Thesis submitted for the degree of Doctor of Philosophy (PhD)

The Economics of Diabetes in Middle-Income-Countries

Till Seuring

School of Medicine
University of East Anglia, UK

date

Abstract

This thesis focuses on the economic analysis of type 2 diabetes (T2D) in middle-income countries. Given its rising prevalence, in-depth country specific analysis is key for understanding the economic consequences of T2D in middle-income countries (MICs). I analyse the economic burden of T2D in terms of labour market consequences, taking into account the heterogeneity of the diabetes population, for both Mexico and China. For China I further investigate the effects of a diabetes diagnosis on health behaviours that may help to curb the adverse consequences of diabetes.

The thesis consists of four essays with the unifying theme of improving our understanding of the causal relationship between diabetes and economic outcomes. Essay (1) provides an updated overview, critically assesses and identifies gaps in the current literature on the economic costs of T2D using a systematic review approach; essay (2) studies the effect of self-reported diabetes on employment probabilities in Mexico, using cross-sectional data and making use of a commonly used instrumental variable approach; essay (3) extends the previous essay via the use of panel data and fixed effects and considering a broader range of outcomes, including wages and working hours; it also makes use of cross-sectional biomarker data that allows for the investigation of measurement error in self-reported diabetes; essay (4) investigates the effect of a diabetes diagnosis on employment and income as well as health behaviours in China, using longitudinal data and applying two distinct identification strategies: fixed effects and marginal structural model estimation.

The findings of the first paper document a considerable increase in studies on the economic costs of diabetes in MICs. It also illustrates that most of the evidence is based on cost-of-illness studies and the literature on labour market and potential earning effects of diabetes in MICs is scarce. The thesis fills part of this void and shows that self-reported diabetes has a considerable impact on employment probabilities of people living in Mexico and China. The findings are robust to the application of different estimation strategies. No consistent evidence of an adverse effect of diabetes on wages or working hours is found, suggesting that diabetes mainly affects the extensive margin. The findings for Mexico indicate that particularly people working in the informal or agricultural, hence less protected and often more physically demanding, sectors bear the brunt of the negative effects of diabetes. Taking into account the undiagnosed population, the adverse effect of diabetes is reduced because undiagnosed diabetes itself does not show an adverse association with any labour market outcome. This suggests that the undiagnosed population is distinctly different from the diagnosed population, likely due to differences in health information and health status. Therefore, research using self-reported diabetes information should limit its claims to the diagnosed population as economic effects are likely different for the undiagnosed. With regards to the effect of a diabetes diagnosis on

health behaviours, the results from China suggest that a diagnosis leads to moderate reductions in body mass index (BMI), waist circumference, alcohol and caloric consumption. Perhaps surprisingly, especially men appear to be able to lose weight and reduce their caloric consumption. Not accounting for unobserved heterogeneity leads to a change in the coefficient sign for the effect of a diagnosis on BMI and waist circumference, while the differences in estimates are less pronounced for other outcomes.

Contents

List of Figures

List of Tables

Abbreviations

ATE average treatment effect

BMI body mass index

COI cost-of-illness

FE fixed effects

GDP gross-domestic-product

HbA1c glycated hemoglobin

HIC high-income country

ICD International Statistical Classification of Diseases and Related Health Problems

IDF International Diabetes Federation

IV instrumental variable

LATE local average treatment effect

LMIC low- and middle-income country

LPM linear probability model

MIC middle-income country

MxFLS Mexican Family Life Survey

NCD non-communicable disease

OOP out-of-pocket

p.p. percentage points

PPP purchasing-power-parity

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RE random effects

UK United Kingdom

WHO World Health Organization

WTP willingness to pay

Publications and statement of authorship

Publications arising from this thesis

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

This publication can be found at: `http://link.springer.com/article/10.1007%2Fs40273-015-0268-9`

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

This publication can be found at: `http://www.sciencedirect.com/science/article/pii/S1570677X15000349`

Statement of jointly authored publications

The research reported is my own original work which was carried out in collaboration with others as follows:

Chapter 1: Written by Till Seuring.

Chapter 2: Till Seuring was the lead author of a paper published as:

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

Till Seuring, Marc Suhrcke and Olga Archangelidi designed the study. The search strategy was designed and executed by Till Seuring. Till Seuring and Olga Archangelidi screened the initial results and extracted the data from the primary studies. Till Seuring drafted the original manuscript which was critically reviewed by Olga Archangelidi and Marc Suhrcke.

Chapter 3: Till Seuring was the lead author of a paper published as:

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

Till Seuring, Yevgeniy Goryakin and Marc Suhrcke designed the study. Till Seuring analysed the data. Till Seuring drafted the original manuscript which was critically reviewed by Yevgeniy Goryakin and Marc Suhrcke.

Chapter 4: Till Seuring, Pieter Serneels and Marc Suhrcke designed the study. Till Seuring analysed the data. Till Seuring drafted the original manuscript which was critically reviewed by Pieter Serneels and Marc Suhrcke.

Chapter 5: Till Seuring and Max Bachmann designed the study. Till Seuring analysed the data. Till Seuring drafted the original manuscript which was critically reviewed by Max Bachmann.

Chapter 6: Written by Till Seuring.

1 General Introduction

- Set stage describing burden of chronic disease/diabetes in world and MICs (Mexico/China) more specifically. (e.g. burden of disease study/high level studies).

- Describe general goal of thesis:

  Identify gaps in literature on the economic burden of diabetes in terms of evidence but also methodology, particularly in MICs, and fill some of the gaps.

- Describe each of the chapters and the motivation behind it

2 The Economic Costs of Type 2 Diabetes: A Global Systematic Review

Abstract

Background There has been a widely documented and recognized increase in diabetes prevalence not only in high-income countries (HICs) but also in low- and middle-income countries (LMICs), over recent decades. It is less clear what is the economic burden associated with diabetes, especially in LMICs. Objective We provide a systematic review of the global evidence on the costs of type II diabetes. Our review seeks to update and considerably expand the previous major review of the costs of diabetes by capturing the evidence on overall, direct and indirect costs of type II diabetes worldwide that was published since 2001. In addition we include a body of economic evidence that has hitherto been distinct from the COI work, i.e. studies on the labour market impact of diabetes. Methods PubMed, EMBASE, EconLit and IBSS were searched (without language restrictions) for studies assessing the economic burden of type 2 diabetes published from January 2001 to October 2014. Costs reported in the included studies were converted to international dollars ($) adjusted for 2011 values. Alongside the narrative synthesis and methodological review of the studies we conduct an exploratory linear regression analysis, examining the factors behind the considerable heterogeneity in existing cost estimates between and within countries. Results We identified 86 COI and 23 labour market studies. COI studies varied considerably in both methods and cost estimates, with most studies not using a control group, though the use of either regression analysis or matching has increased. Direct costs were generally found to be higher than indirect costs. Direct costs ranged from $242 for a study on out-of-pocket (OOP) expenditures in Mexico to $11917 for a study on the cost of diabetes in the USA, while indirect costs ranged from $45 for Pakistan to $16914 for the Bahamas. In LMICs —in much contrast to HICs— a substantial part of the cost burden arose to patients from OOP treatment costs. Our regression analysis revealed that direct diabetes costs are closely and positive associated with a country's gross domestic product (GDP) per capita, and that the USA stood out as having particularly high costs, even after controlling for GDP per capita. Studies on the labour market impact of diabetes were almost exclusively confined to HICs and found strong adverse effects, particularly for male employment chances. Many of these studies also took into account the possible endogeneity of diabetes, which was not the case for COI studies. Conclusions The reviewed studies indicate a large economic burden of diabetes, most directly affecting patients in LMICs. The magnitude of the cost estimates differs considerably between and within countries, calling for the contextualization of the study results. There remains large scope for adding to the evidence base on labour market effects of diabetes in LMICs. Further, there is a need for future COI studies to incorporate more advanced statistical methods in their analysis to account for possible biases in the estimated costs.

2.1 Introduction

Diabetes is a chronic disease that has spread widely, not only in high-income but also in many LMICs over the last decades. The most recent data from the International Diabetes Federation indicate that diabetes affected 382 million people worldwide in 2013, a number that is expected to grow to 592 million by 2035. The estimated global prevalence in 2013 amounts to 8.3 % among people aged 20–79 years, with the world's most populous countries India and China reaching prevalence rates between 9 and 10 %, corresponding to 65 and 100 million in absolute numbers, respectively. Particularly high prevalence rates are found in Mexico (12.6%) and Egypt (16.8%), surpassing the rates of most HICs, including the USA (9.2%) and Germany (8.2%).(International Diabetes Federation, 2014) Taken together, in 2013 about two-thirds of all individuals with diabetes lived in LMICs (International Diabetes Federation, 2014). The rising prevalence of diabetes in LMICs appears to be fuelled by rapid urbanization, nutrition transition and increasingly sedentary lifestyles (Hu, 2011b). The most prevalent form of diabetes by far is type 2 diabetes, affecting about 90 % of people with diabetes while the remaining 10 % mainly have type 1 diabetes or gestational diabetes (International Diabetes Federation, 2014).

Due to its adverse effect on people's health diabetes also imposes an economic burden on individuals and households affected as well as on healthcare systems. The economic burden of diabetes was confirmed by in a review of COI studies on diabetes mellitus, published in 2004, covering the literature up to the year 2000. The authors concluded that the direct and indirect economic burden of diabetes was "large", and that costs had increased over time. However, the review also noted that significant variation in costing methodologies made it near impossible to directly compare the cost estimates. However, the studies reviewed by Ettaro et al. (2004) were almost exclusively focused on the USA, with a small part coming from European HICs and none from LMICs. The aim of this study is therefore to systematically review the literature on the economic costs of diabetes published since 2001 (i.e. the first year not covered by the Ettaro et al. (2004) review), as we expect a considerable number of new studies, also from LMICs. In addition to the COI studies we review the literature on labour market outcomes, with a specific interest in the methodological challenges involved. In doing so we substantively update and expand the scope of the Ettaro et al. (2004) review, allowing us to revisit its findings regarding the evidence base about the economic burden of type 2 diabetes globally.

COI studies generally assess the direct and indirect costs of a particular illness, where the former represent the opportunity cost of resources used for treatment. The indirect costs measure the value of resources lost due the illness, most commonly those caused by losses in productivity due to mortality and morbidity as measured in lost earnings (Segel, 2006). In addition, another approach also focuses on estimating the impact of diabetes on labour market outcomes. However, rather than trying to estimate the monetary losses that arise from a decrease in productivity, these studies typically compare labour market outcomes (e.g. employment probabilities, earnings or lost work days) between people with and without diabetes, while accounting for differences in age, education and other demographic and socioeconomic variables, that might arise between both groups and that could affect labour market outcomes as well as the chances of developing diabetes. The aim of studies in this field is to obtain a clearer picture of how diabetes causally affects these labour market outcomes, without necessarily monetizing the results. Because of the different methodologies and data requirements, these studies tend to differ considerably from traditional COI studies, which is why we reviewed them separately. To the best of our knowledge this is the first review that systematically assesses the studies in this particular field.

2.2 Methods

PRISMA guidelines were used as a basis for the overall study approach.(Moher et al., 2009)

2.2.1 Search Strategy

The electronic search was based on the following search terms: "Diabetes Mellitus"[Mesh] AND ("Costs and Cost Analysis"[Mesh] OR "Cost of Illness"[Mesh] OR "Employment"[Mesh] OR "Labor Market"[All fields] OR "Labour Market"[All fields] OR "Productivity"OR "Willingess to pay"[All fields]). The above search was run in PubMed and was then adapted for searches in EMBASE, EconLit and the International Bibliography of the Social Sciences (IBSS). The search was carried out from October 2012 to October 2014 and restricted to studies published between January 2001 and October 2014, as the earlier review had covered COI studies until 2000 (Ettaro et al., 2004). No language restrictions were applied. The references were downloaded in RIS format where possible and then transferred to Mendeley. Authors were contacted for further information if clarification was needed after the full text analysis.

2.2.2 Inclusion and Exclusion Criteria

Studies were eligible if a monetary estimate of the direct and/or indirect costs of diabetes was presented in the results section or if studies provided an estimate of the impact of diabetes on labour market outcomes (employment chances, labour income, wages and lost work days). We did not exclude studies with a small sample size as this might have discriminated against studies in LMICs. Studies on types of diabetes explicitly different from type 2 diabetes were excluded. However, we included studies that did not explicitly mention the type of diabetes, given that type 2 diabetes accounts for about 90 % of all diabetes cases. Studies exclusively assessing the costs of diabetes complications or the costs of management strategies were excluded as were studies estimating the costs for specific groups with diabetes (e.g. costs for people with poorly controlled diabetes), since we were interested in the costs incurred to populations comprising the whole spectrum of people with type 2 diabetes. Editorials, reviews and studies for which the full text could not be retrieved or only an abstract was available were also excluded.

