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

Till Seuring*[a,b], Pieter Serneels[b], and Marc Suhrcke[c]

[a]*Leibniz Institute for Prevention Research and Epidemiology - BIPS*

[b]*University of East Anglia*

[d]*University of York*

Running head: The impact of diabetes on labor market outcomes in Mexico

Keywords: diabetes, employment, wages, biomarker, Mexico, panel data

JEL: I14, I15, J22, J31, D83

Funding statement: No funding was received to support this project.

*Corresponding author. Leibniz Institute for Prevention Research and Epidemiology - BIPS, Achterstr. 30, 28359 Bremen, Germany, Email: seuring@leibniz-bips.de, Phone: +49 421 218 569 23

Abstract

There remain gaps in the understanding of the economic consequences of diabetes, in particular in a context where diabetes often remains undiagnosed, as appears to be particularly the case in low- and middle-income countries (LMICs). We investigate the impact of diabetes on labor outcomes in Mexico using panel and biomarker data applying fixed effects estimation to account for potential endogeneity and using biomarker information to include previously undiagnosed diabetes. We find strong evidence for adverse effects of self-reported diabetes on the probability of being employed, in particular in agricultural work, but not on wages or hours worked. The employment probability falls gradually with time since diagnosis. In the biomarker analysis we observe that 18% of all observations are false negatives (undiagnosed), i.e. do not report diabetes but exhibit glycated hemoglobin (HbA1c) levels above the clinical diabetes threshold. The estimated employment impact for those that were found to exceed the clinical threshold suggests no effects for men but similar effects for women compared to self-reported diabetes. Further analysis reveals that there is no effect of diabetes on labor outcomes for undiagnosed women or men. The results highlight both the importance of the economic impact of diabetes, and the need to take into account undiagnosed patients.

~~Keywords: diabetes, employment, wages, biomarker, Mexico, panel data~~

~~JEL: I14, I15, J22, J31, D83~~

1. ~~Introduction~~ INTRODUCTION

Diabetes, ~~and particularly its most common variant, type 2 diabetes, has increased worldwide and is expected to continue to rise over the next decades . The condition~~ a disease characterized by elevated blood glucose levels due to the body's inability to use insulin properly, has become a problem for both middle-income countries and high-income countries~~alike~~, with over two-thirds of people with diabetes living in the developing world

(International Diabetes Federation, 2015). ~~Mexicans and Mexican-Americans appear to be particularly affected by diabetes, also in comparison to other Latino populations .~~ In Mexico, diabetes prevalence has grown from 6.7% in 1994 to 14.4% in 2006 (Barquera et al., 2013) and ~~recent estimates suggest~~ a 15.8% prevalence rate in 2015 ~~. Already now, diabetes is~~ and diabetes has become the number one ~~cause of death in Mexico~~reason for death (International Diabetes Federation, 2015). Diabetes increases the risk for heart disease and stroke, blindness, kidney disease and nerve problems, food ulcers and amputations due to elevated glucose levels (Reynoso-Noverón et al., 2011). However, via effective self-management of the disease much if not all of the complications can be avoided (Lim et al., 2011; Gregg et al., 2012).

The observed trend has been attributed to a deterioration in diet and a reduction in physical activity (Barquera et al., 2008; Basu, Yoffe, Hills, & Lustig, 2013), while genetic predisposition among Mexicans with pre-Hispanic ancestry may also play a role (Williams et al., 2013). Recent evidence indicates that the onset of diabetes has been occurring at an ever earlier age in Mexico (Bello-Chavolla, Rojas-Martinez, Aguilar-Salinas, & Hernández-Avila, 2017)~~. With treatment as ineffective as it currently is—only a minority~~ , which given that only a minority of patients achieves adequate blood glucose control (Barquera et al., 2013)~~—the earlier onset will increase the likelihood of~~ , will likely lead to an increase in complications during the productive lifespan.

~~Diabetes describes various conditions characterized by high blood glucose values, with the predominant disease being type 2 diabetes accounting for about 90of all diabetes cases . The elevated blood glucose levels that are a result of~~ Further, the ~~body's inability to use insulin properly to maintain blood glucose at normal levels, can entail a range of adverse health effects for the individual concerned. However, via effective self-management of the disease much if not all of the complications can be avoided . In the absence of this management—or in the case of inadequate treatment—over time the constantly elevated blood glucose levels can lead to heart disease and stroke, blindness, kidney disease and~~

~~nerve problems, food ulcers and amputations . Consequently, diabetes can reduce an individual's economic activity, including its productivity and labor market participation.~~ high diabetes burden in Mexico coexist with high levels of infectious diseases, exposing the health system to a 'double-disease burden' that increases the pressure to identify treatment priorities and to efficiently use the existing resources.

~~The effect of diabetes~~ Despite the catastrophic impact of diabetes on health, its economic consequences, in particular in low- and middle-income countrys (LMICs) have received less attention. This is particularly true for its effects on labor outcomes (Seuring, Archangelidi, & Suhrcke, 2015), which has been studied predominantly in high-income countries, ~~where diabetes was associated with reductions~~ suggesting substantial economic losses for individuals and households affected by diabetes (Brown, Pagán, & Bastida, 2005; Brown, 2014; Brown et al., 2011; Minor, 2011, 2013; Minor & MacEwan, 2016; Latif, 2009). For LMICs less evidence is available. Liu and Zhu (2014) exploit a natural experiment in China and find a significant reduction in income for those with a recent diagnosis of diabetes. A study for Mexico using cross-sectional data from 2005, finds a significant ($p<0.01$) reduction in employment probabilities ~~as well as wages and labor supply .~~ [1] for males of 10 percentage points and for females 4.5 percentage points ($p<0.1$) (Seuring, Goryakin, & Suhrcke, 2015). While these studies have provided useful evidence, ~~many of the complexities of the relationship between diabetes and labor outcomes remain unaddressed. Especially~~ in order to address the potential endogeneity of diabetes most have relied on the same instrumental variable (IV) strategy—the family history of diabetes—to exploit the genetic component of the disease in order to establish a causal relationship. However, because family history of diabetes may also proxy for other genetically transferred traits, including unobserved abilities, as well as intrahousehold or intergenerational dynamics that impact labor outcomes directly, its validity remains at least debatable. However, panel data methods to account for time-invariant unobserved

[1] ~~We know of only two studies for middle-income countries, one for Mexico and one for China , which are discussed later in greater detail.~~

individual characteristics~~(e.g.~~ , haven not yet been used. Time-invariant unobservables such as health endowments or risk preferences ~~) may~~ could adversely affect health in general and the propensity to develop type 2 diabetes in particular (van Ewijk, 2011; Sotomayor, 2013; Li et al., 2010) ~~; they may also affect employment probabilities, wages or working hours—either~~ as well as labor outcomes—either directly through their effects on contemporaneous productivity (Currie & Vogl, 2013), or indirectly by limiting educational attainment and human capital accumulation (Ayyagari, Grossman, & Sloan, 2011). ~~Further, given the chronic nature of the condition, a better understanding of the time and~~ They thereby present one of the major sources of a potential bias that could be accounted for by the use of individual level fixed effects (FE), which does not rely on the strong assumptions of an IV.

