayyagari, P., Grossman, D., Sloan, F., Mar. 2011. Education and health: evidence on adults with diabetes. International Journal of Health Care Finance and Economics 11 (1), 35–54.

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

Barquera, S., Hernandez-Barrera, L., Tolentino, M. L., Espinosa, J., Ng, S. W., Rivera, J. A., Popkin, B. M., Dec. 2008. Energy Intake from Beverages Is Increasing among Mexican Adolescents and Adults. Journal of Nutrition 138 (12), 2454–2461.

Basu, S., Yoffe, P., Hills, N., Lustig, R. H., Feb. 2013. The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data. PLoS ONE 8 (2), e57873.

Beagley, J., Guariguata, L., Weil, C., Motala, A. a., 2014. Global estimates of undiagnosed diabetes in adults. Diabetes Research and Clinical Practice 103 (2), 150–160.

Bell, A., Jones, K., 2015. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. Political Science Research and Methods 3 (01), 133–153.

Bellamy, L., Casas, J.-P., Hingorani, A. D., Williams, D., 2009. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 373 (9677), 1773–9.

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

Bound, J., Jaeger, D. a., Baker, R. M., 1995. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. Journal of the American Statistical Association 90 (430), 443–450.

Brown, H. S., Pagan, J. A., Bastida, E., 2005. The impact of diabetes on employment: Genetic IVs in a bivariate probit. Health Economics 14 (5), 537–544.

Brown, T. T., 2014. How effective are public health departments at preventing mortality? Economics and human biology 13, 34–45.

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

Cawley, J., Maclean, J. C., Hammer, M., Wintfeld, N., Dec. 2015. Reporting error in weight and its implications for bias in economic models. Economics & Human Biology 19, 27–44.

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

Currie, J., Vogl, T., 2013. Early-Life Health and Adult Circumstance in Developing Countries. Annual Review of Economics 5 (1), 1–36.

Eriksson, A.-K., van den Donk, M., Hilding, A., Ostenson, C.-G., Sep. 2013. Work Stress, Sense of Coherence, and Risk of Type 2 Diabetes in a Prospective Study of Middle-Aged Swedish Men and Women. Diabetes Care 36 (9), 2683–2689.

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

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

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

Health Quality Ontario, 2014. Health Quality Ontario. Point-of-care hemoglobin A1c testing: an evidence-based analysis. Ontario health technology assessment series 14 (8), 1–30.
URL http://www.hqontario.ca/evidence/publications-and-ohtac-recommendations/ontario-health-technology-assessment-series/eba-point-of-care-a1cPermission

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

International Diabetes Federation, 2014. Diabetes Atlas 2014 Update, 6th Edition. International Diabetes Federation.

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

Latif, E., May 2009. The impact of diabetes on employment in Canada. Health Economics 18 (5), 577–589.

Li, Y., He, Y., Qi, L., Jaddoe, V. W., Feskens, E. J. M., Yang, X., Ma, G., Hu, F. B., Oct. 2010. Exposure to the Chinese Famine in Early Life and the Risk of Hyperglycemia and Type 2 Diabetes in Adulthood. Diabetes 59 (10), 2400–2406.

Lim, E. L., Hollingsworth, K. G., Aribisala, B. S., Chen, M. J., Mathers, J. C., Taylor, R., 2011. Reversal of type 2 diabetes: Normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. Diabetologia 54 (10), 2506–2514.

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

Meza, R., Barrientos-Gutierrez, T., Rojas-Martinez, R., Reynoso-Noverón, N., Palacio-Mejia, L. S., Lazcano-Ponce, E., Hernández-Ávila, M., 2015. Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates. Preventive Medicine.
URL http://linkinghub.elsevier.com/retrieve/pii/S0091743515003278

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

Minor, T., Dec. 2013. An investigation into the effect of type I and type II diabetes duration on employment and wages. Economics & Human Biology 11 (4), 534–544.