2.2.3 Data Extraction and Analysis

Data extraction was carried out by two investigators (TS and OA). After duplicates were removed, titles and abstracts were scanned by one researcher (TS) to identify studies suitable for a full text review. The process was checked by a second researcher (OA) on a random subsample of 2000 studies of the retrieved references. The full text was subsequently retrieved for the identified studies and they were reviewed by two researchers (TS and OA), with disagreements resolved by discussion. Finally, 109 studies were identified (see Figure 2.1) that fulfilled the inclusion criteria and data extraction was carried out using a pre-defined extraction table. Primary outcomes were the total costs, the direct costs, and the indirect costs of type 2 diabetes and the respective per capita estimates of these outcomes, as well as the impact of type 2 diabetes on employment chances, income, wages and lost work days. Secondary outcomes comprised the methodology used to assess the monetary costs of type 2 diabetes, the range of cost factors included in the analysis, as well as the methodology used to assess the labour market impact of diabetes. Further extracted information included the year of publication, year of data collection, the time horizon, the country or region studied, the data source, sample size and age as well as information on whether the study distinguished between types of diabetes.

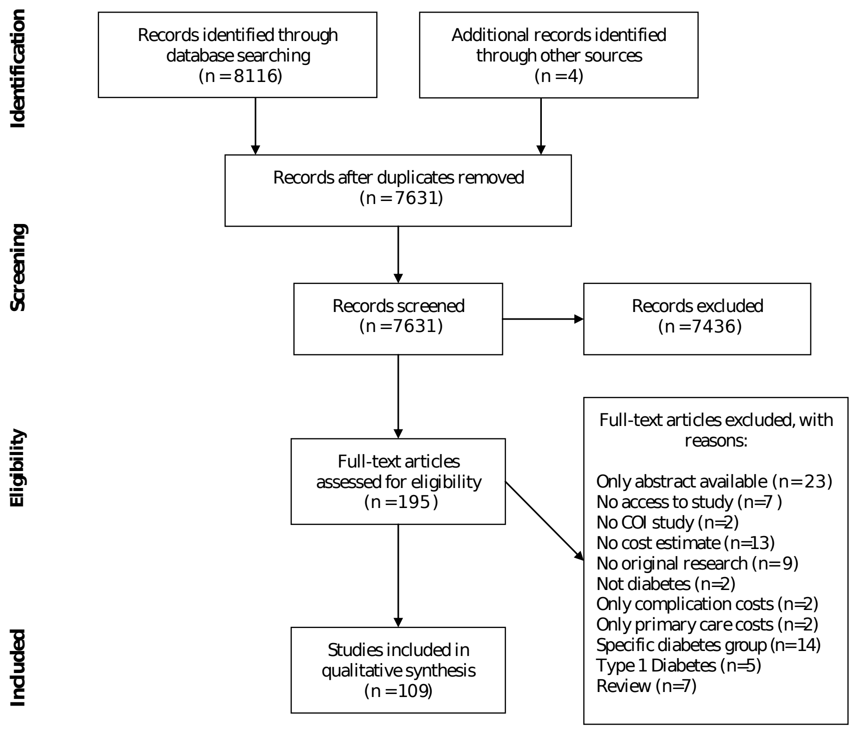
We present the COI study results in per capita values to facilitate comparability across countries. For studies presenting overall population level estimates rather than per capita costs information, we calculated those costs, whenever possible, using the diabetes prevalence mentioned in the respective study. If no total cost estimate was presented but information on direct and indirect costs was available, then direct and indirect costs were added up to produce a total cost estimate. We converted costs into purchasing-power-parity (PPP) adjusted estimates, also called international dollars and henceforth denoted with the $ sign, in order to further increase comparability. Since some studies did not present the data in the country's local currency but in USA$ or some other major currency, we used the exchange rate given in the article to convert the estimates back into the local currency. If no exchange rate was provided in the study itself, the average exchange rate (midpoint exchange rate according to OANDA historical exchange rates - [http://www.oanda.com/currency/historical-rates/]) for the reported year. The PPP adjusted estimates for the year 2011 were then calculated using the Campbell and Cochrane Economics Methods Group Evidence for Policy and Practice Information and Coordination Centre (CCEMG-EEPPI Centre) cost converter (Shemilt et al., 2010). For all additional analyses carried out in the following sections only studies for which a mean cost estimate was presented or could be calculated, were included. Further, in the case of a study presenting estimates for more than 1 year, only the estimate for the most recent year was used for the analysis. For studies presenting both incremental and total cost estimates, only the incremental cost estimate was taken into account.

Studies were further classified into two groups according to the level of economic development of the investigated country—(1) high-income and (2) LMICs (LMICs)—according to the historical World Bank income group classification of the respective country in the year that data collection for the respective study had taken place (World Bank, n.d.). Where necessary due to space constraints we used abbreviations for country names, as detailed in Table 2.7.

In order to explore the factors involved in the variation of direct costs reported in COI studies, we first plotted the direct per capita costs in relation to the GDP per capita of the respective country and provided an estimate of the relationship using linear regression. We then conducted an exploratory regression analysis, with the annual direct cost per patient as the dependent variable to investigate what other factors might explain the variation in direct cost estimates. The set of independent variables comprised (1) the estimation approach in each study, (2) the year of data used, (3) GDP per capita of the studied country in international dollars, (4) an indicator of whether the study was conducted in the USA, (5) an indicator of whether the study was deemed to be nationally representative, and (6) a variable indicating whether the study had explicitly taken diabetes-related complications into account. The year of the used data was considered because the development of social security systems and treatment methods may affect how the direct costs evolve over time.

We categorized this variable into groups: studies using data from before 1995, 1995 to 1999, 2000 to 2004, 2005–2009 and 2010–2004. The dummy variable for studies on the USA was included to account for the generally higher healthcare expenditures in the USA compared which other HICs with similar per capita income levels (Laugesen and Glied, 2011). Accounting for national representativeness should cancel out any effects that might be driven by those studies that estimate costs for sub-national, regional- or city-level population samples. Including an estimator for diabetes complications should account for the possible underestimation of diabetes costs in studies excluding complications. We exclude country estimates extracted from multi-country studies in our preferred specification, as their inclusion would lead to an over-statement of the cost effect of the estimation method employed in the given multi-country study.

Figure 2.1: PRISMA flowchart.



2.3 Results

Due to the differences in methodologies, we first present the findings on the identified COI studies and subsequently turn to studies on labour market outcomes.

2.3.1 COI studies on Type 2 Diabetes

Number of Studies

We identified a total of 86 relevant COI studies (see Table 2.8 for a detailed description of the included studies), of which 62 focused on HICs, 23 on LMICs, and one multi-country study covered both HICs and LMICs. Studies in LMICs increased over time, with the majority of the LMIC studies being published between 2007 and 2014. Six of the selected studies were multi-country studies, of which two (Kirigia et al., 2009; Smith-Spangler et al., 2012) did not provide detailed cost estimates for every country in the study and one did not provide a year for the estimated costs, so that we could not calculate estimates in international dollars (Boutayeb and Boutayeb, 2014). Therefore, we could not include these particular studies in our country-specific analysis.

Regional Distribution

In terms of geographic regions, most studies were carried out on countries in Latin America and the Caribbean (n=38) and Europe (n=37), followed by the USA and Canada (n=26), East Asia and Pacific (n=11), the Middle East and North Africa (n=5), South Asia (n=4), Sub-Saharan Africa (n=4) and Australia (n=1).[1] The USA was the most studied country (n=19), followed by Canada (n=7) and Germany (n=5). Mexico (n=6) and China (n=4) were the most frequently studied LMICs.

Data Sources

Especially in LMICs, self-administered surveys represented a popular method to retrieve data on the cost of diabetes. These were mostly limited regionally, i.e. to a city or hospital, and usually only representative of these regional diabetes populations but not of a national population. In HICs, databases of insurance and healthcare providers were the main source of information in most studies. These data tended to be representative either at a national or at some sub-national level. As a result, the size of the samples in HICs was

[1] The number of countries studied is higher than the number of articles reviewed due to four multi-country studies (Abdulkadri et al., 2009; Barceló et al., 2003; Boutayeb and Boutayeb, 2014; Jönsson, 2002), estimating costs for multiple countries.

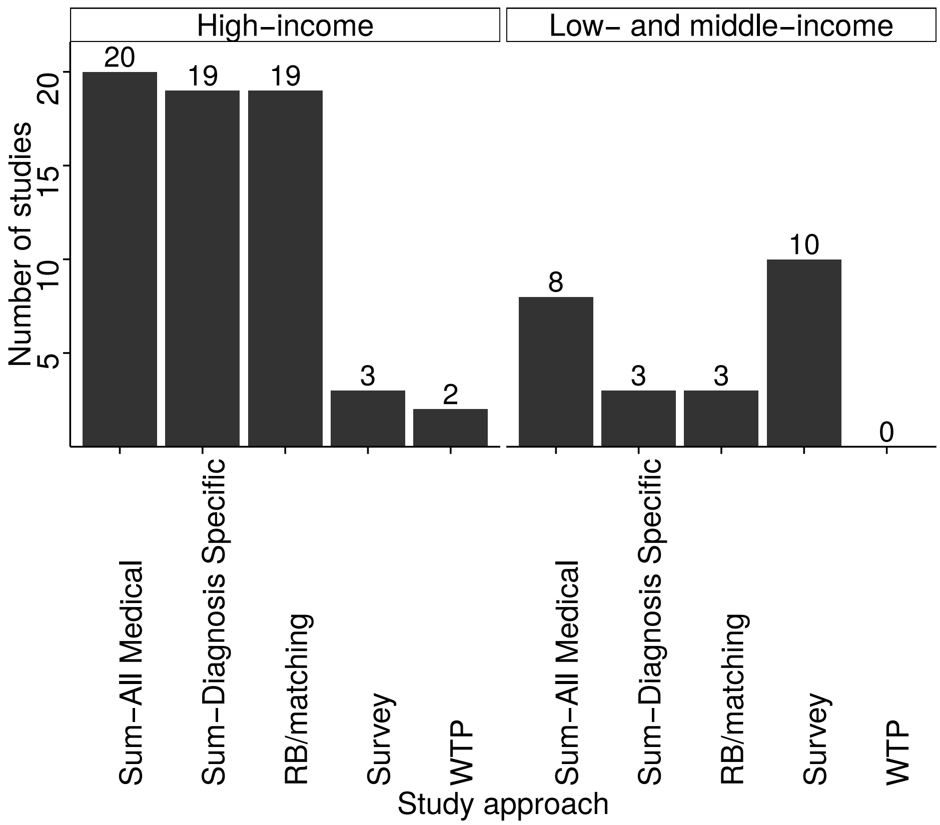
mostly between 1,000 and several million. By contrast, studies in low- and lower-middle-income countries were generally characterized by smaller sample sizes, ranging from 35 (Suleiman et al., 2006) to about 2,433 (Yang et al., 2012) in the studies reviewed here.

Variation in Costing Approaches

As discussed in more detail in Text Box 1, a range of costing approaches can be found in the COI literature. Figure 2.2 shows that the most common costing method for the direct costs of diabetes in HICs was the sum-all medical approach for people with diabetes without using control groups (Arredondo and Barcelo, 2007; Arredondo, Zúñiga, and Parada, 2005; Arredondo and De Icaza, 2011a; Arredondo and Zúñiga, 2004; Barceló et al., 2003; Bjegovic et al., 2007; Boutayeb and Boutayeb, 2014; Brandle et al., 2003; Camilo González et al., 2009; Chi et al., 2011; Condliffe et al., 2013; Horak, 2009; Jönsson, 2002; Kirigia et al., 2009; Lau et al., 2011; Lee et al., 2006; Lucioni et al., 2003; Maciejewski and Maynard, 2004; Martin et al., 2007; Morsanutto et al., 2006; Nakamura et al., 2008; Nolan et al., 2006; Ohinmaa et al., 2004; Oliva et al., 2004; Peele et al., 2002; Pohar, Majumdar, et al., 2007; Redekop et al., 2002; Ringborg et al., 2008; Zhou et al., 2005). The disease-attributable costing approach (Abdulkadri et al., 2009; Ballesta et al., 2006; Bastida and Pagán, 2002a; Buescher et al., 2010; Dall, Nikolov, et al., 2003; Davis et al., 2006; Honkasalo et al., 2014; Johnson et al., 2006; Lin et al., 2004; Mata et al., 2002; Rodríguez Bolaños et al., 2010; Simpson et al., 2003; Solli et al., 2010; Suleiman et al., 2006; Tunceli, Wade, et al., 2010) and the attributable-fraction approach were also used widely, though mainly in the USA (Bolin et al., 2009; Dall, Mann, et al., 2008; Dall, Zhang, et al., 2010; Dawson et al., 2002; Honeycutt et al., 2009; Lesniowska et al., 2014; Schmitt-Koopmann et al., 2004). The incremental cost approach was applied primarily in studies on HICs (Birnbaum et al., 2003; Bruno et al., 2012; Chodick et al., 2005; Durden et al., 2009; Esteghamati et al., 2009; Honeycutt et al., 2009; Köster, Ferber, et al., 2006; Köster, Huppertz, et al., 2011; Köster, Schubert, et al., 2012; Linden et al., 2009; Marchesini et al., 2011; Norlund et al., 2001; O'Connell et al., 2012; Pohar and Johnson, 2007; Ramsey et al., 2002; Ricordeau et al., 2003; Rodbard et al., 2010; Smith-Spangler et al., 2012; Trogdon and Hylands, 2008; Tunceli, Wade, et al., 2010; Wiréhn et al., 2008; Yang et al., 2012). For LMICs, the survey approach was the most used (Biorac et al., 2009; Chan et al., 2007; Chatterjee et al., 2011; Druss et al., 2001; Elrayah-Eliadarous et al., 2010; Javanbakht et al., 2011; Khowaja et al., 2007; Al-Maskari et al., 2010; Ramachandran et al., 2007; Tharkar et al., 2010; Wang, Fu, Pan, et al., 2009; Wang, Fu, Zhuo, et al., 2010; Wang, McGreevey, et al., 2009).