Other limitations of the previous literature include only limited evidence on the severity of the potential labor market penalties ~~is important.~~

~~Especially in a middle-income country setting, large parts of the population remain undiagnosed~~ over the duration of diabetes. Further, undiagnosed diabetes—a particular problem in LMICs (Beagley, Guariguata, Weil, & Motala, 2014)~~, implying that studies relying~~ — has mostly remained unaccounted for. Studies mainly relied on self-reported diabetes data~~may leave undetected~~ , leaving undetected potentially important differences between those with diagnosed and undiagnosed diabetes. ~~It is at least conceivable, for instance, that a diabetes diagnosis in itself may exert an effect independently of that resulting from the actual disease. People aware of their condition could be less inclined to continue working if this interferes with their disease management; or they may be suffering from psychological~~ These could be caused by the effects of stress, depression ~~, or anxiety , caused by the sudden awareness of being sick; they may also use the diagnosis~~ or anxiety resulting from a diagnosis (Liu & Zhu, 2014), as well as the disease being used as a justification ~~for decreasing their~~ to reduce labor supply (Kapteyn, Smith, & Van Soest, 2009). ~~Hence~~ Further, it is likely that a diabetes diagnosis is related to the transition from an asymptotic to symptomatic state with the appearance of diabetes complications, leading to a selection of people in worse health and at a later diabetes stage into the diagnosed/self-reporting population. Further, a diagnosis may also be related to socioeconomic factors affecting the access to health care and its quality, leading to a selection of better educated and wealthier people into the diagnosed population. Therefore, labor market effects might well be distinct for people with self-reported diabetes versus those unaware of their condition.

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes, critically improving upon previous work by paying close attention to the challenges of unobserved heterogeneity, to the chronic nature of diabetes and to those undiagnosed. ~~To this end we~~ We use three waves of panel data from the Mexican Family Life Survey (MxFLS), covering the period 2002–2012. Applying individual level fixed effects (FE)~~for the first time in this literature~~, we take account of time-invariant heterogeneity when assessing the impact of self-reported diabetes and the time since diagnosis on labor outcomes.[1] We also make use of rich and novel biomarker data from the most recent wave of the MxFLS, to explore the role of undiagnosed ~~diabetes—an issue that has remained unexplored in the existing literature despite its considerable importance.~~

~~Our results of the panel data analysis for self-reported diabetes suggest an economically important decrease in the employment probability of those who are aware of their disease. Wages and working hours, however, remain unaffected. Further analysis for the long term indicates that employment probabilities are reduced in particular immediately after diagnosis and then further after the first 10 years.~~

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

[1] A recent review of the economic cost of diabetes confirms the scarcity of evidence for low- and middle-income countries (Seuring, Archangelidi, & Suhrcke, 2015).

~~in contrast to those self-reporting diabetes, men and women unaware of their condition do not experience adverse labor market effects.~~

2. ~~Diabetes and labor outcomes—what do we know?~~

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

~~More studies have investigated the effects of diabetes on labor outcomes in high-income countries. Brown et al. (2005) consider an elderly population of Mexican-Americans living close to the Mexican border in the US, and find 7 percentage points lower employment rates for men with self-reported diabetes, while for women, the negative relationship becomes insignificant when using estimation. In a similar vein, Brown et al. (2011), again considering a cross-section of Mexican-Americans, detect a negative relationship between the level of on one hand and the probability of employment or male wages on the other hand. Women remain again unaffected.~~

~~Slightly different results are obtained in two other studies, this time for a more representative US population: using a sample comprised exclusively of women, Minor (2011) finds a significant negative effect of self-reported diabetes on female employment and earnings but not on working hours. In this study self-reported diabetes turns out as endogenous and the simple probit estimates are downward biased compared to estimates. In a subsequent~~

~~study by the same author, employment probabilities decline shortly after diagnosis for men and after about 10 years for women, while wages are not affected by the time since diagnosis .~~

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

~~To the best of our knowledge, only one labor market related study has considered undiagnosed diabetes, if for a small sample: Minor and MacEwan (2016) find no statistically significant relationship between undiagnosed diabetes and employment probabilities, but negative and statistically significant results when using self-reported diabetes, for both men and women. When merging both the undiagnosed and self-reporting respondents into one diabetes group, the effect on employment remains statistically significant but decreases in magnitude~~diabetes.

~~While these studies suggest substantial economic losses for individuals and households affected by diabetes, most of the existing evidence suffers from methodological limitations, in particular due to the use of cross-sectional data and the limited possibilities to account for unobserved characteristics. The papers attempting to address this bias rely on the family history of diabetes as the identifying instrument to exploit the genetic component of the disease. However, it remains debatable whether the instrument indeed satisfies the exclusion restriction, as it may also proxy for other genetically transferred traits, including unobserved abilities, as well as intrahousehold or intergenerational dynamics that impact labor outcomes directly.~~[2]

[2] ~~It is plausible that diabetes might deteriorate parental health in such a way that the offspring either has to give up its employment to provide care, or has to increase labor supply to compensate for lost income, as also argued by Seuring, Goryakin, and Suhrcke (2015).~~

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

~~This paper makes headway to overcome these key limitations in two ways. First, we apply estimations to three waves of panel data, allowing to control for unobserved time invariant characteristics. We further consider the effect of self-reported diabetes on the type and sector of employment, and the long term effects in the years after diagnosis. Second, we use biomarker data for a large subsample of the population to carry out a comparison between the effect of self-reported and clinically tested diabetes. This also allows us to infer about the effects for undiagnosed patients, who suffer from diabetes according to a clinical test, but are unaware of this.~~

2. ~~Context and Data~~DATA

~~Mexico is a middle-income country that ranks among the the countries with the highest levels of obesity and diabetes prevalence in the world . A recent study showed that diabetes accounted for one-third of all deaths among those 35 to 74 years old in Mexico City, with renal disease, cardiac disease, infection, acute diabetic crisis, and other vascular diseases being the biggest contributors to the elevated mortality risk of people with diabetes . The high diabetes burden in Mexico coexist with high levels of infectious diseases, exposing the health system to a 'double-disease burden' that increases the pressure to identify treatment priorities and to efficiently use the existing resources. Earlier studies point to a failry poor performance of the health system when it comes to diagnosing and effectively treating diabetes in Mexico, with most patients achieving only poor glucose control and suffering from other untreated risk factors such as hypertension . In addition, about half of the diabetes population has been estimated to be unaware of the condition .~~

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

Our outcome variables of interest ~~comprise the labor outcomes employment~~ are employment status, hourly wage, weekly working hours and occupation.[2] ~~About half of the respondents in the sample live in rural areas~~ Descriptive statistics for the entire panel sample show that 86% of men report some form of employment compared to 37% of women (see Table I). Interestingly, men do not report considerably higher hourly wages than women but work more hours per week. Men also work more often in agricultural jobs while women are more likely to be self-employed or in non-agricultural wage employment. Women also have lower educational attainment on average.

~~In the~~ The first part of the analysis ~~we focus~~ focuses on the relationship of labor outcomes with self-reported diabetes[3]. For the pooled data of all three waves (~~Table~~ table I), diabetes was self-reported by 5% of men and 6% of women, respectively. This is consistent with Barquera et al. (2013), who observe a prevalence of diagnosed diabetes in Mexico of 7.5% in 2006, using a slightly older sample that also included respondents beyond 64 years of age. Apart from self-reported diabetes information that is available in all rounds, we also use information on the self-reported year of diagnosis as well as biometrically measured glycated hemoglobin (HbA1c) levels for a subsample of respondents. Throughout, our analysis focuses on the working age population (15–64), and ~~excludes~~ exclude pregnant

[2] Employment status is defined as having worked or having carried out an activity that helped with the household expenses the last week and working for at least four hours per week. We explicitly include those employed informally, for instance people working in a family business. We tested if changing the definition of being employed to having worked at least ten hours per week affects the results~~, and this~~ . This only leads to marginal changes~~in the coefficients and standard errors~~, ~~keeping~~ not affecting the interpretation of the results~~unchanged~~. ~~The hourly~~ Hourly wage was calculated by adding up ~~the~~ reported monthly income from the first and second job (if any)~~and~~ , dividing it by the average number of weeks per month~~providing us with an estimate of the average earnings per week~~, which ~~is~~ was then divided by ~~the~~ weekly working hours~~to arrive at the hourly wage~~. ~~labor~~ Labor income was reported in two ways: either responding to questions on wages, income from piecework, tips, income from extra hours, meals, housing, transport, medical benefits and other earnings, or by reporting the aggregate labor income for the whole month. We adjusted the calculated wage for inflation from the year of the interview up to 2013 and ~~take~~ took the log of real wages. Due to a considerable number of missing or zero income reports, the sample used for the wage estimation is smaller than the sample for working hours. Working hours reflect working hours of the first and second job (if applicable).