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

Nouwen, A., Nefs, G., Caramlau, I., Connock, M., Winkley, K., Lloyd, C. E., Peyrot, M., Pouwer, F., 2011. Prevalence of depression in individuals with impaired glucose metabolism or undiagnosed diabetes: A systematic review and meta-analysis of the European Depression in Diabetes (EDID) research consortium. Diabetes Care 34 (3), 752–762.

O'Neill, D., Sweetman, O., 2013. The consequences of measurement error when estimating the impact of obesity on income. IZA Journal of Labor Economics 2 (1), 3.

Paddison, C. A. M., Eborall, H. C., French, D. P., Kinmonth, A. L., Prevost, A. T., Griffin, S. J., Sutton, S., Feb. 2011. Predictors of anxiety and depression among people attending diabetes screening: A prospective cohort study embedded in the ADDITION (Cambridge) randomized control trial. British Journal of Health Psychology 16 (1), 213–226.

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

Pforr, K., 2014. femlogit - Implementation of the multinomial logit model with fixed effects. The Stata Journal 14 (4), 1–16.

Reynoso-Noverón, N., Mehta, R., Almeda-Valdes, P., Rojas-Martinez, R., Villalpando, S., Hernández-Ávila, M., Aguilar-Salinas, C. a., 2011. Estimated incidence of cardiovascular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey. Cardiovascular Diabetology 10 (1), 1.

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

Schaller, J., Stevens, A. H., Sep. 2015. Short-run effects of job loss on health conditions, health insurance, and health care utilization. Journal of Health Economics 43, 190–203.

Schneiderman, N., Ironson, G., Siegel, S. D., 2005. Stress and Health: Psychological, Behavioral, and Biological Determinants. Annual Review of Clinical Psychology 1 (1), 607–628.

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

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

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

Slade, A. N., 2012. Health Investment Decisions in Response to Diabetes Information in Older Americans. Journal of Health Economics 31 (3), 502–520.

Sotomayor, O., 2013. Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes. Economics & Human Biology 11 (3), 281–293.

Strauss, J., Thomas, D., 1998. Health, Nutrition, and Economic Development. Journal of Economic Literature 36 (2), 766 – 817.

Thoolen, B. J., De, R. D. T., Bensing, J. M., Gorter, K. J., Rutten, G. E., 2006. Psychological outcomes of patients with screen-detected type 2 diabetes: the influence of time since diagnosis and treatment intensity. Diabetes Care 29, 2257–2262.

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

Villalpando, S., de la Cruz, V., Rojas, R., Shamah-Levy, T., Avila, M. A., Gaona, B., Rebollar, R., Hernández, L., 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., Jacobs, S. B. R., Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., García-Ortíz, H., Gómez-Vázquez, M. J., Burtt, N. P., Aguilar-Salinas, C. a., González-Villalpando, C., Florez, J. C., Orozco, L., Haiman, C. a., Tusié-Luna, T., Altshuler, D., 2014. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 506 (7486), 97–101.

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), 25.

Yeo, C.-P., Tan, C. H.-C., Jacob, E., Sep. 2009. Haemoglobin A1c: evaluation of a new HbA1c point-of-care analyser Bio-Rad in2it in comparison with the DCA 2000 and central laboratory analysers. Annals of Clinical Biochemistry 46 (5), 373–376.

Yuan, X., Liu, T., Wu, L., Zou, Z.-Y., Li, C., Apr. 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.

Zajacova, A., Dowd, J., Schoeni, R. F., Wallace, R. B., 2010. Consistency and precision of cancer reporting in a multiwave national panel survey. Population Health Metrics 8 (1), 20.

Zhang, X., Zhao, X., Harris, A., 2009. Chronic diseases and labour force participation in Australia. Journal of Health Economics 28 (1), 91–108.

Zhao, M., Konishi, Y., Glewwe, P., 2013. Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption. Journal of Health Economics 32 (2), 367–385.