By contrast, almost all indirect cost assessments followed the same methodology, i.e.

Figure 2.2: Number of COI studies, by costing approach and income group.



Notes: For LMICs no willingness to pay (WTP) study is counted, because the only study (Tharkar et al., 2010) presenting a WTP estimate for a LMIC used primarily a different approach to estimate costs, and the WTP estimate was only presented additionally. Therefore this study was not counted under WTP here. Two studies are counted twice as they give estimates for a sum-diagnosis specific and a RB/matching approach.

the human capital approach. This approach considers all forgone labour earnings of a patient or caregiver that are attributable to diabetes. A minority of three studies (Chang, 2010; Gyldmark and Morrison, 2001; Tharkar et al., 2010), estimated the indirect costs using the WTP approach, which tries to measure how much individuals would be willing to pay to reduce the risk of an illness (Segel, 2006), here diabetes (or certain complications associated with it). One of the studies included WTP estimates in addition to the direct and indirect costs measured by the human capital approach (Tharkar et al., 2010) but did not include the WTP estimate in the overall cost estimate, while the other two studies estimated exclusively the WTP (Chang, 2010; Gyldmark and Morrison, 2001).

Study Perspective

Studies also varied in their perspective, again compromising direct comparability of the cost estimates across studies. Overall, most studies either took a societal (n=32) or healthcare system perspective (n=48). The former generally takes into account the direct and indirect monetary costs that arise to society, including costs to the healthcare system, costs due to lost productivity and sometimes OOP costs (Segel, 2006). The latter was especially common in HICs where many studies assessed the cost of diabetes to private or public health insurances. In LMICs, studies often took the patient perspective (n=5), estimating OOP expenditures and in some cases productivity losses, directly arising to the diabetes patient.

Text box 1 COI methodologies

Methodologies for COI studies can broadly be categorized into two main categories:(1) estimating the total disease costs and (2) estimating the incremental costs (Akobundu et al., 2006). Studies can then be divided further according to the specific approach used for estimation. Our categorization builds on that by Akobundu et al. (2006) in their review of COI methodologies.

1. Total disease costs

    a) Sum-All Medical: captures all medical expenditures of a person diagnosed with diabetes, irrespective of the relation of the expenditures with diabetes.

    b) Sum-Diagnosis Specific: includes the costs that are related to diabetes. This can be done by using a disease-attributable costing approach, using administrative claims databases to identify the cost of diabetes by respective International Statistical Classification of Diseases and Related Health Problems (ICD) codes that link the expenditures to a primary or secondary diagnosis of diabetes as the reason for the healthcare utilization. Alternatively, a similar technique used at the population level is the attributable-fraction approach, where the relative contribution of, e.g., diabetes, to the risk of developing another disease (e.g. renopathy or cardiovascular disease) is used to determine how much of the costs of this disease can be attributed to diabetes.

    c) Survey approach: while not specifically mentioned by Akobundu et al. (2006), for this review we create a separate category capturing studies using surveys of people with diabetes. This category differs from the two approaches a) and b) above in that estimations rely solely on the individual, reported experience of people with diabetes, without use of any diagnostic data at an aggregate level. The survey approach was also used as a separate category in the earlier review on diabetes COI studies by Ettaro et al. (2004).

2. Incremental disease costs

    There are two main approaches for the estimation of incremental medical costs:

    a) Regression approach: a statistical technique which can account for observable differences between the group with diabetes and the control group (i.e. those without diabetes) to find—ideally—the independent effect of diabetes on healthcare costs. The differences typically accounted for are age, region and gender.

    b) Matching approach: uses a control group to directly compare those with diabetes to those without diabetes after matching each person of the 'treatment' group to a 'similar' person of the control group, using various categories like age, region and gender to—again—find the independent effect of diabetes on healthcare cost (Akobundu et al., 2006).

All of the above approaches can be used in prevalence or an incidence based study. In the former case the costs of diabetes are estimated for a certain point in time, typically one year, while the latter approach estimates costs over a person's lifetime or several years, always starting with the point at which the disease is diagnosed. Both approaches may also be combined in studies estimating the future cost burden of type 2 diabetes by first taking a prevalence approach to

Costing Components

Of the 75 studies that reported the cost components they used to estimate direct costs, 72 took into account outpatient hospital visits, 70 inpatient hospital visits, 63 physician visits, 58 drug costs, 51 laboratory costs for diagnostic tests and check-ups, 37 equipment costs and 21 non-medical and transportation costs. A total of 46 studies had at least included the costs of hospital, outpatient and physician visits as well as drugs (see Table 2.9 for a detailed description of cost components used in each study).

Cost Estimates of Diabetes Using a Prevalence Approach

Two basic epidemiological approaches exist for the estimation of COI, and they are not directly comparable. The incidence approach follows people with diabetes, usually starting with their diagnosis at a common base year, estimating yearly costs for a sample of people at the same disease stage, finally giving an estimate of diabetes costs over a certain time period, such as from diagnosis to death or over a distinct period of, for example, 10 years. This approach can also document how costs of diabetes change and develop over the progression of the disease (Larg and Moss, 2011). By contrast, the prevalence approach estimates the costs of diabetes for a cross-section of people with diabetes at a certain point in time, normally a year, who are at different stages of the disease. It is most suitable for assessing the total economic burden of diabetes at a certain point in time. Due to this difference in time periods and the used data, the estimates of prevalence-based studies are not directly comparable with those of incidence-based studies. Hence, we present the cost estimates separately, starting with the prevalence approach.

Table 2.2 shows the range of direct cost estimates by estimation approach and income status. As can be observed, direct cost estimates varied widely, both between and within the different estimation approaches. Cost estimates for direct costs, irrespective of the costing method applied and the cost components included, ranged from $242 for Mexico Arredondo, Zúñiga, and Parada (2005) in 2010 to $11,917 for the USA Condliffe et al. (2013) in 2007. Also, studies from LMICs generally indicated smaller direct costs than studies from HICs.

For indirect costs, studies using the human capital approach estimated costs ranging from $45 for Pakistan (Khowaja et al., 2007) in 2006 to $16,914 for the Bahamas (Barceló et al., 2003) in 2000. Three studies estimated indirect costs by using the WTP approach and found costs ranging from $191 in a study on the WTP for a health insurance for type 2 diabetes in Denmark in 1993 (Gyldmark and Morrison, 2001), a WTP $4,004 per year for a cure of type 2 diabetes (Chang, 2010) in Taiwan and an annual payment of $4,737 to halt disease progression/prevent future complications of diabetes in India (Tharkar et al., 2010).

Societal costs of Type 2 Diabetes, which are estimated by studies combining direct and indirect costs, ranged from $544 in a study on the economic costs of diabetes in Iran (Esteghamati et al., 2009) in 2001 to $18,224 for the Bahamas (Barceló et al., 2003) in 2000.

Table 2.1: Summary of direct costs by estimation approach and income status in international dollars $ (2011) for prevalence-based studies.

	High-income countries				Low- and middle-income countries			
	Sum-all medical costs	Sum-diagnosis specific	RB / matching	own survey	Sum-all medical costs	Sum-diagnosis specific	RB / matching	own survey
Min	1117	907	264	1495	242	662	443	456
Max	11917	9346	8306	5585	4129	4672	1136	3401
N	25[a]	19[a]	18	3	27[a]	5[a]	2	10

[a] Includes country estimates from multi-country studies; RB Regression based

In order to improve the cross-country comparability of the costs of diabetes we plotted the results from studies providing a direct per capita cost estimate against the GDP per capita estimate of the respective country (we limited this comparison to studies using samples representative of their entire population). Figure 2.3 confirms the expectation that costs do increase with economic wealth: GDP per capita explains about one-third of the variation in cost estimates (see r2 in Figure 2.3). Also, studies on the USA seem to estimate costs consistently higher than would be expected on the basis of its GDP per capita.

The USA, however, spend consistently more than what would be expected on the basis of its GDP per capita. Again, the wide variation in estimated costs for many countries underscores the point that the studies need to be contextualized and may not be directly comparable per se. On the whole—though by no means always—the matching and regression as well as the sum-diagnosis specific approaches appear to produce lower cost estimates than especially the total cost results, particularly so for HICs. In an inevitably crude attempt to quantitatively explore the driving factors behind the heterogeneity in cost estimates, we estimated a simple linear regression model with per capita direct costs as the dependent variable; explanatory variables included GDP per capita, the estimation approach employed by the study, the number of included cost components, a dummy for studies carried out in the USA, the year of data collection, the representativeness of the study and if the study included diabetes complications as explanatory variables. The results, displayed in Table 2, show a strong relationship between GDP per capita and expenditures for diabetes, with every additional international dollar in per capita GDP translating into an average increase in direct diabetes expenditures of about $0.04. The estimation approach is not found to matter significantly, nor is the year of study. Estimates from USA studies put the costs at over $3,000 higher (on average) than studies from other countries, indicating that costs in the USA may indeed be unusually high. The number of costing components and the inclusion of complications likely also explain some of the variance in estimates, although they are just below and above the 10 % significance level, respectively. Overall, the included independent variables explain about 56 % of the variation in direct cost estimates.[2]

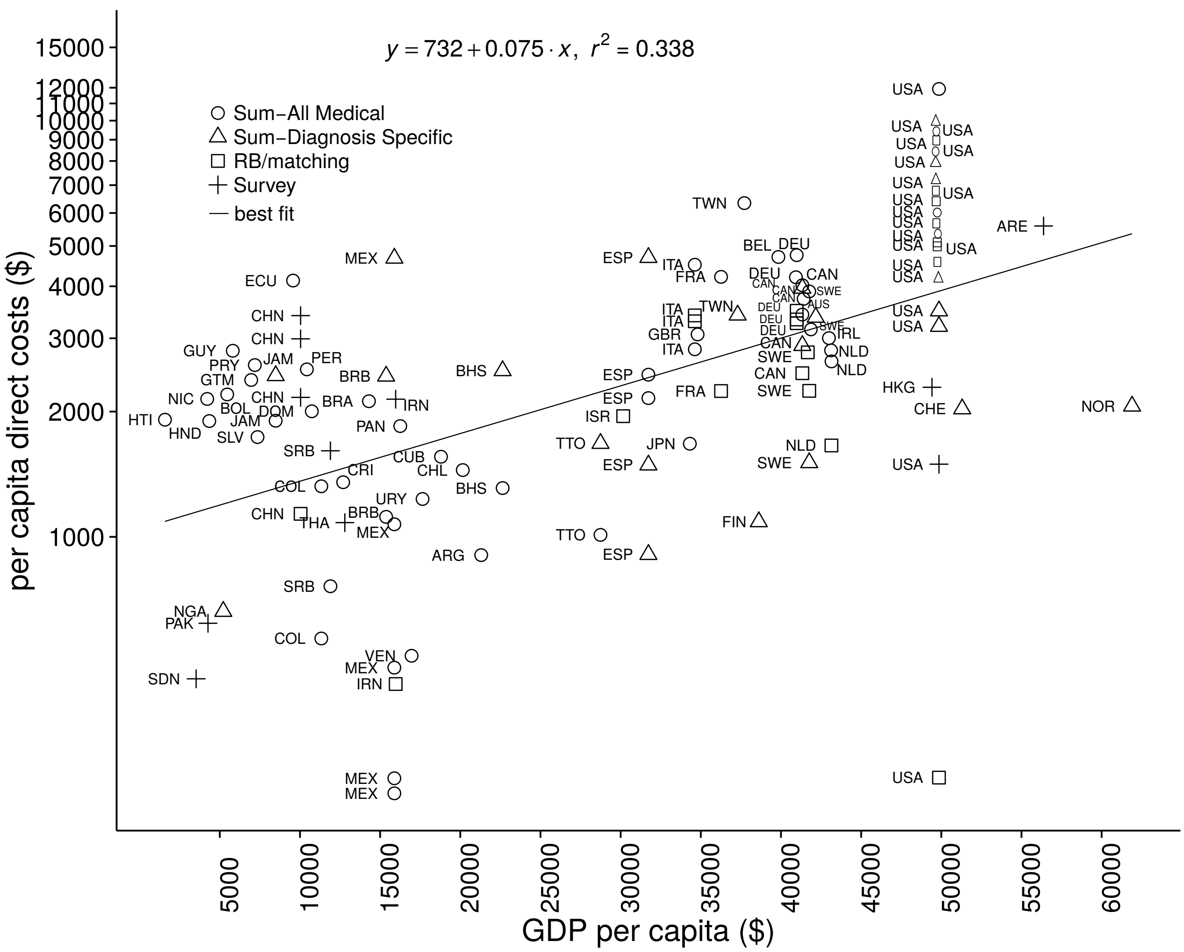
The sensitivity of the cost results to the estimation approach was also examined by two studies that investigated the effect of different estimation techniques in diabetes COI studies. Honeycutt et al. (2009) compared the use of a regression-based and an attributable-fraction approach and found that the cost estimate of the former exceeded the latter by 43 %. Tunceli, Wade, et al. (2010) compared the matching and the diabetes (disease)-attributable costs approach and found a 14–29 % higher cost estimate using matching, depending on the used assumptions. Both studies concluded that an incremental cost approach results in a higher, and likely more exact, estimate of the direct costs of diabetes than disease-attributable approaches. The authors attributed this to the fact that a regression or matching approach can assign costs to diabetes that cannot be linked to diabetes otherwise. Those approaches are therefore in a position to account for all costs of co-morbidities caused by diabetes, while this is not automatically the case with the other approaches.