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

women and those in school.[4]

Table I about here

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

~~A further, and no less important,~~ The second part of the analysis focuses on another source of measurement error ~~with~~ related to self-reported diabetes~~is~~ : the omission of those with undiagnosed diabetes, i.e., the false negatives. ~~The information on~~ Thanks to the availability of biometrically measured blood glucose values for a subsample of the 2009-2012 wave, containing data for over 6000 respondents~~allows identification of~~ , we are bale to identify respondents with undiagnosed diabetes. This allows us to explore measurement error in self-reported diabetes and differences between the effects of self-reported and undiagnosed diabetes on labor outcomes.

~~The~~ Summary stats for the biomarker data indicate that a large share of the sample (27%) have an HbA1c indicative of diabetes (table I), defined by the World Health Organization (WHO) as levels equal to or above 6.5% (World Health Organization, 2011). A second striking observation is the large proportion of false negatives, namely 18% of all observations, implying that 68% of males and females who test positive do not self-report and hence are unaware of their condition. ~~This may lead to biased estimates based on self-reported diabetes.~~

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

3. ~~Estimation strategy~~ESTIMATION STRATEGY

To investigate the relationship between self-reported diabetes and three labor outcomes: employment, wages and weekly working hours, respectively, we estimate the following fixed effects (FE) model.[5]

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

where Y_{it} is a binary variable taking a value of 1 if respondent i reports being in employment at time t and 0 otherwise, $Diabetes_{it}$ is a binary variable taking a value of 1 at time t if the respondent reports having ever received a diagnosis of diabetes[6], X_{it} is a vector of control variables, c_i represents an individual fixed effect, γ_t represents year dummies, while u_{it} is the error term.

For the relationship of diabetes with wages and working hours, our empirical models are estimated conditional on being in employment. Y_{it} represents the log hourly wage or the weekly working hours over the last year, for respondent i at time t.

~~The control variables in both specifications include dummy variables to capture the effects of living in a small, medium or large city~~ We control in detail for levels of urbanization, with rural as the reference category, ~~and state dummies. They also include a marital statusdummy and~~ for state level effects, for marital status, the number of children residing in the household below the age of 6, ~~to control for both the impact of marriage and children. To account for the effect of changes in household wealth on diabetes and employment probabilities, we use standard~~ a quadratic age term and calendar year dummies as well as for household wealth based on an indicator created using principal component analysis of

[5] We also estimated random effects models but do not present them here as the Hausman test suggested the use of the FE model throughout. Results are obtainable upon request.

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

multiple indicators of household assets and housing conditions ~~to create an indicator for household wealth~~ (Filmer & Pritchett, 2001)[7]. ~~The models also include a quadratic age term and calendar year dummies to capture the non-linear effect of age and a time trend, respectively.~~

~~While using individual level does not allow to fully identify a causal relationship, it does improve considerably on existing estimates which are typically obtained from cross-sectional analysis, or from estimation that tend to be weakly identified. The FE model does control~~ Importantly, while the FE model controls for unobserved personal characteristics, ~~although~~ omitted time-variant variables and simultaneity may still affect the relationship of interest. ~~With respect to employment status, one potential concern could be that job loss affects lifestyle choices leading to~~ So could job loss lead to lifestyle choices that cause changes in the probability to develop diabetes~~that~~, which could in turn ~~again~~ affect labor outcomes. ~~Existing~~ While we cannot exclude this possibility, existing work for high-income countries finds no evidence for this kind of reverse causality (Bergemann, Grönqvist, & Gudbjörnsdottir, 2011; Schaller & Stevens, 2015)~~. Another possible channel might be that stress at work leads to a higher propensity of developing type 2 diabetes . However, while stress levels may change over time, a person's coping mechanisms to deal with stress are typically considered to be time-invariant . But while we cannot exclude a role of time-variant unobserved factors or simultaneity, time-invariant variables (including genetic predisposition and stable personality traits) may be more important. The~~ , suggesting that the FE approach ~~should then limit the biasresulting from these time-invariant confounding factors~~will help to considerably reduce any existing bias.

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

3.1. ~~*Labor outcomes and time since diagnosis*~~Labor outcomes and time since diagnosis

In light of the chronic nature and irreversibility of diabetes, there is good reason to explore the long term effects post diagnosis. We estimate the following model:

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

where $\beta_1 Dyears_{it}$ is continuous indicating years since first reported diabetes diagnosis.[8] To capture possible non-linearities in the relationship we also consider a spline function that allows for ~~the effect to vary~~ non-linear effects over time.

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

with $g(Dyears_{it}) = \sum_{n=1}^{N} \delta_n \cdot max\{Dyears_{it} - \eta_{n-1}\} I_{in}$ and $I_{in} = 1[\eta_{n-1} \leq Dyears_{it} < \eta_n]$, with η_n being the place of the n-th node for $n = 1, 2, \ldots, N$. The coefficient δ_n captures the effect of diabetes for the n-th interval. The effects are linear if $\delta_1 = \delta_2 =, \ldots, = \delta_n$.[9] Based on visual inspection (Figure 1 on page 52) we choose three nodes that seem to best account for possible non-linearities in the relationship between diabetes duration and labor outcomes best. These are located at 4, 11 and 20 years after diagnosis. The first four years capture any immediate effects of the diagnosis, the years five to eleven any effects during time of adaptation to the disease and the last term (beyond 11 years) accounts for the long-term effects.

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

[9] Because the year of diagnosis was only reported in the third wave, time since diagnosis is not available for those who were not interviewed in the third round. A reported diagnosis in the year of the interview is counted as 'one year since diagnosis'.

3.2. ~~*Labor outcomes and biometrically measured diabetes*~~Labor outcomes and biometrically measured diabetes

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

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

~~The health information revealed at a diabetes diagnosis may also have an effect in itself. It might, for instance, affect the patient's state of mind which in turn may affect his or her economic decision making and behavior. Two studies found evidence that patients with a diabetes diagnosis and subsequent treatment are more prone to psychological conditions, including depression and anxiety compared to people without diabetes.~~

~~Since undiagnosed diabetes is not found to be associated with psychological conditions, this suggests a possible causal relationship . Health information may also lead to a change in behavior. Slade (2012) shows how patients, when learning about their diabetes diagnosis, change their consumption of alcohol and smoking and start to loose weight. This is in line with evidence for other chronic diseases (see Baird, Gong, McIntosh, and Özler (2014), Gong (2015 Recent evidence for China also suggests that receiving a diabetes diagnosis in itself reduces labor income, possibly through psychological effects of the diagnosis .~~[10]

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

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

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 X_i + c_i + \cancel{u} v_i \tag{4}$$

where v_i are community fixed effects which reflect ~~community~~ local unobserved characteristics such as access to healthcare, poverty and unemployment in the community. ~~These factors are included as they potentially affect the propensity to develop diabetes, to receive a diagnosis, and the labor outcomes of the individuals in the community.~~[10]

[10] ~~Further evidence on the effect of health news on labor outcomes is provided by Dillon, Friedman, and Serneels (2014), who in a very different context, and using a randomized intervention, find that the news stemming from a diagnosis of malaria affect productivity and income, but not labor supply among sugar cane cutters in Nigeria.~~

[10] We did not use household fixed effects since the average number of observations per household was close to one, as most households had only one member providing biomarker information.

In a second step we then estimate the relations between diabetes (as defined by the HbA1c biomarker) and labor outcomes, using the following equation:

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

where $Dbio^d$ is equal to 1 if HbA1c $\geq 6.5\%$.

To estimate the effect of undiagnosed diabetes, in Eq. 6 we add self-reported diabetes back in and interact it with the biomarker.

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i + \beta_3 Dsr_i * Dbio_i + \beta_4 X_i + v_i + u_i. \tag{6}$$

The interaction term changes the interpretation of β_1 and β_2, with β_1 now representing the effect of those aware of their condition but with HbA1c levels below the diabetes threshold and β_2 those with undiagnosed diabetes, i.e. not self-reporting diabetes but with HbA1c levels equal to or above the threshold. The interaction term β_3 shows the effect for those with self-reported diabetes and HbA1c levels above the threshold. We then test if $\beta_1 + \beta_3 = \beta_2$, i.e. if self-reported diabetes is significantly different from undiagnosed diabetes.

~~In a final step we~~ We further investigate the effect of the severity of diabetes on labor outcomes, replacing $Dbio^d$ with $Dbio^c$, a variable that is 0 for ~~HbA1c~~ HbA1c $< 6.5\%$ and takes the actual value of HbA1c for those with an $HbA1c \geq 6.5\%$ (Eq. 7). This will allow us to investigate the effect of a one percentage point increase in HbA1c levels for people with undiagnosed diabetes (β_2) as well as those with self-reported diabetes above the diabetes threshold (β_3).

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i^c + \beta_3 Dsr_i * HbA1c_i + \beta_4 X_i + v_i + u_i. \tag{7}$$

4. ~~Results~~RESULTS

4.1. ~~*Labor outcomes and self-reported diabetes*~~Labor outcomes and self-reported diabetes

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

~~The results in~~ . No significant relationship between wages and working hours is found (Columns 3–6 of Table II~~show no significant relationship between self-reported diabetes and wages or working hours . To assess whether this result differs by the~~).

Because results for wages and working hours may differ by type of work ~~(i.e. work in agriculture requiring strenuous, physical efforts may be more affected by diabetes complications than more sedentary work) we include interaction terms between diabetes and agricultural employment , and between diabetes~~ due to differing physical demands, we interact diabetes with agricultural employment and self-employment, ~~respectively,~~ using non-agricultural wage employment as the comparison group~~(and restricting our sample to those employed only)~~. The results in ~~Table~~ table II panel B ~~, show that the type of work appears not to be a discriminating factor, as none of the interaction terms shows up as significant. In the working hours regression,~~ do not suggest such an effect, however. While the interaction term ~~of diabetes with agricultural work indicates reduced work supply for this group relative~~ for diabetes on work hours in agriculture indicates a reduction compared to non-agricultural ~~workers and employees. However, the Wald test to assess~~

~~the overall significance of the interaction term, cannot reject the null of no interaction effects~~ employment, a Wald test does not indicate a significant difference to agricultural workers without diabetes ($p = .15$).

Table II about here

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

To assess whether diabetes affects the selection into different types of work, we estimate FE models of the probability of being in non-agricultural wage employment, agricultural employment or self-employment, respectively. ~~The results in Table III only indicate statistically significant~~ We only find evidence for negative effects on the probabilities of females to work in agriculture, possibly due to the higher physical requirements ~~.~~(table (III).[11]

Table III about here

4.2. ~~*Labor outcomes and time since diagnosis*~~Labor outcomes and time since diagnosis

Given the chronic ~~(and life-long)~~ nature of diabetes, we investigate how soon after the first diagnosis it impacts labor outcomes. ~~Over time, ever more severe complications tend to develop if diabetes remains poorly treated, and the economic effects may therefore~~

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

~~increase as time passes. These effects may also vary over time as problems that have led to the diagnosis, as well as psychological effects related to the diagnosis may have stronger impacts immediately post diagnosis. Similarly, management of the disease may help delay the onset of complications until after a number of years, reducing health, labor supply and productivity only years after the initial diagnosis.~~

Using non-parametric kernel-weighted local polynomial regression, Figure 1 shows that the probability of employment for men ~~shows a~~ more or less ~~steady decline that becomes more pronounced~~ steadily declines as time progresses. For women, a first drop-off occurs right after diagnosis; thereafter no consistent pattern is observed.[12] ~~A similar analysis for wages shows~~ For wages and working hours somewhat less clear dynamics are observed, with a possibly long term negative trend for women but not for men. ~~An analogous picture is obtained for working hours.~~

Figure 1 about here

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

Table IV about here

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

~~The results for wages indicate a reduction in female wages of~~ Female wages are reduced by about 7% per year after diagnosis in the linear FE model~~. For men we find no consistent~~

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

~~linear effect . Panel B indicates that there may be a reduction in wages~~ , while no effect is found for men. The non-linear results suggest that wages are reduced 5–11 years after the initial diagnosis for both men and women~~. We also find associations for women with~~ , and potentially after more than 20 years ~~of diabetes, but these estimates may be spurious due to the considerably reduced number of observations in this group.~~for women only.[13] Interestingly, ~~the~~ these reductions in wages ~~found in the non-linear specification~~ appear exactly at ~~the time~~ points where employment probabilities are less affected. This may suggest ~~that at this time~~ reductions in productivity ~~affect wages but are not so severe that they would cause~~ being compensated by lower wages rather than job loss. ~~For working hours there appears to be~~ There is no consistent relationship ~~with the~~ of time since diagnosis ~~, neither for men nor for women.~~

~~In summary, the results suggest a fairly constant decrease in the probability of employment for both men and women and in earnings for women, which contrast with estimates for the USA , where no such relationship has been observed. Minor (2013) finds a reduction in employment probabilities of 82 percentage points for females after 11 to 15 years and a reduction of 60 percentage points for males after 2-5 years, indicating very large employment penalties in comparison to our results for Mexico.~~[14] with working hours.

4.3. ~~*Cross-sectional biomarker analysis*~~Cross-sectional biomarker analysis

Table V presents a cross tabulation of self-reported diabetes and biomarker results. Overall, for 80% of the observations the ~~self reports~~ self-reports are consistent with the biomarker

[13] The results for over 20 years may be spurious as they suffer from a low number of observations There are only 9 (3) observations for male (female) wages with more than 20 years since diagnosis in wave 3, and 17 (7) in the pooled sample. For male (female) working hours there are 12 (7) observations with more than 20 years since diagnosis in wave 3, and 20 (12) for the pooled sample.

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

results (diagonal cell %), 18% are false negatives or undiagnosed and 2% false positives~~. Due to the nature of the condition and the possible efficacy of its management the false negatives may well include cases where the person received an actual diabetes diagnosis but consequently achieved blood glucose levels below the diabetes threshold due to successful management of the disease~~ , though the latter may include cases that received a diabetes diagnosis and have managed to reduce their HbA1c levels to non-diabetes levels via medication and/or lifestyle changes (Flores-Hernández et al., 2015). There are no considerable differences between men and women (results not shown).

Table V about here

Table VI presents the results from estimating Eq. 4, Eq. 5, Eq. 6 and Eq. 7. ~~The results in panel~~ Panel A of Table VI ~~show~~ shows that the earlier longitudinal results using self-reported diabetes carry over to the cross-sectional biomarker sample. The coefficients in panel B indicate that the relationship with employment becomes much weaker when using diabetes defined by the biomarker instead of self-reported diabetes, in particular for men. Results in Panel C are obtained from estimating Eq. 6, where we interact self-reported diabetes with biometrically measured diabetes, which allows us to identify the effect of undiagnosed diabetes.

~~The results suggest that there~~ There does not appear to be a statistically significant negative relationship between undiagnosed diabetes (~~in Table VI Panel C~~ expressed in the 'Biomarker diabetes but not self-reported' coefficient) with any labor outcome. The coefficients for the interaction term are negative throughout, though only statistically significant for male wages and female working hours~~, suggesting that being both tested positive and being aware of one's condition has an effect on male wages and especially female working hours, respectively. Only for male employment probabilities, the effects appear to be similar for those with self-reported diabetes regardless whether they are above or below the diabetes threshold. Because the self-reported diabetes group below the threshold is likely comprised of both well controlled people with diabetes as well as~~

~~misreports, the coefficient remains difficult to interpret~~. To establish if the effects of self-reported diabetes ($\beta_1 + \beta 3$) and undiagnosed diabetes (β_2) are significantly different, an F-test is conducted. Overall, we find little evidence of a statistically significant difference, apart from female working hours and potentially male employment probabilities.