Direct and Indirect Costs of Diabetes

To assess the relative importance of direct and indirect costs across countries, we plotted direct against indirect costs from studies that provided both estimates and drew a 45°line depicting the equal share of direct and indirect costs (see Figure 2.4).

[2] In a sensitivity analysis, we included the results from multi-country studies providing country estimates in the regression analysis. The only major difference to the presented analysis is that the inclusion of complications as well as the number of included cost components were now significant at the 1 and 5 % significance level, respectively. The effect size and significance of the other estimates did not change considerably.

Figure 2.3: GDP to direct costs ratio by estimation approach.



Notes: The line depicts the best fit based on the linear regression of direct costs on GDP per capita in international dollars.

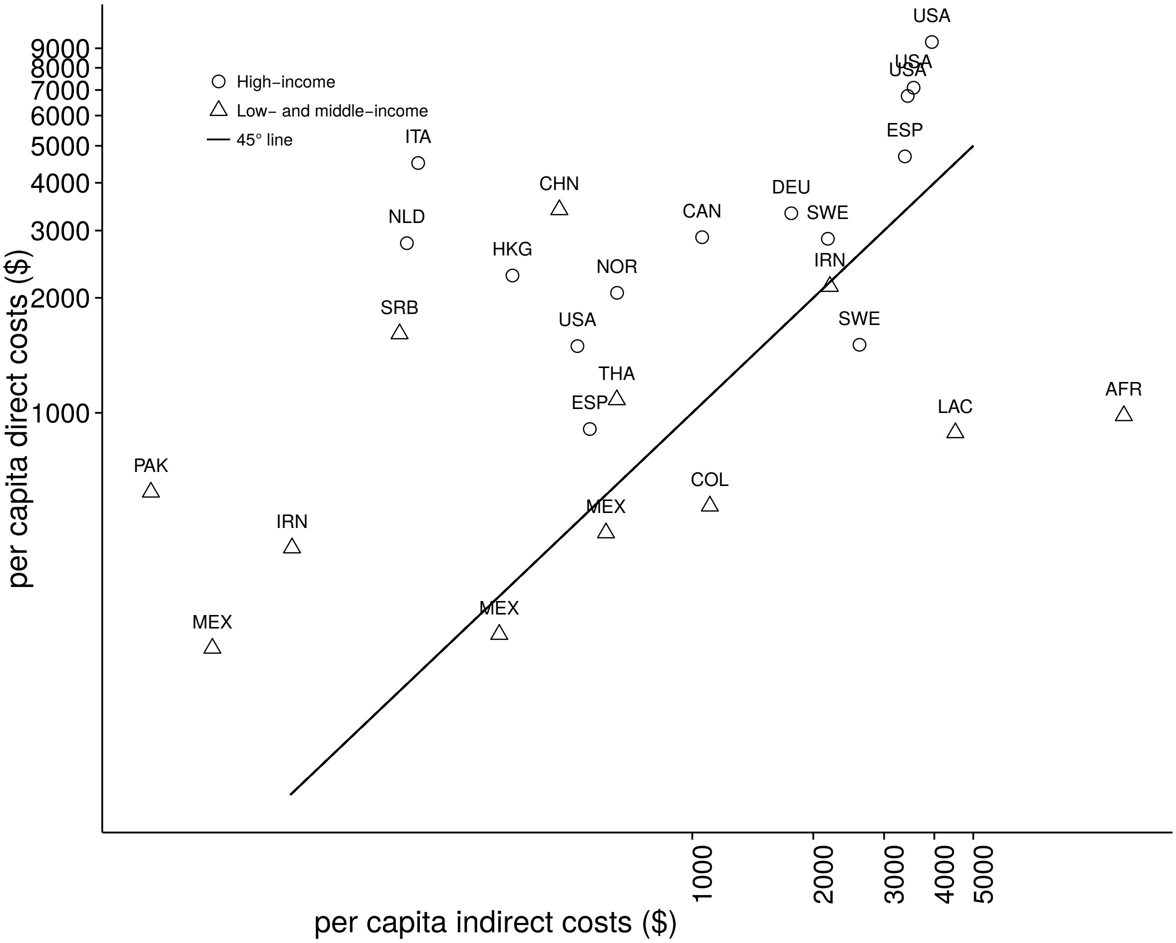
Table 2.2: Relationship between direct costs and study characteristics (robust linear regression).

	Estimate	Std. Error
Constant	2133	1773.922
GDP per capita ($)	0.045**	0.017
Estimation Approach		
Sum-All medical (Ref.)		
Sum-Diagnosis Specific	−413.880	528.766
RB/matching	−719.868	526.896
Survey	−689.806	671.020
At least four costing components	702.966*	403.968
USA study	3111.067***	533.534
Year of study		
<1995 (Ref.)		
1995-1999	−1744.799	1632.498
2000-2004	−816.647	1586.966
2005-2009	−1021.685	1592.595
2010-2014	−2744.739	1839.689
Study representative	−598.670	409.070
Complications	666.803	414.727
R-squared adj.	0.559	
N	70	

Standard errors in parenthesis. Ref. reference category.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Most studies found a larger share for direct costs in comparison with indirect costs (observations above the 45°line in Figure 2.4). This is especially true for HICs, where only a study on Sweden (Bolin et al., 2009) found a larger share for indirect costs. For LMICs, a study on Colombia (Camilo González et al., 2009) found considerably higher indirect costs, as did the multi-country study of Barceló et al. (2003) and a study on various countries in the African region (Kirigia et al., 2009), which both found higher indirect costs for almost every country in the study and also on average for the entire regions, represented as the mean overall study estimate in Figure 2.4. Both studies used similar approaches to estimate costs, and indirect cost estimates were likely so high because evidence from only a few countries within the region were used as a basis for estimating indirect costs for every other country in the respective study. Further, the studies took the countries' per capita gross national product as a proxy for earnings, which might have led to an over-estimation of the indirect costs (Kirigia et al., 2009).

Figure 2.4: Direct and indirect cost relation in studies estimating total costs of type 2 diabetes.



Notes: The 45°line depicts the points where direct and indirect costs would be equal. Above the line direct costs are higher than indirect costs and vice versa. For better visibility both coordinate axes are expressed in log scale

Studies Using the Incidence Approach

The four studies that used an incidence approach (see Table 2.3 estimated the cost of diabetes either over a person's lifetime (Birnbaum et al., 2003; Camilo González et al., 2009) or over a certain period after diagnosis Johnson et al. (2006) and Martin et al. (2007). Camilo González et al. (2009) modelled the lifetime (direct and indirect) costs of a typical diabetes patient in Colombia, arriving at a mean cost estimate of $54,000. The second study providing lifetime estimates by Birnbaum et al. (2003), estimated incremental lifetime healthcare costs for USA females with diabetes of $283,000.

Two studies followed patients over a limited time period and found different patterns in the development of Type 2 Diabetes-attributable healthcare costs. In Germany costs increased from $1634 in the first year after diagnosis to $4881 in the seventh year (Martin et al., 2007). In Canada, Johnson et al. (2006) found the highest costs in the year of diagnosis with $7635, up from $2755 the year prior to diagnosis. In the year after diagnosis costs decreased to $4273 and then only increased slightly to $4618 in year ten. In Germany and Canada, costs related to complications or hospital visits were the most important components and in Germany increased steadily over time. In Canada costs related to prescriptions increased the most.

Country level costs prediction studies

Four studies projected costs of diabetes over a certain period of time (Davis et al., 2006; Lau et al., 2011; Ohinmaa et al., 2004; Wang, McGreevey, et al., 2009), making assumptions about the future development of diabetes prevalence and population ageing (see Table 2.4). For Canada, a 1.7-fold increase from 2000 to 2016 (Ohinmaa et al., 2004) and a 2.4-fold increase from 2008 to 2035 in diabetes healthcare costs was estimated (Lau et al., 2011). Taking a health care system perspective, both studies found that the estimated increase would be mostly driven by an ageing population. For Australia, Davis et al. (2006) estimated a 2.5- to 3.4-fold increase in diabetes attributable healthcare costs from 2000 to 2051, depending on the underlying assumptions about population ageing and diabetes prevalence rates. For China, Wang, McGreevey, et al. (2009) extrapolated total costs of diabetes from the year 2007 to 2030, estimating the costs of diabetes to increase 1.8-fold, solely accounting for the expected increase in prevalence.

2.3.2 The Impact of Diabetes on Employment Chances and Productivity

Besides studies that determined the cost of diabetes by costing related expenditures, another body of research has investigated —using econometric techniques—the impact of diabetes on 'productivity', a term used here to comprise outcomes including employment probabilities and lost work days and income or earnings. A recent study systematically reviewed evidence on the impact of diabetes on the ability to work, focusing on studies assessing the impact of diabetes on early retirement, lost work hours, absenteeism and presenteeism (Breton et al., 2013). We focused particularly on studies exploring the impact of diabetes on employment probabilities and earnings—both issues that were not covered

Table 2.3: Incidence studies on the costs of diabetes

Ref.	Country	Time horizon	Population	Approach	Results
Johnson et al. (2006)	Canada	1992–2001	Incidence T2D patients from Saskatchewan Health's administrative database in Canada	Sum-all medical	Highest total healthcare costs at year of diagnosis with CAN$7343 ($7635), then increased from a low of CAN$3880 ($4034) 3 years after diagnosis to CAN$4441 10 years thereafter ($4618).
Camilo González et al. (2009)	Colombia	32 years	Hypothetical average Columbian T2D patient	Sum-all medical	Total lifetime costs (32 year period) of average diabetes patient, including direct and indirect costs, 57.565 million Colombian pesos ($54,351).
Martin et al. (2007)	Germany	1995–2003	Newly diagnosed T2D patients from randomly drawn practices across Germany	Sum-all medical	EUR 1,288 ($1635) for the first treatment year after diabetes diagnosis and increased to EUR 3845 ($4880) in the seventh year.
Birnbaum et al. (2003)	United States	1997–1998	Women employed by nationwide operating company and hypothetical women above age 64 receiving Medicare	RB/matching	$282973 incremental lifetime direct healthcare costs, using incidence-based, steady-state methodology.

Table 2.4: Country level costs prediction studies

Ref.	Country	Population	Approach	Time horizon	Results
Davis et al. (2006)	Australia	Australian population	Sum diagnosis Specific	2000–2051	If age and sex specific prevalence remains unchanged a 2.5-fold increase; if age and sex specific prevalence allowed to change as well a 3.4-fold increase.
Ohinmaa et al. (2004)	Canada	Canadian population	Sum-all medical costs	2000–2016	1.7-fold increase.
Lau et al. (2011)	Canada	Four Alberta Health and Wellness databases	Sum-all medical costs	2008–2035	2.4-fold increase.
Wang, McGreevey, et al. (2009)	China	In patients and outpatients in 20 hospitals	Own survey	2007 and 2030 (projection)	Increase from $73 billion in 2007 to $132 billion in 2030 (1.8 fold increase).

in the mentioned review—and we took a more detailed look at the empirical challenges posed by the issue of endogeneity (see the Appendix for a more detailed discussion of endogeneity).