Table VI about here

To explore whether the adverse effects increase with ~~diabetes severity, proxied by HbA1c levels , we replace the indicator variable for diabetes with a variable that takes the value zero for levels below and the actual value of HbA1c for those above the diabetes threshold~~higher HbA1c levels we estimate the model detailed in Eq. 7. The results in panel D support the findings from panel C, showing negative coefficients for a 1 percentage point increase in ~~HbA1c~~ HbA1c for those with self-reported diabetes and ~~HbA1c~~ HbA1c levels in the diabetes range, again, however, only statistically significant for male wages and female working hours. ~~No significant effects are found for the undiagnosed.~~

~~Overall, the biomarker results indicate that undiagnosed diabetesis very prevalent in Mexico, but appears to have no statistically significant effects on labor outcomes, in contrast to self-reported diabetes (though the coefficients are not always significantly different from each other in statistical terms)~~For undiagnosed diabetes, again no effects are found.

To further investigate the smaller impact of undiagnosed diabetes, we examine potential mediators~~of the impact of diabetes~~. People self-reporting diabetes may have a different profile in terms of diabetes complications or self-reported health compared to those undiagnosed. First, we therefore include a range of indicators for other chronic diseases that are often related to diabetes. More specifically, we control for being overweight or obese (based on anthropometric measures of body mass index (BMI)) and self-reported hypertension and heart disease. If those diagnosed with diabetes are more likely to experience adverse labor outcomes because they are more likely to suffer from one of these conditions, then accounting for them should lead to a reduction in the coefficient of self-reported diabetes. ~~Table VII, panel A~~Panel A of table VII, however, indicates only small reductions in the coefficients of self-reported diabetes ($\beta_1 + \beta_3$), suggesting no major contributions of these factors to the economic burden of diabetes.

Table VII about here

Controlling for subjective health instead of chronic diseases, the results reported in Table VII, panel B, show a somewhat bigger change in coefficients. The overall effect of self-reported diabetes on the probability of male employment is now comparatively smaller and becomes statistically insignificant. The effect for male undiagnosed diabetes remains close to zero. For women, the coefficient for self-reported diabetes is also reduced, while the effect for undiagnosed diabetes remains the same. We do not observe major changes for all other outcomes.

~~This suggests that the discrepancies in health between the two groups play a role in explaining some of the differences between self-reported and undiagnosed diabetes, with the former being in a somewhat worse health state. However, other unobserved factors appear to be still important.~~

~~It is interesting to contrast these results with those obtained from Brown et al. (2011), one of two other studies that analyze biomarker data. Using data for a Mexican American population the paper finds that once diabetes is diagnosed, the level of blood glucose has little additional effect on labor outcomes. This is similar to our findings and may not be surprising given that HbA1c levels only provide a picture of blood glucose levels over the last three months and may therefore not be representative of blood glucose levels in the longer period before and after the diabetes diagnosis, which determine how soon complications appear and how severe they will be.~~

~~In the same vein, Minor and MacEwan (2016) find for a USA population that people with undiagnosed diabetes experience smaller employment penalties than those self-reporting the condition. Their results suggest, however, considerably bigger effects than we do when~~

~~estimating the impact of biometrically measured diabetes. One possible explanation for the difference is that the undiagnosed population made up a much smaller share of the overall diabetes population compared to our context, and is therefore likely to have a more distinct profile.~~

5. ~~Conclusion~~DISCUSSION

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

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

The paper finds evidence for adverse effects of self-reported diabetes on the probability of being employed, but not on wages or hours worked, using fixed effects estimation. Considering different types of work, the relationship between self-reported diabetes, wages and hours worked remains weak, but the results also suggest occupational selection with women with self-reported diabetes less likely to work in agriculture. These results confirm earlier findings for Mexico regarding the employment impact of diabetes that used cross-sectional information, however, they also suggest a comparatively larger impact of diabetes on female employment probabilities. For wages and working hours they present first evidence.

Analysis of the long term impact suggests that the employment probability falls gradually over the years after having been diagnosed with this chronic condition. Overall, in particular female employment chances and potentially also female wages are reduced. This contrasts with estimates for the USA (Minor, 2013), where only a non-linear relationship has been observed, indicating reductions in employment probabilities for females after 11 to 15 years and after 2-5 years for males.[14]

Making use of the biomarker data allows us to both test reporting error (false negatives) and the effects of undiagnosed diabetes. We find that diabetes based on biomarkers is less related to reduced employment compared to self-reported diabetes, in particular for men. Further analysis shows that this is due to the non-existing relationship of undiagnosed diabetes with employment. For the USA, Minor and MacEwan (2016) find, similar to us, no statistically significant effects of undiagnosed diabetes on employment, while the effect of diagnosed diabetes is significant. Our results further suggest, that the reason for the difference between diagnosed and undiagnosed diabetes are not mediated by current HbA1c levels, but rather overall health status. This is supported by findings for Mexican-Americans in the USA, where no changes in labor outcomes were found with increasing HbA1c levels (Brown et al., 2011). This may not be surprising given that HbA1c levels are only informative for the last three months, and are not the only indicator for the severity of diabetes. They may further be affected by treatment, so that people with already severe diabetes complications and a long diabetes duration may have lower HbA1c levels due to their medical treatment and experience with managing the disease.

Our findings bear several implications. First, the impact of self-reported diabetes on labor outcomes in Mexico is mostly limited to its effect on employment probabilities, though there is some indication that it could also reduce wages over time. Second, its effect on employment is much stronger for females. The reasons for this remain unclear but it could be due to lower wages or working hours for women, making a drop out less costly, or lower formal employment rates, making it easier to discriminate against people

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

with diabetes. Other evidence suggests, that diabetes in women may be more severe due to women being in worse metabolic health compared to men when they cross the diabetes threshold. Third, when interpreting labor market impact estimates relying on self-reported diabetes, one cannot assume that the results extend to those with undiagnosed diabetes. ~~However, combining self-reported and undiagnosed in one diabetes category may not be ideal either, as doing so will fail to account for the heterogeneity between the groups in terms of health information, their actual time of living with diabetes and consequently their subjective as well as true health status, leading to a potentially important loss of information. By contrast, accounting for~~ Studies should therefore, when possible, account for both groups separately, acknowledging their inherent differences, ~~allows~~ to gain information about the distribution of the economic burden across the two groups.

Our results add further weight to the case for reducing the incidence and progression of diabetes. ~~On top of the well-documented health benefits, it appears there are considerable gains to be had by increasing the productive lifespan of people. This is of particular importance in low- and middle-income countries, where parental health shocks, related job loss and increasing health expenditures can have repercussions across the entire household. Other family members, including children, may be forced to increase their labor supply and to reduce non-health expenditures in order to prevent a deterioration of the household's economic situation. This can lead to forgone investments into child education, showcasing the potential for adverse long-term effects of health shocks due to diabetes . Moreover, the~~ The large proportion of previously undiagnosed cases indicates that diagnosis—at least in Mexico—still happens too late or not at all. This reduces the possibilities to prevent complications via treatment and self-management, thereby increasing the risk of severe complications appearing earlier. Hence, much of the health and economic burden may be prevented by earlier diagnosis and ~~, given the generally limited success in achieving good blood-glucose control in Mexico, better~~ ensuing effective treatment of those already diagnosed with diabetes. Further, there is a particular need to explore why women experience such strong economic effects. Ultimately, there is a need to invest in the prevention of diabetes. Taxation of sugar sweetened beverages may be one promising way forward (Colchero, Popkin, Rivera, & Ng, 2016), though the long-term effects remain to be demonstrated. Further, considering the double-disease burden of non-communicable and communicable diseases ~~and malnutrition in many~~ in many low- and middle-income countries (LMICs), investments in maternal and child health may not only reduce the current disease burden but ~~would likely reduce~~ also the future incidence of diabetes, given the established links between early life heath ~~status~~ and later life incidence of diabetes and other chronic diseases (Sotomayor, 2013; Hanson, Gluckman, Ma, Matzen, & Biesma, 2012; Li et al., 2010).