Tables 2.5 and 2.6 synthesize the relevant information from the 22 identified studies on the effect of diabetes on employment and other labour market outcomes. Almost all studies were conducted on HICs, mainly the USA (n=13) and European countries (n=4). Only one study focused on a LMIC. investigating the effect of diabetes on labour income in China.

Employment Chances

Most studies examined the impact of diabetes on employment probability (n=17), applying a range of econometric techniques. These have evolved over time, and more recent studies took into account the possibility that diabetes might be endogenous: it is conceivable that especially personal traits such as motivation and drive could influence the propensity to develop type 2 diabetes as well as a persons' job market opportunities. Further, being employed or unemployed could also lead to changes in lifestyles, due to changes in income, stress or leisure time, that could themselves affect the chances of developing diabetes (Brown, Pagan, et al., 2005). Of the studies that tried to account for this problem (Brown, Pagan, et al., 2005; Harris, 2009; Latif, 2009a; Lin, 2011; Minor, 2011a; Zhang et al., 2009a), the majority used an instrumental variable (IV) technique. This approach allows for the consistent estimation of the effect of diabetes on employment if a variable can be found that is causally related to diabetes without affecting the employment chances through any other unobserved pathway apart from its effect on diabetes. (see Text Box in Online Resource 2). In the case of type 2 diabetes all studies used the family history of diabetes as an IV to exploit the fact that the development of type 2 diabetes is much more likely for individuals whose biological parents have also had diabetes. It is argued that, while controlling for education, age and other observable demographic and socioeconomic factors (e.g. wealth, regional and ethnic differences and the number of children in the household), having a family member with diabetes should not affect the person's employment status or other labour market outcomes, while strongly predicting the onset of type 2 diabetes.

Table 2.5: Studies estimating the relationship between diabetes and employment (2001 – 2014)

Ref	Survey year	Country	Age	Effect on employment	
				Males	Females
Harris (2009)	1999-2000	Australia	>24	Exogenous: 10.8 percentage points reduction to be in labour force; endogenous: 7.1 percentage points reduction and test indicates endogeneneity.	Exogenous: 10 percentage points to be in labour force; endogenous: Nine percentage points reduction and test indicates endogeneneity.
Zhang et al. (2009a)	2001, 2004-2005	Australia	18-64	50-64: 11.5 percentage points less likely to be in labour force; 18-49: 3.9 percentage points less likely, all effects increase when other chronic diseases are present.	No significant effect for diabetes alone; significant negative effect if other chronic diseases are present.
Latif (2009a)	1998	Canada	15-64	Exogenous: 19 percentage points less likely to be employed; endogenous: not significant and positive and test indicates endogeneity.	Exogenous: 17 percentage points less likely to be employed, endogenous: not significant and positive and test indicates exogeneity.

Ref	Survey year	Country	Age	Effect on employment	
				Males	Females
Kraut et al. (2001)	1983-1990	Canada	18-64	With complications 2 times less likely to be in labour force; no significant effect on employment for those in labour force.[a]	
Norlund et al. (2001)	1992-1993	Sweden	>24	14.2 percentage points higher retirement rate (22.9 compared to 8.7).[a]	
Alavinia and Burdorf (2008)	2004	Sweden, Denmark, Netherlands, Germany, Austria, Switzerland, France, Italy, Spain, Greece	50-65	For whole dataset: no effect of diabetes on being unemployed, but increased odds ratio of 1.33 on being retired. No information on effects by country.[a]	
Lin (2011)	2005	Taiwan	45-64	Exogenous: 9 percentage points less likely to be employed; endogenous: 19 percentage points less likely to be employed; test on whole sample indicates endogeneity.	Exogenous: 11 percentage points less likely to be employed, endogenous: not significant and negative.

Ref	Survey year	Country	Age	Effect on employment	
				Males	Females
Brown, Pagan, et al. (2005)		United States	>44	Exogenous: 7.4 percentage points less likely to be employed; endogenous: 10.6 percentage points less likely but test indicates exogeneity.	Exogenous: 7.5 percentage points less likely to be employed; Endogenous: no significant effect found and test indicates endogeneity.
Minor (2011a)	2006	United States	>19 at diagnosis		Exogenous: 25.2 percentage points less likely to be employed, endogenous: 45.1 percentage points less likely to be employed.
Vijan et al. (2004a)	1992-2000	United States	51-61	More likely to be retired in 1992 (adjusted OR 1.3). Over 8 years follow up spent 0.14 incremental years in retirement.[a]	
Bastida and Pagán (2002a)	1996-1997	United States	>44	7.5 percentage points less likely to be employed.	No significant effect on employment chances found.

Ref	Survey year	Country	Age	Effect on employment	
				Males	Females
Brown, Perez, et al. (2011)	2008	United States	35-64	Diabetes negatively related to employment (5 percentage points reduction); better diabetes management (HbA1c) positively affects employment probabilities; HbA1c lowering of 10% increases employment probability by 0.44 percentage points.	No significant effect on employment chances found.
Tunceli, Bradley, et al. (2005)	1992,1994	United States	51-61	9 percentage points less likely to work without complications controlled for, with complications controlled for 7.1 percentage points less likely.	5.9 percentage points less likely to work without complications controlled for, with complications controlled for 4.4 percentage points less likely but not significant.
Tunceli, Zeng, et al. (2009)	1997-2005	United States	20-44 and 45-64	20-44: proportion with work limitations 3.1% higher; 45-64: proportion not working is 8.1% higher; the proportion work disabled is 3.4% higher; proportion with work limitations is 5.7% higher (all compared to similar age group without diabetes).[a]	

Ref	Survey year	Country	Age	Effect on employment	
				Males	Females
Valdmanis et al. (2001)	1990-1995	United States		Unemployment rate for persons with diabetes was 16% compared with 3% among matched comparison group.[a]	
Ng et al. (2001)	1989	United States	>29 at diagnosis	3.6% less likely of being employed (exogenous), 12% for those with complications.[a]	
Minor (2013)	1979-2010	United States	>14	Average reduction of employment probability of 28 percentage points; strongest employment penalty in first 5 years after diagnosis.	Average reduction of employment probability of 36 percentage points; strongest employment penalty in first 15 years after diagnosis.

[a] No gender differentiation in study

Because IV estimation has worse asymptotic properties than single equation regression results when endogeneity is not an issue, studies tested for the existence of endogeneity to determine which results to rely on for inference (Brown, Pagan, et al., 2005; Latif, 2009a; Lin, 2011; Minor, 2011a). Interestingly, the reviewed studies found diabetes to be endogenous for either males (Latif, 2009a) or females (Brown, Pagan, et al., 2005; Minor, 2011a), but never for both. Further, the use of an IV sometimes increased the estimated effect(Lin, 2011; Minor, 2011a) whereas in other cases the effect turned insignificant (Brown, Pagan, et al., 2005; Latif, 2009a). As a result, no unambiguous conclusions can be drawn as to how endogeneity affects diabetes and whether or not it causes biased estimates. Most of the relevant studies also explored whether accounting for body mass index (BMI) or other diabetes-related chronic conditions would substantially alter the result and found this not to be the case (Brown, Pagan, et al., 2005; Latif, 2009a; Minor, 2013).

Overall, studies more commonly found a significant adverse impact of diabetes on males, ranging from no effect in Canada (Latif, 2009a) to a 19 percentage point reduction in Taiwan (Lin, 2011). Conversely, no effect was found for women in Taiwan (Lin, 2011), Australia (Zhang et al., 2009a) or for Mexican Americans in Texas (Brown, Pagan, et al., 2005). However, a 45 % decrease in employment chances was observed for women in the USA (Minor, 2011a). Extending the scope and looking at how diabetes duration affected labour market outcomes, using panel data from the USA, one study found that the main adverse effect on employment chances materialized within the first 5 years after diagnosis for men and 11–15 years after diagnosis for women (Minor, 2013).

Productivity

For earnings, no effect was found for Mexican-American men in Texas (Bastida and Pagán, 2002a), while the highest loss was found for women in the USA ($21392 per year) (Minor, 2011a). Again looking at diabetes duration, a wage penalty was only found for USA men 6–10 years after diagnosis, reducing their wage by about 18 percentage points (Minor, 2013). The only study on a non-HIC, China, tried to tease out the psychological effect of a diabetes diagnosis on subsequent labour income, finding a reduction of 22 % in income for males, but not for females. Further, those with an HbA1c between 8–10 % experienced the most severe income penalty (29 %). The study further showed that the adverse effect of a diabetes diagnosis was concentrated among the poorest third of the study population (Liu and Zhu, 2014). Another study investigated the effect on earning losses for caregivers of people with diabetes in the United Kingdom (UK), finding a reduction of $2,609 per year, while the person with diabetes experienced a loss of $1,744 per year (Holmes et al., 2003). For income, a reduction of $6,250 per year was found for older USA adults who had been followed between the years 1992 and 2000 (Vijan et al., 2004a). In terms of lost workdays and work hours due to diabetes, the effects ranged from no impact on lost work days on older people (Vijan et al., 2004a) and females in the USA (Minor, 2011a) to 3.2 lost work days in a USA population within a 2-week period if complications were present (Ng et al., 2001).

Table 2.6: Studies estimating the relationship between diabetes and other productivity outcomes (2001 – 2014)

Ref.	Survey year	Country	Age	Effect on other productivity outcomes	
				Males	Females
Kraut et al. (2001)	1983–1990	Canada	18–64	Effect on earnings only when complications are present: reduced to 72% of total income of controls.a	
Liu and Zhu (2014)	2009, 2011	China	not given	16.3% decrease in annual income; strongest effect for those in lower income quintiles.	16.3% decrease in annual income; strongest effect for those in lower income quintiles.
Herquelot et al. (2011)	1989–2007	France	Male 40–50, females 35–50 in 1989	1.7 HR to transition from employed to disabled, 1.6 HR to be retired, 7.3 HR to be dead; between age 35 and 60 each person with diabetes lost 1.1 years of time in workforce.[a]	
Leijten et al. (2014)	2010–2013	Netherlands	45–64	Diabetes reduced work ability measured using Work Ability Index (WAI) by 2%. No significant effect on productivity was found.[a]	
Norlund et al. (2001)	1992–1993	Sweden	>24	9.4 more sick days.[a]	
Holmes et al. (2003)	1999	United Kingdom	<65	GBP 869 lost earnings per year with diabetes; GBP 1300 for carers of people with diabetes.[a]	

Ref.	Survey year	Country	Age	Effect on other productivity outcomes	
				Males	Females
Minor (2011a)	2006	United States	>19 at diagnosis		Exogenous: $2865 loss in earnings per year, Endogenous: $19655; Exogenous: 2 working hours less per week, no significant effect on missed workdays per year, endogenous: no significant effect on working hours or workdays missed.
Vijan et al. (2004a)	1992–2000	United States	51–61	Lost income of $50004 from 1992–2000 per capita or $6250 per year, for whole USA population of same age $85.6 billion or $10.7 billion per year; people with diabetes more likely to have taken sick days in 1992 (adjusted OR 1.3).[a]	
Collins et al. (2005)	2002	United States	working age	No significant effect on work days.[a]	
Bastida and Pagán (2002a)	1996–1997	United States	>44	No significant effect on earnings.	Women with diabetes earn 84% less.

Ref.	Survey year	Country	Age	Effect on other productivity outcomes	
				Males	Females
Brown, Perez, et al. (2011)	2008	United States	35–64	Wages reduced by 0.74% due to diabetes; for every 10% reduction in A1C wages rise by 0.62 %. A1C >8 was related to decreasing wages.	No significant effect of diabetes on female earnings; no effect of blood sugar management for women, A1C levels just below 6 to just above 7 were related to lower wages.
Lenneman et al. (2011)	2005–2009	United States	>16	Lost earnings per year of $2146.[a]	
Tunceli, Bradley, et al. (2005)	1992, 1994	United States	51–61	No significant effect on number of work days.	2.5 more lost workdays per year.
Valdmanis et al. (2001)	1990–1995	United States		71% of the persons with diabetes had an annual income of less than $20000 compared with 59% of the matched respondents.[a]	
Ng et al. (2001)	1989	United States	>29 at diagnosis	No significant effect on work days for T2D, for those with complications 3.2 days lost within two weeks	

Ref.	Survey year	Country	Age	Effect on other productivity outcomes	
				Males	Females
Brown, Estrada, et al. (2005)	NA	United States	>45	For every dollar of labor income lost by adults with diabetes, a further income reduction of $0.48 occurs in the community. Total output reduction for upper bound estimate is $300 million for the local economy.[a]	
Minor (2013)	1979–2010	United States	>14	no general effect of type 2 diabetes on wages; some evidence of wage penalty of about 18% 6–10 years after diagnosis	No strong evidence found for wage penalty for females

[a] No gender differentiation in study

In terms of the methodology used, these studies tended to rarely account for endogeneity, and they mostly used standard regression or matching methods to estimate the impact of diabetes. Three studies (Bastida and Pagán, 2002a; Brown, Perez, et al., 2011; Minor, 2011a) corrected for the possibility of a sample selection bias, to account for systematic differences between the working population and the overall population. Only one study additionally applied IV methods and found diabetes to be endogenous, so that its effects on earnings were dramatically understated using naive regression results (Minor, 2011a). For working hours and days missed due to illness, the same study found no indication of endogeneity. Only one study applied an approach other than IV to account for endogeneity, using a difference-in-difference model and exploiting a recent diagnosis of diabetes, which was the result of the collection of biomarkers in the survey used, 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).