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

Appendix

A Strategies to deal with inconsistent self-reporting over time

Reporting error can pose a considerable challenge in the use of self-reported data. Fortunately, the MxFLS data provide several possibilities to assess the amount of misreporting and apply corrections before estimating the labor market effects of diabetes. In what follows we describe how we have dealt with inconsistencies in self-reported diabetes over time.

Throughout the surveys, self-reported diabetes was measured by the question 'Have you ever been diagnosed by diabetes'. If they answered 'yes', they were asked if they received treatment for diabetes and the type of treatment they received.

One of the key advantages of panel data is the repeated measurement which results in more than one data point allowing to uncover inconsistencies for cases with multiple observations. Very little is known about inconsistencies in self-reported diabetes over time. However, Zajacova, Dowd, Schoeni, and Wallace (2010) assess the consistency of a self-reported cancer diagnosis over time in the USA. The study found that 30% of those who had reported a cancer diagnosis at an earlier point failed to report the diagnosis at a later point in time. A more recent diagnosis was found to be reported with greater consistency possibly due to increasing recall problems as time since diagnosis advanced.

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

As discussed at the end of section 2., for diabetes, the main concern with mismeasurement is related to false negatives. False positives are deemed less of a problem since incentives to report diabetes when one does not have it seem to be very limited—although we cannot exclude this. A study from China finds that the vast majority (98%) of those who self-report diabetes are tested positive for diabetes, while only a minority of those who are tested positive for diabetes (40%) actually self-report the disease (Yuan, Liu, Wu, Zou, & Li, 2015). Our data showed a similar pattern, with a negligible proportion (3%) of the respondents who are tested negative self-reporting to suffer from diabetes, while the majority of those who are tested positive (68%) do not self-report suffering from diabetes.

We used the above information to infer the "true" diabetes status for those with inconsistent reports. For respondents present in all three waves, we corrected inconsistencies as reported in Table VIII. We assumed that if diabetes was reported only once in the first two waves (either in 2002 or 2005) and then not reported again in the ensuing waves, this diabetes report was likely to be false (see lines 3 and 4 in Table VIII) and that the person never had received a diagnosis. If a diabetes diagnosis was however reported in two of the three waves (in 2002 and 2009 but not 2005, or in 2002 and 2005 but not in 2009) we assumed that the respondent had diabetes in all three waves (see lines 1 and 2 in Table VIII). For cases where we only had information from two waves, we assumed that if a diabetes diagnosis had been reported in a prior wave they also had diabetes in the ensuing wave, even if it was not reported in the latter (see lines 5 and 6 in Table VIII), given that most diabetes self-reports tend to be correct.

Table VIII about here

We then tested if those respondents we categorized as not having a diabetes diagnosis based on above rules were actually more likely to not have diabetes, using the biomarker data from wave 3. Of those with inconsistencies in their diabetes self-reports, 95 were present in the biomarker sample (46 with two self-reports (from lines 3 and 4 in Table VIII) and 49 with one self-report of diabetes (from lines 1 and 2 in Table VIII)). Figure 2 illustrates the difference between both groups and suggests that indeed those with two self-reports of diabetes are much more likely to have HbA1c values above the diabetes threshold. A t-test comparing the mean HbA1c for the two groups indicates that those with two self-reports also have significantly ($p<0.001$) higher HbA1c levels than those with only one self-report of diabetes (9.7% vs. 7.1%). Further, of those with one self-report, only 30% have an HbA1c$\geq$ 6.5% compared to 87% of those with two self-reports. Based on these results it appears that we did minimize misclassification of people into diabetes or no-diabetes.

Alternatively we also test if using an alternative strategy, i.e. assuming that everybody who reported a diabetes diagnosis once had diabetes in any later wave, would lead to different estimation results. We do not find this to be the case and find only minor differences in the point estimates of the coefficients (results available on request).

Figure 2 about here

B Studies on diabetes and labor market outcomes

Table IX about here

References

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

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

Barquera, S., Campos-Nonato, I., Aguilar-Salinas, C., Lopez-Ridaura, R., Arredondo, A., & 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. (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. (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.

Bello-Chavolla, O. Y., Rojas-Martinez, R., Aguilar-Salinas, C. A., & Hernández-Avila, M. (2017). Epidemiology of diabetes mellitus in Mexico. *Nutrition Reviews*, *75*(suppl 1), 4–12.

Bergemann, A., Grönqvist, E., & Gudbjörnsdottir, S. (2011). The effects of job displacement on the onset and progression. *Netspar Discussion Paper*, (25).

Brown, H. S., Pagán, J. A., & Bastida, E. (2005). The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit. *Health Economics*, *14*(5), 537–544.

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

Brown, T. T. (2014). How effective are public health departments at preventing mortality? *Economics & Human Biology*, *13*, 34–45.

Colchero, M. A., Popkin, B. M., Rivera, J. A., & Ng, S. W. (2016). Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. *British Medical Journal*, *352*, h6704.

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

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

Dillon, A., Friedman, J., & Serneels, P. M. (2014). Health information, treatment, and worker productivity: Experimental evidence from malaria testing and treatment among Nigerian sugarcane cutters. *World Bank Policy Research Working Paper*, (7120).

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.

Flores-Hernández, S., Saturno-Hernández, P. J., Reyes-Morales, H., Barrientos-Gutiérrez, T., Villalpando, S., & Hernández-Ávila, M. (2015). Quality of Diabetes Care: The Challenges of an Increasing Epidemic in Mexico. Results from Two National Health Surveys (2006 and 2012). *Plos One*, *10*(7), e0133958.

Gong, E. (2015). HIV Testing and Risky Sexual Behaviour. *The Economic Journal*, *125*(582), 32–60.

Gregg, E. W., Chen, H., Wagenknecht, L. E., Clark, J. M., Delahanty, L. M., Bantle, J., . . . Bertoni, A. G. (2012). Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes. *Journal of the American Medical Association*, *308*(23), 2489.

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

International Diabetes Federation. (2015). *Diabetes Atlas* (7th ed.). International Diabetes Federation.

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

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

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

Lim, E. L., Hollingsworth, K. G., Aribisala, B. S., Chen, M. J., Mathers, J. C., & 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. (2014). Will knowing diabetes affect labor income? Evidence from a natural experiment. *Economics Letters*, *124*(1), 74–78.

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

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

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

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

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. (2015). Short-run effects of job loss on health conditions, health insurance, and health care utilization. *Journal of Health Economics*, *43*, 190–203.

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

Seuring, T., Goryakin, Y., & Suhrcke, M. (2015). 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.

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

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.

Williams, A. L., Jacobs, S. B. R., Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., … Tusié-Luna, T. (2013). 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.

Yuan, X., Liu, T., Wu, L., Zou, Z.-Y., & Li, C. (2015). Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study. *British Medical Journal Open*, *5*(4), e006633–e006633.

Zajacova, A., Dowd, J., Schoeni, R. F., & 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.

Table I. Descriptive statistics for panel and biomarker sample.