2.4 Discussion

The objectives of this systematic review were to identify new evidence on the economic impact of type 2 diabetes that emerged since 2001 and extend the scope of the review by including studies on the labour market impact of diabetes. We identified studies from a great variety of countries, with large differences in cost estimates across and within countries.

2.4.1 General Findings and Developments Since the 2004 Review of Diabetes COI Studies

An obvious development since the last review is the emergence of COI studies on LMICs. The economic burden related to diabetes found in these studies indicated a strong direct impact on those affected by diabetes. This is reflected in the substantial burden of OOP treatment costs incurred by patients (Arredondo and Barcelo, 2007; Chatterjee et al., 2011; Elrayah-Eliadarous et al., 2010; Esteghamati et al., 2009; Khowaja et al., 2007; Ramachandran et al., 2007; Smith-Spangler et al., 2012; Suleiman et al., 2006; Tharkar et al., 2010; Wang, Fu, Pan, et al., 2009; Wang, Fu, Zhuo, et al., 2010), with considerable proportions of the annual income being spent on diabetes care. This relative cost burden was generally higher for people with relatively lower household incomes (Khowaja et al., 2007; Ramachandran et al., 2007; Tharkar et al., 2010). Health insurance coverage had some protective effects against OOP expenditures, but mainly for those with higher incomes, while the poor often lacked coverage (Khowaja et al., 2007; Ramachandran et al., 2007; Tharkar et al., 2010). Once people were covered by health insurance their risk of incurring catastrophic expenditures decreased significantly (Smith-Spangler et al., 2012). An important cost factor that was predominantly investigated in studies on LMICs were non-medical costs for transportation, informal healthcare or food which were found to considerably add to the experienced diabetes cost burden (Chatterjee et al., 2011; Esteghamati et al., 2009; Tharkar et al., 2010; Wang, Fu, Pan, et al., 2009; Wang, McGreevey, et al., 2009).

In terms of the costing methodology applied in COI studies, the number of studies estimating the excess costs of diabetes increased since the Ettaro et al. (2004) review. Those studies either used regression analysis or matching to adjust for the differences between people with diabetes and those without, accounting at least for age and gender, but often also for other socioeconomic, geographic and demographic differences. Other widely used approaches to estimate direct healthcare costs from the perspective of the healthcare system or private insurance included the disease-attributable and—slightly less frequently—the attributable-fraction approach. For cost assessment in LMICs, studies often either estimated total healthcare costs or carried out self-administered surveys. While Ettaro et al. (2004) suggested an increased use of disease-attributable approaches to arrive at more exact estimates of the costs of diabetes, the evidence found in this review indicates that using an incremental cost approach via matching or regression analysis could provide more accurate results, due to its ability to capture costs otherwise not directly traceable to diabetes. Nonetheless, the use of the estimation technique always hinges on the availability of appropriate data, with regression or matching analyses requiring information on people without diabetes to be used as a control group. Therefore the estimation approach needs to be tailored to the available data.

Compared with the evidence reviewed by Ettaro et al. (2004), the field has generally advanced with respect to the analysis of costs in different ethnic and age groups. Two studies investigated differences between racial groups in the USA, showing that while ethnic minorities spend less on diabetes healthcare than Whites, this difference seems to be mainly based on differences in access to care between Whites and Blacks or Hispanics (Buescher et al., 2010; Lee et al., 2006). In terms of age, studies found an increase in healthcare costs with age as well as with, in some cases, the duration of diabetes. A recurring problem was that many studies did not distinguish diabetes types, making it difficult to exactly attribute the costs to the respective diabetes types.

To explore the reasons for the wide heterogeneity in direct cost estimates across studies, we performed a regression analysis, which indicated that an important determinant for the cost variation across countries could be the economic wealth of the country (proxied by GDP per capita), similar to what was found in a review of indirect costs of various chronic diseases (Zhao, Xie, et al., 2013), possibly due to differences in the availability and affordability of diabetes care between HICs and LMICs (Cameron, Ewen, et al., 2009; Cameron, Roubos, et al., 2011).

Further, studies on the USA seem to estimate consistently higher costs than studies on other countries, even when accounting for differences in GDP per capita. The higher direct costs of diabetes estimated for the USA are in line with the generally higher healthcare expenditures in the USA compared with countries with similar income levels, and could be the result of exceptionally high service fees (Laugesen and Glied, 2011) and prices paid in the USA healthcare system (Lorenzoni et al., 2014; Squires, 2012).

Because of the small sample size on which our analysis was based, these results must be interpreted with caution, and other factors could still be important. For instance, other evidence suggests that different costing approaches have a considerable effect on diabetes cost estimates (Honeycutt et al., 2009; Tunceli, Wade, et al., 2010). Furthermore, the perspective taken, different data sources and populations investigated and decisions on the cost components included are likely important in explaining within-country heterogeneity. In particular, the inclusion of diabetes complications and decisions about which complication(s) to include, as well as the extend to which costs for these diseases are attributable to diabetes, can significantly affect the results. Not all studies in the review provide extensive information about how they include complications and some do not include them at all.

Finally, the quality of the data used could have affected the cost estimates. Many studies in LMICs relied on self-reported data from small household surveys, limiting their generalizability and leading their results to be prone to recall bias. Further, these studies often identified people with diabetes via their use of healthcare institutions, which excluded a potentially important section of the population in LMICs unable to access formal care, possibly leading to an overestimation of the average diabetes-related costs.

2.4.2 Labour market studies

Turning to the effects of diabetes on the labour market, the existing studies showed, almost consistently, with the exception of Canada (Latif, 2009a) and one study on the USA (Minor, 2013), that the employment probabilities of men were affected more adversely by the disease than those of women. However, while most studies have tried to tentatively explain these gender differences, the reasons for this have not been investigated in depth. The studies also showed that, when interpreting this research, it is important to consider whether a study has tried to account for unobservable factors or reverse causality, as otherwise the results might be misleading. Nonetheless, all studies using IV techniques used similar instruments to achieve identification, providing scope for further research using different identification strategies to further explore how endogeneity might affect the results. What has been apparent is the lack of research on labour market outcomes of diabetes in LMICs, with only one study investigating the effect of diabetes on labour income in China (Liu and Zhu, 2014). This deficit might be due to a limited availability of suitable data sources containing sufficient information to allow for a similar investigation of the topic.

The potential for rich, good-quality data sources to aid the investigation of the economic impact of diabetes can be illustrated by the several studies that used data from the Lower Rio Grande Valley in Texas. These studies demonstrate the evolution of methodology and data from the use of single equation regression models (Bastida and Pagán, 2002a) to the use of IV methods (Brown, Pagan, et al., 2005) and—finally—biometric data on blood glucose values (Brown, Perez, et al., 2011). While the first two methods allowed the investigation of the general effect of diabetes on employment chances, the latter was able to assess the impact according to how diabetes was managed by the patient, as proxied by the measured biomarkers. The study found that the main adverse effect was due to having diabetes regardless of how it was managed and that improvements in management only had minor positive effects. The authors concluded that investments in the prevention of diabetes would likely be more effective than improved diabetes management.

The latter study and the study by Liu and Zhu (2014) also show how biometric data (e.g. blood glucose values) can be used to arrive at a deeper understanding of the economic effects of diabetes. This information makes it possible to investigate the impact of diabetes according to the severity of the disease and also allows for the consideration of previously undiagnosed people with diabetes, increasing the policy relevance of the research.

2.4.3 Comparison of COI and Labour Market Studies: Common Themes and Lessons Learned

The results of both fields, COI and labour market studies, show a considerable adverse impact of diabetes in terms of costs to society, health systems, individuals and employers and in terms of a reduction in the productive workforce and productivity in general. Both research strands particularly indicate that the adverse effects of diabetes increase with diabetes duration as well as with the severity of the disease, judged by the high complication costs estimated in COI studies and the larger employment and income penalties for those with a longer disease duration or higher blood glucose levels.

Nonetheless, several lessons can be learned for each field from advancements in the other field. Future COI studies would, for instance, benefit from the more frequent use of biomarker data. This would allow for a more precise analysis of the costs of diabetes according to the severity of the disease and help inform researchers and policy makers about the possible economic effects of achieving certain treatment goals, e.g., a reduction in blood glucose values.

Also, and in contrast to the labour market outcomes literature, the endogeneity problem has hitherto not been addressed in any form in studies estimating direct healthcare or productivity costs, despite it being an equally important challenge in this domain. A possible bias could arise if some people developed diabetes as a result of an unobserved accident or illness, likely resulting in an overestimation of the costs. Endogeneity could also be introduced if people with diabetes became poorer as a result of the disease and consequently were not able to spend as much on their treatment as they would like to, leading to an underestimation of the true monetary cost of diabetes. Furthermore, an endogeneity bias would be introduced if diabetes was correlated with poverty so that diabetes prevalence would be disproportionately high in subgroups with less resources and consequently less access to care. This would lead to an underestimation of the healthcare costs of diabetes. Endogeneity in COI studies has recently been addressed for the estimation of healthcare costs of obesity, suggesting that direct costs would have been underestimated, had the study not accounted for endogeneity (Cawley and Meyerhoefer, 2012). It appears that, on the basis of the studies identified in our review, a similar—worthwhile—approach could and should be applied to the case of Type 2 Diabetes.

Yet the labour market studies also stand to gain from adopting certain approaches that are more common in COI studies. To date, only few labour market studies have used the incidence approach found for COI studies to follow people with diabetes over a certain time period from their diagnosis onwards, in order to further explore how the effect of diabetes on employment and productivity measures develops over time.

Some further recommendations may be derived for future COI and labour market studies on diabetes:

1. For COI studies the estimation of incremental costs—wherever possible—appears to be most suitable for diabetes, as it more accurately accounts for costs of co-morbidities and for less obviously related disease costs (Honeycutt et al., 2009; Tunceli, Wade, et al., 2010). More information that can guide researchers in their choice of methods already exists and should be referred to when performing a COI study (Akobundu et al., 2006).

2. If possible, the use of convenience samples of people with diabetes visiting a health care institution should be avoided, particularly in LMICs, as it excludes those not able or willing to visit a clinic for treatment due to economic reasons, leaving out a potentially important proportion of diabetes patients.

3. The interpretation of the COI results always hinges on the amount of information provided about, among others, the aim of the study, the perspective adopted and the cost components included as well as the estimation approach used. A discussion of how these choices might affect the estimates should also be part of every COI study. Researchers should therefore consult available guidance from the literature that sets out what information should ideally be included in a COI study (Larg and Moss, 2011) to increase the transparency and usability of their research.

4. For labour market studies more evidence from LMICs is needed. There is scope for for exploring existing household datasets from LMICs that contain information on diabetes (Seuring et al., 2014). In some cases, panel data are (or may come)

available, which would allow the investigation of the effects of diabetes over time as well as to improve the degree of causal inference by controlling for unobserved heterogeneity.

5. As for labour market studies, other ways of achieving identification should be explored to reduce the reliance on IV methods using the family history of diabetes as the sole instrument. The increasing richness of information provided in recent data sets could be used to this effect, also taking into account other quasi-experimental econometric methods (Craig et al., 2012).

2.4.4 Limitations

A possible limitation of this review is the decision to refrain from excluding studies based on certain quality criteria, such as study design, costing methodology, sample size or reporting standards. This might have resulted in the inclusion of lower quality studies with less reliable estimates, compromising the comparability across countries, particularly between LMICs and HICs, as study designs differed considerably. On the other hand our overarching objective was to ensure a truly globally comprehensive overview of the literature on the economic impact of diabetes, including evidence from LMICs, which, for reasons often beyond the control of the researchers, may have been of limited quality and thus would have been excluded, had we applied stringent quality benchmarks. Further, any attempt to apply a quality threshold would have faced the challenge of dealing with the absence of a formal checklist to follow in critically appraising the quality of COI studies. Rather than interpreting it as a limitation, we see the identification and synthesis of LMIC studies as a unique added value of this review, when compared to the Ettaro et al. (2004) review.