	Panel		Biomarker	
	Males	Females	Males	Females
Dependent variables				
Employed	0.86	0.37	0.86	0.34
	(0.34)	(0.48)	(0.35)	(0.47)
Hourly wage (Mexican Peso)	42.47	40.49	36.30	35.23
	(485.87)	(142.08)	(53.69)	(43.63)
Weekly working hours	46.82	38.99	46.00	38.15
	(16.79)	(18.90)	(16.89)	(19.65)
Agricultural worker	0.22	0.04	0.25	0.03
	(0.41)	(0.20)	(0.43)	(0.18)
Self-employed	0.19	0.28	0.21	0.32
	(0.39)	(0.45)	(0.41)	(0.47)
Non-agricultural worker or employee	0.59	0.68	0.53	0.64
	(0.49)	(0.47)	(0.50)	(0.48)
Diabetes variables				
Self-reported diabetes	0.05	0.06	0.09	0.12
	(0.22)	(0.24)	(0.29)	(0.32)
Diabetes duration if self-reported diabetes (years)	7.49	7.83	7.48	7.99
	(6.01)	(7.83)	(6.07)	(7.03)
Glycated hemoglobin (HbA1c)			6.46	6.58
			(1.89)	(2.02)
HbA1c $\geq$ 6.5%			0.26	0.28
			(0.44)	(0.45)
Undiagnosed diabetes			0.18	0.18
			(0.39)	(0.39)
Control variables				
Age	36.03	36.29	42.78	42.79
	(13.62)	(13.17)	(14.28)	(13.94)
Rural village of $<$ 2,500	0.44	0.43	0.50	0.46
	(0.50)	(0.50)	(0.50)	(0.50)
Married	0.54	0.54	0.60	0.56
	(0.50)	(0.50)	(0.49)	(0.50)
Number of children (age $<$ 6) in household	1.48	1.57	1.18	1.22
	(1.45)	(1.47)	(1.29)	(1.32)
Indigenous group	0.19	0.19	0.19	0.18
	(0.39)	(0.39)	(0.39)	(0.39)
Education				
Secondary	0.30	0.30	0.26	0.26
	(0.46)	(0.46)	(0.44)	(0.44)
High school	0.16	0.13	0.14	0.12
	(0.36)	(0.34)	(0.34)	(0.33)
Higher education	0.11	0.09	0.12	0.09
	(0.32)	(0.29)	(0.32)	(0.28)
Wealth index	−0.00	−0.02	0.09	−0.00
	(1.02)	(1.00)	(1.07)	(1.02)
N	21388	273410	2785	3623

Notes Mean values, standard deviations in parenthesis.

Table II. Labor outcomes and self-reported diabetes

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

Notes Individual fixed effects regression. Robust standard errors in parentheses. Reference category: dependent non-agricultural worker or employee. All models include variables for states, urbanization level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table III. Selection into types of work and self-reported diabetes

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

Notes Individual fixed effects regression. Robust standard errors in parentheses. All models include variables for states, urbanization level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. The hausman test for female agricultural employment could not be calculated due to the random effects model results being equivalent to pooled OLS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

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

Notes The table presents the results of three estimation methods. Panel A presents the results of the linear specifications. Panel B presents the results of the non-linear specifications. Robust standard errors in parentheses. All models include variables for states, urbanization level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year. The OLS and pooled OLS models additionally control for age. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table V. Number of observations with diabetes (HbA1c $\geq$ 6.5%) and self-reported diabetes.

	$HbA1c < 6.5\%$	HbA1c $\geq$ 6.5%	Total
No self-reported diabetes (N)	4544	1181	5725
Row %	79%	21%	100%
Column %	97%	68%	89%
Cell %	**71%**	**18%**	-
Self-reported diabetes (N)	129	554	683
Row %	19%	81%	100%
Column %	3%	32%	11%
Cell %	**2%**	**9%**	-
Total (N)	4673	1735	6408

Table VI. Biomarker results

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Panel A: Diabetes (self-reported)						
Self-reported diabetes	$-.051^{**}$	$-.044^{*}$	$-.010$	$-.040$	$-.293$	$-.751$
	(.026)	(.023)	(.065)	(.113)	(1.305)	(2.178)
Panel B: Diabetes (biomarker)						
Biomarker diabetes (HbA1c $\geq$ 6.5)	$-.012$	$-.031^{*}$	$-.007$	$-.057$	$-.088$	1.153
	(.016)	(.018)	(.044)	(.070)	(.844)	(1.462)
Panel C: Interacting self-reported and biomarker diabetes						
Self-reported diabetes but tested negative (β_1)	$-.030$	$-.001$	0.328^{*}	$-.002$	1.756	6.183
	(.056)	(.050)	(.192)	(.226)	(3.248)	(4.356)
Biomarker diabetes but not self-reported (HbA1c $\geq$ 6.5) (β_2)	0.006	$-.017$	0.017	$-.054$	0.168	2.577
	(.018)	(.020)	(.050)	(.078)	(.960)	(1.640)
Self-reported diabetes and biomarker diabetes (β_3)	$-.029$	$-.042$	$-.396^{*}$	$-.010$	-2.511	-10.883^{**}
	(.062)	(.058)	(.209)	(.259)	(3.594)	(5.153)
All self-reported ($\beta_1 + \beta_3$)	$-.059^{**}$	$-.043$	$-.068$	$-.012$	$-.755$	-4.700^{*}
	(.029)	(.031)	(.084)	(.136)	(1.570)	(2.777)
F-test (p-value): $\beta_1 + \beta_3 = \beta_2$	0.111	0.564	0.462	0.818	0.674	0.056
Panel D: HbA1c levels						
Self-reported diabetes	$-.050$	$-.013$	0.223^{*}	0.029	1.650	3.464
	(.065)	(.041)	(.117)	(.178)	(2.504)	(3.527)
HbA1c if $\geq$ 6.5	0.001	$-.003$	0.002	$-.005$	$-.005$	0.256
	(.002)	(.002)	(.005)	(.008)	(.104)	(.192)
Self-reported diabetes $\times$ HbA1c if $\geq$ 6.5	$-.001$	$-.002$	$-.029^{**}$	$-.005$	$-.231$	$-.746^{*}$
	(.007)	(.005)	(.014)	(.022)	(.283)	(.408)
N	2785	3623	1803	884	2302	1144

Notes Results are based on community level fixed effects. Robust standard errors in parentheses. All models include variables for states, urbanization level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year to account for the multiple years of data collection for the third wave. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural wage employment as the base). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Panel A: Controlling for other chronic diseases						
Self-reported diabetes but tested negative (β_1)	$-.016$	0.008	0.340*	0.033	2.021	6.769
	(.056)	(.050)	(.195)	(.227)	(3.277)	(4.390)
Biomarker diabetes but not self-reported (HbA1c ≥ 6.5) (β_2)	0.004	$-.015$	0.013	$-.036$	0.124	2.372
	(.018)	(.021)	(.050)	(.079)	(.962)	(1.649)
Self-reported diabetes and biomarker diabetes (β_3)	$-.035$	$-.050$	$-.398$*	0.030	-2.662	-11.520**
	(.062)	(.058)	(.210)	(.260)	(3.609)	(5.173)
All self-reported ($\beta_1 + \beta_3$)	$-.051$*	$-.043$	$-.058$	0.063	$-.640$	-4.752*
	(.029)	(.031)	(.085)	(.138)	(1.584)	(2.807)
F-test: $\beta_1 + \beta_3 = \beta_2$	0.180	0.535	0.548	0.592	0.729	0.063
N	2785	3621	1803	882	2300	1142
Panel B: Controlling for self-reported health						
Self-reported diabetes but tested negative (β_1)	$-.013$	0.007	0.353*	0.043	2.189	4.672
	(.056)	(.050)	(.193)	(.225)	(3.257)	(4.374)
Biomarker diabetes but not self-reported (HbA1c ≥ 6.5) (β_2)	0.005	$-.018$	0.020	$-.052$	0.116	2.728*
	(.018)	(.020)	(.050)	(.078)	(.961)	(1.639)
Self-reported diabetes and biomarker diabetes (β_3)	$-.028$	$-.041$	$-.416$**	0.024	-2.545	-9.725*
	(.062)	(.058)	(.210)	(.259)	(3.604)	(5.198)
All self-reported ($\beta_1 + \beta_3$)	$-.041$	$-.034$	$-.064$	0.066	$-.356$	-5.053*
	(.029)	(.031)	(.085)	(.139)	(1.586)	(2.831)
F-test: $\beta_1 + \beta_3 = \beta_2$	0.256	0.719	0.473	0.521	0.831	0.043
N	2785	3621	1803	883	2302	1143

Notes Community level fixed effects. Robust standard errors in parentheses. All models include variables for states, urbanization level of education, marital status, number of children < 6, wealth, health insurance status, age squared and one dummy variable for each calendar year to account for the multiple years of data collection for the third wave. The wage and working hour models additionally control for type of work (agricultural and self employed with non-agricultural wage employment as the base). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VIII. Inconsistencies in diabetes self-report in MxFLS.

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

Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
China	2009, 2011	Employed population	Panel	HbA1c	Find a significant reduction of 16.3 % in income for those with a recent diagnosis in China.	NA	Use difference-in-difference model, exploiting a recent diagnosis of diabetes as a result of biomarker collection within the used survey, as a natural experiment to measure how income developed between those who were newly diagnosed and those without diabetes in the years following diagnosis	Liu and Zhu (2014)
Mexico	2005	Working age population	Cross-section	Self reported	A significant ($p<0.01$) reduction in employment probabilities for males by about 10 % points and for females by about 4.5 % points ($p<0.1$)	Diabetes exogenous for men and women based on Hausman test ($p >.10$)	Probit and bivariate probit model using parental diabetes as IV	Seuring, Goryakin, and Suhrcke (2015)

Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	1996-1997	Elderly population of Mexican Americans close the Mexican border	Cross-section	Self reported	Significant adverse relationship, with 7 % points lower employment rates for men - for women, the negative relationship becomes insignificant when using instrumental variable (IV) estimation	Diabetes endogenous for women but not men based on Hausman test	Bivariate probit	Brown et al. (2005)
USA	2008	Mexican-American working age adults	Cross-section	HbA1c levels	Find a negative relationship between HbA1c levels and the probability of employment as well as male wages. No effects found for women.	Exogeneity assumed	Probit and Heckman selection model	Brown et al. (2011)
USA	2006	Women 20 - 65	Cross-section	Self reported	Exogenous: 25.2 % points less likely to be employed, endogenous: 45.1 % points less likely to be employed.	Self-reported diabetes endogenous and estimates upward biased compared to IV estimates	Probit and Heckman selection model; unclear which model is used for IV estimates	Minor (2011)

Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	1979 - 2010	Follows young adults in 1979 throughout their adult life	Panel	Self-reported year of diagnosis	Average reduction of employment probability of 28 % points for men and 36 % points for women; employment probabilities decline shortly after diagnosis for men and after about 10 years for women, while wages are not affected by the duration of diabetes	Exogeneity assumed	Uses sibling and job fixed effects model (no individual fixed effects) using logit model for selection into employment and ordinary least squares for wages	Minor (2013)

Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
USA	2001 - 2008	Men and women 18 - 65	Panel	Self-reported and HbA1c levels for subsample	No statistically significant relationship between undiagnosed diabetes and the probability of employment. Self-reported diabetes significantly related with lower employment probabilities for men (-11 % points) and women (-19 % points). Using only biomarker information (HbA1c >6.4 %), statistically significant reductions in employment probabilities for men (-8.3 % points) and women (-11 % points). No significant effects of undiagnosed diabetes on hours worked. Increase in HbA1c by 1 % point related to 1.3 % points lower employment probabilities for men. No effect for women.	Exogeneity assumed	Probit model for binary outcomes, OLS for continuous outcomes; all applied to pooled data	Minor and MacEwan (2016)

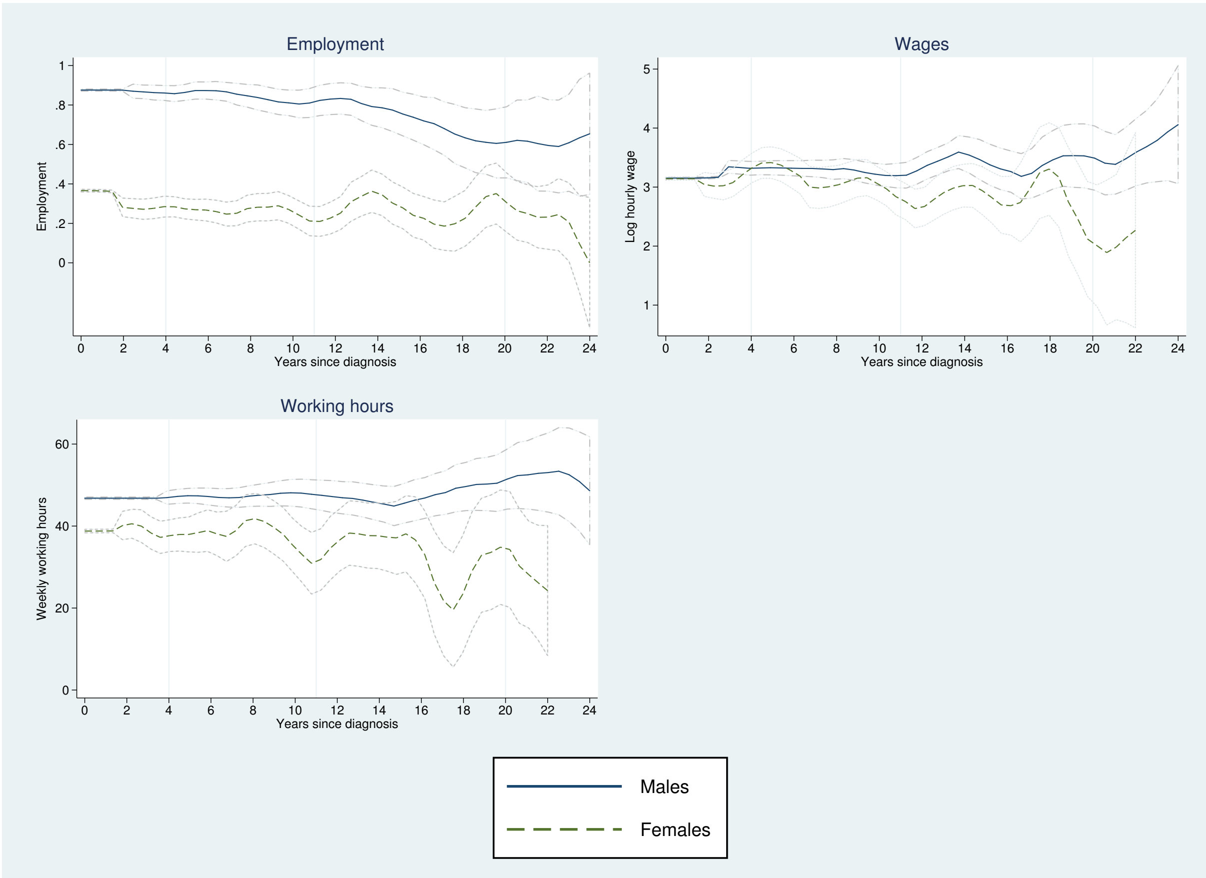
Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
Canada	1998	Men and women 15 - 64	Cross-section	Self reported	For men: Exogenous 19 % points less likely to be employed; endogenous: not significant and positive; test indicates endogeneity For women: Exogenous: 17 % points less likely to be employed; endogenous: not significant and positive and test indicates exogeneity	Diabetes endogenous for men, resulting in upwards biased estimates; exogenous for women	Instrumental variable strategy using bivariate probit model and family history of diabetes as the instrument	Latif (2009)

Table IX. Studies estimating the relationship between diabetes and labor market outcomes

Country	Year	Population	Panel or cross-sections	Measurement of diabetes	Main finding	Finding on bias due to endogeneity	Estimation method	Reference
Australia	1999 - 2000	Men and women age >24	Cross-section	Self reported	Reduced labor market participation for men (-7.1 % points) and women (-9 % points) as a result of diabetes, with the effects appearing overstated (-10.8 % points for men and -10 % points for women) if the endogeneity of diabetes is unaccounted for	Overestimation if endogeneity unaccounted for	Endogenous multivariate probit model	Zhang et al. (2009)

Figure 1: Employment, wages, working hours and years since self reported diabetes: Kernel-weighted local polynomial regression



Notes The dashed lines show 95% confidence intervals.

Figure 2: Kernel density of HbA1c values for those with one inconsistent and two inconsistent reports.