Notably, we also abstained from any language restrictions, which would have particularly excluded evidence from Spanish speaking and Eastern European countries. Taken together, these factors have resulted in a large number of included studies, allowing for an (albeit exploratory) statistical investigation of the heterogeneity in diabetes cost estimates as a complement to the narrative analysis. We therefore feel that the advantages of refraining from too stringent inclusion criteria more than outweigh the possible negative consequences of including potentially lower-quality studies.

Further, our search was limited to studies after the year 2000. While for COI studies a previous review covered the literature until 2000, this is not the case for the literature on labour market effects of diabetes and we therefore cannot exclude the possibility of having missed some relevant (if old) studies.[3]

2.5 Conclusion

This review has provided an updated and considerably expanded picture of the literature on the global economic impact of type 2 diabetes. The results show a considerable impact of diabetes in terms of costs to society, health systems, individuals and employers and in terms of a reduction in the productive workforce and productivity in general. Studies on the costs of diabetes now provide evidence from HICs as well as LMICs, using a variety of study designs to estimate the costs of diabetes. The evidence indicates a particularly strong and direct economic impact of type 2 diabetes on people's livelihoods in lower-income settings. Studies on labour market outcomes so far have been confined, almost exclusively, to HICs, leaving space for further studies in LMICs to provide additional evidence of the effect of diabetes in these countries. An issue not yet covered in diabetes COI studies—in striking contrast to labour market outcome studies—has been the possible bias introduced by endogeneity, providing an opportunity for advancing research in this area.

Acknowledgments The work of MS on this paper was partially funded by the Centre for Diet and Activity Research (CEDAR), a UK CRC Public Health Research Centre of Excellence. Funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council (MRC), Medical Research Council, the National Institute for Health Research, and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Author Contributions TS, OA and MS planned the work and finalized the manuscripts. TS carried out the data base search, extracted, analysed and interpreted the data and produced the draft of the manuscript. OA also performed data extraction and contributed to the production of the first draft. MS oversaw the development of the work, contributed to the various drafts of the manuscript and provided guidance. TS is the guarantor for the overall content.

[3] We have checked the references of our included labour market studies for any relevant studies published before 2001. We could find only one relevant study from 1998 investigating how employment chances and family income were affected by diabetes in the USA, comparing samples from 1976, 1988 and 1992 and finding significant adverse effects of diabetes on employment chances but not on family income (Kahn, 1998). The effect for women decreased somewhat between 1976 and 1992, while the effect increased for men. The study did not account for the possible endogeneity of diabetes nor selection bias when estimating the effects on income.

What is endogeneity?

Endogeneity is a statistical problem that occurs in regression models if the assumptions about the flow or direction of causality are incorrect. If endogeneity is ignored, it could be that claims about causality between two variables or the magnitude of the effect are false. In general, one can only be certain about a causal relationship of the effect of x on y if the following three conditions are met (Antonakis et al., 2012):

- y follows x temporally
- y changes as x changes (and this relationship is statistically significant)
- no other causes should eliminate the relation between x and y.

There are three major causes of endogeneity that violate the conditions above.

1. Omitted variables When a regression is run to determine the causal effect of variable x on variable y, but there are unobserved variables that affect variables x or x and y simultaneously, the estimated effect of x on y will be biased. For the case of type 2 diabetes and employment chances, there is the danger that, e.g., personal traits like ambition, which are hard to observe, could influence the probability of developing type 2 diabetes through their effect on a person's lifestyle, but they could also simultaneously affect the chances of employment through their influence on a person's determination to find work or to perform well at work. If we are not able to control for this, then our estimate of the effect of diabetes on employment chances might, at least partially, represent the effect of personal traits on employment chances. As a result, our estimate of the effect of diabetes is biased and does not represent the true size of the relationship between the two variables.

2. Simultaneity Simultaneity is present, if our outcome variable y and our variable of interest x influence each other simultaneously, so that y not only is affected by x but x is also affected by y. In the case of type 2 diabetes and labour market outcomes, not only diabetes could influence employment chances or work related income, but also resulting changes in lifestyle due to employment or an increase in income could affect the probabilities of developing diabetes. Due to an increase in income people could change their diet or change towards a less active lifestyle which in turn would make them more likely to develop type 2 diabetes.

3. Measurement error Measurement errors occur when the independent variable x is imprecisely measured. Here this would be the case if people in a survey did not remember if they have been diagnosed with type 2 diabetes and gave a wrong answer.

There are several solutions to the problem of endogeneity, but only using instrumental variable (IV) techniques has the potential to deal with all three causes of endogeneity at once. Endogeneity is a problem, because the variable of interest, here diabetes, is correlated with the error term of the estimated model, which includes all omitted variables as well as the effect of y on x and if measurement error is present, the true values. To do this, one needs to find a suitable instrument that needs to fulfil the following conditions:

- it has to be causally related to the endogenous variable x and
- it should not be correlated to the dependent variable y other than through its correlation with x.

This instrument is then used in a first regression to obtain predicted values of the problematic endogenous regressor. Because the instrument is not correlated with the error term, these predicted values of the endogenous variable will be uncorrelated as well and can then be used in a second regression to predict the dependent variable y. The estimated coefficients of this second stage can then be regarded as consistent estimates.

In the case of type 2 diabetes and labour market outcomes, an instrument has to predict the development of diabetes without being otherwise causally related to any of the labour market outcomes, be it employment chances, wages or some other measure of productivity. The instrument of choice so far has been the family history of diabetes. It has been shown that a considerable part of the risk of developing type 2 diabetes is hereditary (Hemminki et al., 2010; Herder and Roden, 2011; The Interact Consortium, 2013). This fact is exploited when the instrument is used and it is assumed that this is the only pathway through which a family history of diabetes affects a person's diabetes risk, and also that, e.g., parental diabetes does not affect the person's labour market outcomes directly.

The most common estimation techniques for the estimation of IV regressions are the linear IV model and the bivariate probit model. The latter is often deemed more apt for models where both the outcome as well as the instrumental variable are binary, so either 0 or 1, which is the case for employment as an outcome variable as well as diabetes family history as an instrument. Nonetheless, there is some discussion in the econometrics literature regarding the best method to estimate these cases, as it also has been argued that because the linear IV technique does not depend on the assumption of normality of the error terms, in contrast to the bivariate probit model, its results are more reliable in the case of non-normality, but can sometimes lead to imprecise estimators which can no longer be interpreted meaningfully (Chiburis et al., 2012). Both methods can be found in the reviewed papers.

Country codes

Table 2.7: Country Codes

Country	Country code	Country	Country code
35 developing countries	LMIC	Jamaica	JAM
Argentina	ARG	Japan	JPN
Australia	AUS	Latin America and Caribbean	LAC
Bahamas	BHS	Mexico	MEX
Barbados	BRB	Netherlands	NLD
Belgium	BEL	Nicaragua	NIC
Bolivia	BOL	Nigeria	NGA
Brazil	BRA	Norway	NOR
Canada	CAN	Pakistan	PAK
Chile	CHL	Panama	PAN
China	CHN	Paraguay	PRY
Colombia	COL	Peru	PER
Costa Rica	CRI	Serbia	SRB
Cuba	CUB	Spain	ESP
Czech Republic	CZE	Sudan	SDN
Denmark	DNK	Sweden	SWE
Dominican Republic	DOM	Switzerland	CHE
Ecuador	ECU	Taiwan	TWN
El Salvador	SLV	Thailand	THA
Europe	EUR	The Bahamas, Barbados, Jamaica, Trinidad and Tobago	CARICOM
France	FRA	Trinidad and Tobago	TTO
Germany	DEU	United Arab Emirates	ARE
Guatemala	GTM	United Kingdom	GBR
Guyana	GUY	United States	USA
Haiti	HTI	Uruguay	URY
Honduras	HND	Venezuela	VEN
Hong Kong	HKG	WHO African Region	AFR

Table 2.7: Country Codes

Country	Country code	Country	Country code
India	IND		
Iran, Islamic Rep.	IRN		
Ireland	IRL		
Israel	ISR		
Italy	ITA		

2.6 Tables

Table 2.8: COI study characteristics and cost estimates

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Smith-Spangler et al. (2012)	2002–2003	35 LMIC	121051	General pop.	Patient	RB/M	$				3 at 50th percentile to 157 at 95th percentile	3.40 at 50th percentile to 178 at 95th percentile		
Boutayeb and Boutayeb (2014)	NA	Various Arab countries	NA	General pop.	Healthc. system	SAM	USD				UDD 529[j]			
Barceló et al. (2003)	2000	ARG	1250300	General pop.	Societal	SAM	ARS	16547	1130	15416[b]	597[a]	904[a]	8145[a]	12330[a]
Davis et al. (2006)	2000–2051	AUS	1294	General pop.	Healthc. system	SDS	AUD		1514 (2000), 2282 (2051)		3496[a] (2000)	3379[a] (2000)		
Barceló et al. (2003)	2000	BHS	12800	General pop.	Societal	SAM	BSD	43	25.2	16	1605	2507	1009	1575
Abdulkadri et al. (2009)	2001	BHS	10435	General pop.	Societal	SDS	BSD	233	17	216[b]	836[a]	1310[a]	10789[a]	16914[a]
Abdulkadri et al. (2009)	2001	BRB	28438	General pop.	Societal	SDS	BBD	75	69.2	5	2455	2433	204	202

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Barceló et al. (2003)	2000	BRB	23300	General pop.	Societal	SAM	BBD	307	26	281[b]	1099[a]	1117[a]	11880[a]	12076[a]
Jönsson (2002)	1999	BEL	735 patients	General pop.	Healthc. system	SAM	EUR		1561		3295	4704		
Jönsson (2002)	1999		7000 (overall)	General pop.	Healthc. system	SAM	EUR				2834	Not possible because no country specific estimate		
Barceló et al. (2003)	2000	BOL	153900	General pop.	Societal	SAM	BOB	901	338	563[b]	3435[a]	2199[a]	5717[a]	3659[a]
Barceló et al. (2003)	2000	BRA	4532600	General pop.	Societal	SAM	BRL	54892	9598	45294[b]	1595[a]	2118[a]	1595[a]	9993[a]
Lau et al. (2011)	2008–2035	CAN	147498 with diabetes	Four Alberta Health and Wellness databases	Healthc. system	SAM	CAD		5934 (2007); 20032 (2035)		4563[a]	4023[a]		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Pohar, Majumdar, et al. (2007)	1993–2001	CAN	57774	Saskatchewan Canadians (excluding Indians)	Healthc. system	SAM	CAD				large urban: 3563 (1993), 3454 (2001), small urban:3321 (1993), 3427(2001), rural:3368 (1993), 3289 (2001)	large urban: 2665 (1993), 3591 (2001), small urban: 3453 (1993), 3563 (2001), rural: 3502 (1993), 3420 (2001)		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Pohar and Johnson (2007)	2001	CAN	5284 (Indians) + 41630 (general pop.) with diabetes, 11692 (Indians) + 98680 (general pop.) without diabetes	Registered Indians according to the Indian Act	Healthc. system	RB/M	CAD				Excess costs: Indians 2227, General pop. 2378 (total costs with diabetes: 3,622 for Indians/ 3253 in general pop., controls: 1,395 for Indians/ 875 for general pop.)	Excess costs: Indians 2316, General pop. 2473: (total costs with diabetes: 3766 for Indians/ 3382 in general pop., controls: 1450 for Indians/ 910 for general pop.)		
Barceló et al. (2003)	2000	CHL	496500	General pop.	Societal	SAM	CLP	5890	719	5171[b]	320601[a]	1447[a]	2307131[a]	10416[a]
Wang, Fu, Zhuo, et al. (2010)	2007	CHN	1478	T2D patients in these Chinese hospitals	Healthc. system	Survey	RMB				4564 (median), 7926 (mean)	1246 (median), 2164 (mean)		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Wang, Mc-Greevey, et al. (2009)	2007 and 2030 (projection)	CHN	2040	In-patients and out-patients with DM in 20 hospitals	Societal	Survey	RMB	72916 (2007), 132472 (2030)	67946 (2007), 123187 (2030)	4982 (2007), 9058 (2030)	11555	3401	1586	467
Yang et al. (2012)	2009–2010	CHN	1232 (diabetes), 1201 (no diabetes)	General pop.	Healthc. system	RB/M	RMB				4135 (3.38 times greater than controls)	1136 (3.38 times greater than controls)		
Wang, Fu, Pan, et al. (2009)	2007	CHN	2054	T2D patients in these Chinese hospitals	Healthc. system	Survey	RMB				4800 (median), 10164 (mean)	1412 (median), 2991 (mean)		
extcite-Gonzalez2009b	32 years	COL	NA	Hypothetical average Columbian type 2 DM patient	Societal	SAM	COP	5.3	1.8	3.5	611750	570	1187000	1106
Barceló et al. (2003)	2000	COL	937700	General pop.	Societal	SAM	COP	7737	1241	6496[b]	923826[a]	1323[a]	4836001[a]	6928[a]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Barceló et al. (2003)	2000	CRI	154900	General pop.	Societal	SAM	CRC	1026	210	817[b]	192194[a]	1353[a]	749278[a]	5274[a]
Barceló et al. (2003)	2000	CUB	592400	General pop.	Societal	SAM	CUP	1721	923	798[b]	1219[a]	1558[a]	1054[a]	1347[a]
Horak (2009)	2007	CZE		Insured in health-care system (63.1% of pop.)	Healthc. system	SAM	CHK		190					
Gyldmark and Morrison (2001)	1993	DNK	948	General pop.	Societal	WTP	DKK						1128 (mean), 300 (median)	191 (mean), 51 (median)
Barceló et al. (2003)	2000	DOM	254100	General pop.	Societal	SAM	DOP	1410	509	901[b]	14580[a]	2003[a]	25801[a]	3545[a]
Barceló et al. (2003)	2000	ECU	267300	General pop.	Societal	SAM	USD	2830	1104	1727[b]	873[a]	4129[a]	1366[a]	6460[a]
Barceló et al. (2003)	2000	SLV	219400	General pop.	Societal	SAM	SVC	1385	381	1004[b]	626[a]	1737[a]	1650[a]	4577[a]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Honkasalo et al. (2014)	2005–2010	FIN	1890 with T2D	People with T2D in two cities in Finland	Healthc. system	SDS	EUR				1038	1087		
Ricordeau et al. (2003)	1998, 2000	FRA	704423 (1998), 1145603 (2000) with diabetes	Metropolitan France	Healthc. system	RB/M	EUR		2784 (1998), 3268 (2000)		1529 (1998), 1655 (2000)	2107 (1998), 2241 (2000)		
Jönsson (2002)	1999	FRA	751 patients	General pop.	Healthc. system	SAM	EUR		5478		3064	4214		
Jönsson (2002)	1999	DEU	809 patients	General pop.	Healthc. system	SAM	EUR		1653		3576	4752		
Köster, Ferber, et al. (2006)	2001	DEU	306736 (26971 with diabetes)	General pop.	Societal	RB/M	EUR		Excess: 19364 (total: 40650)		Excess 2507 (total 5262)	Excess: 3329 (total: 6987)	Excess 1328 (total: 5019)	Excess: 1763 (total: 6664)
Köster, Huppertz, et al. (2011)	2000–2007	DEU	320000 (2000) to 275000 (2007)	AOK Hessen	Healthc. system	RB/M	EUR		17299 (2000), 25614 (2007)		2400 (2000), 2605 (2007)	3493 (2007), 3218 (2000)		
Martin et al. (2007)	1995–2003	DEU	3268	Newly diagnosed T2D patients	Healthc. system	SAM	EUR				3210	4075		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Köster, Schubert, et al. (2012)	2000–2009	DEU	not given, only DM patients stated (30472)	AOK Hessen	Healthc. system	RB/M	EUR		21230 (2000), 26226 (2009)		2779 (2000), 2611 (2009)	3471 (2000), 3261 (2009)		
Barceló et al. (2003)	2000	GTM	368700	General pop.	Societal	SAM	GTQ	2535	878	1657[b]	6131[a]	2382[a]	11572[a]	4495[a]
Barceló et al. (2003)	2000	GUY	28400	General pop.	Societal	SAM	GYD	141	80	62[b]	131041[a]	2800[a]	102135[a]	2182[a]
Barceló et al. (2003)	2000	HTI	79500	General pop.	Societal	SAM	HTG	249	152	97[b]	12782[a]	1912[a]	8175[a]	1223[a]
Barceló et al. (2003)	2000	HND	193000	General pop.	Societal	SAM	HNL	772	366	405[b]	8750[a]	1898[a]	9680[a]	2100[a]
Chan et al. (2007)	2004	HKG	147	T2D patients attending the DM outpatient clinic at a public hospital	Societal	Survey	USD				11638	2288	1817[e]	357[e]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Ramachandran et al. (2007)	1998, 2005	IND	556 with T2D (urban=309, rural=247)	T2D patients in India	Patient	Survey	INR				Median values: 10000 (urban), 6260 (rural)	Median values: 773 (urban), 484 (rural)		
Tharkar et al. (2010)	2009	IND	718	Diabetes patients in Chennai city	Societal	Survey	INR		268		25391 (median)	1557 (median)	4970 (median)	305 (median)
Javanbakht et al. (2011)	2009	IRN	4500	Diabetes patients from Tehran and Fars province	Societal	Survey	IRR	9611[h]	5187[h]	4420[h]	8358592	2142	8578816	2199
Esteghamati et al. (2009)	2004, 2005	IRN	710 (T2D), 904 (controls)	Pop. in Teheran	Societal	RB/M	IRR	401 (Teheran); 2117[h] (Iran)	327 (Teheran); 1727[h] (Iran)	74 (Teheran), 390[h] (Iran)	876622 (Teheran)	443 (Teheran)	200146 (Teheran)	101 (Teheran)
Nolan et al. (2006)	1999	IRL	701	T2D patients of four Irish hospitals	Healthc. system	SAM	EUR				2469	2867		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Chodick et al. (2005)	2001	ISR	24632	Insured patients in HMO	Healthc. system	RB/M	ILS		433		6002 (2001), 3926 (1999)	1950 (2001), 1275 (1999)		
Lucioni et al. (2003)	1998	ITA	1263	T2D patients from randomly drawn practices across Italy	Societal	SAM	EUR	8289[d]	7930	359	2991	4588	135[ac]	208[ac]
Bruno et al. (2012)	2003–2004	ITA	33792 (diabetes) and 863123 (no diabetes)	Turin pop.	Healthc. system	RB/M	EUR				2465 (3361 (diabetes), 896 (no diabetes)	3328 (4537 (diabetes), 1210 (no diabetes)		
Morsanutto et al. (2006)	2001–2002	ITA	299	T2D patients who visited a diabetologic center in Italy (DC)	Healthc. system	SAM	EUR				1910	2823		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Marchesini et al. (2011)	2006	ITA	311979	People with DM at 22 local health districts	Healthc. system	RB/M	EUR				2589	3296		
Abdulkadri et al. (2009)	2001	JAM	186036	General pop.	Societal	SDS	JMD	556	454	102	44647	2439	10046	549
Barceló et al. (2003)	2000	JAM	181400	General pop.	Societal	SAM	JMD	1037	345	693[a]	32251[a]	1901[a]	64787[a]	3818[a]
Nakamura et al. (2008)	1990–2001	JPN	4535	Community dwelling in Shiga	Healthc. system	SAM	JPY				189060 (diabetes), 99900 (non-diabetes)	1674 (diabetes), 884 for (non-diabetes)		
Barceló et al. (2003)	2000	LAC	Diabetes prevalence of 15.2 million	Pop. from all countries in Latin America and Caribbean	Societal	SAM	USD	82304	13529	68774[b]	703[a]	887[a]	3576[a]	4512[a]
Barceló et al. (2003)	2000	MEX	3738000	General pop.	Societal	SAM	MXN	30677	4006	26671[b]	4994[a]	1072[a]	33249[a]	7135[a]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Arredondo, Zúñiga, and Parada (2005)	2004, 2006	MEX	951417 estimated cases	All users of health care in public institutions	Societal	SAM	MXN	290[d]	229	61k	1472[a]	242[a]	386[a]	64[a]
Arredondo and De Icaza (2011a)	2010	MEX	Whole pop.	Population demanding services at Mexican health-care institutions for T2D	Societal	SAM	MXN	1066	470	596	4016[a]	485[a]	5090[a]	610[a]
Arredondo and Barcelo (2007)	2005	MEX	Whole pop.	General pop.	Patient	SAM	MXN		284 OOP expenditures (52% of overall expenditures)					

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Arredondo and Zúñiga (2004)	2003, 2005	MEX	Whole pop.	General pop. using public health-care institutions	Societal	SAM	MXN	532 (2005)	235 (2005)	297 (2005)	1467[a] (2005)	263[a] (2005)	1852[a] (2005)	331[a] (2005)
Rodríguez Bolaños et al. (2010)	2002, 2004	MEX	497	IMSS insured	Healthc. system	SDS	MXN		661 (2004)		35622[a] (2004)	4672[a] (2004)		
Redekop et al. (2002)	1998	NLD	1371 with T2D	T2D patients in the Netherlands	Societal	SAM	NLG	1014[d]	953	61	4023	2780	282[a]	195[a]
Linden et al. (2009)	2000–2004	NLD	2.5 million (641200 with diabetes)	Dutch people with diabetes	Healthc. system	SDS	EUR		571 (2000), 1063 (2004)		974 (2000), 1283 (2004)	1259 (2000), 1658 (2004)		
Jönsson (2002)	1999	NLD	909 patients	General pop.	Healthc. system	SAM	EUR		671		1827	2761		
Barceló et al. (2003)	2000	NIC	136100	General pop.	Societal	SAM	NIO	442	292	150[b]	7922[a]	2145[a]	4082[a]	1105[a]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Suleiman et al. (2006)	July 2003–June 2004	NGA	35	Diabetes patients in out-patient clinic in Nigeria	Patient	SDS	NGN				29366	662		
Solli et al. (2010)	2005	NOR	4.6 million from register data of entire pop.	General pop.	Societal	SDS	NRK	319	242	76	20492[a]	2061[a]	5067[a]	650[a]
Khowaja et al. (2007)	2006	PAK	345	Diabetes patients in Karachi	Societal	Survey	PKR				11580[f]	620[f]	840[e]	45[e]
Barceló et al. (2003)	2000	PAN	120500	General pop.	Societal	SAM	PAB	926	222	704[b]	866[a]	1846[a]	2741[a]	5840[a]
Barceló et al. (2003)	2000	PRY	94300	General pop.	Societal	SAM	PYG	738	244	495[b]	2661903[a]	2587[a]	5397747[a]	5245[a]
Barceló et al. (2003)	2000	PER	606800	General pop.	Societal	SAM	PEN	5627	1533	4094[b]	2890[a]	2526[a]	7717[a]	6746[a]
Lesniowska et al. (2014)	2009	POL	Whole pop.	All Polish diabetes patients	Healthc. system	SAM	RSD	3396	1910	1486				

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Biorac et al. (2009)	2007	SRB	99	T2D patients in health centre in Svilajnac	Societal	Survey	RSD	7579[h]			47865	1610	5548	187
Bjegovic et al. (2007)	2002	SRB	360433 people with T2D in Serbia	Serbian T2D patients	Healthc. system	SAM	RSD		280		12457[a]	761[a]		
Mata et al. (2002)	1998	ESP	1004	Diabetes patients from 29 primary healthcare centres	Healthc. system	SDS	EUR				771	1488		
Ballesta et al. (2006)	1999	ESP	517	People with DM in region of Cadiz	Societal	SDS	EUR				2560	4690	1844	3379
Oliva et al. (2004)	2002	ESP	1675304 to 2010365 depending on assumed prevalence	Diabetes patients in National Health System	Healthc. system	SAM	EUR		4010 (6% prev.)–4461 (5% prev.)		1290 (6% prev.)–1476 (5% prev.)	2155 (6% prev.)–2466 (5% prev.)		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Jönsson (2002)	1999	ESP	1004 patients	General pop.	Healthc. system	SAM	EUR		3679		1305	2453		
Bastida, Aguilar, et al. (2002)	1998	ESP	Whole pop. (exact number not given)	Canary Island pop. with diabetes	Societal	SDS	Pts (pre Euro)	75	47	28	78240	907	47928[b]	556[b]
Elrayah-Eliadarous et al. (2010)	2005	SDN	822	Patients with T2D in Khartoum state in Sudan	Patient	Survey	USD				438	456		
Bolin et al. (2009)	1987 and 2005	SWE	Whole pop.	General pop.	Societal	SDS	SEK	499 (1987), 1045 (2005)	223 (1987), 383 (2005)	276 (1987), 662 (2005)	12102 (1987), 12287 (2005)	1484 (1987), 1507 (2005)	15000[a] (1987), 21253[a] (2005)	1840[a] (1987), 2606[a] (2005)
Norlund et al. (2001)	1993	SWE	70786 (1677 with diabetes)	Southern Sweden	Societal	RB/M	SEK				19411	2855	14777	2174
Wiréhn et al. (2008)	2005	SWE	415990 (19226 with diabetes)	Whole Östergötland population	Healthc. system	RB/M	EUR				18293	2243		
Jönsson (2002)	1999	SWE	773 patients	General pop.	Healthc. system	SAM	SEK		929		24927	3319		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Ringborg et al. (2008)	2004	SWE	8230	Diabetes patients in Uppsala county	Healthc. system	SAM	SEK				33210	3888		
Schmitt-Koopmann et al. (2004)	1998	CHE	1479	T2D patients from randomly drawn practices across Switzerland	Healthc. system	SDS	CHF		561		3004	2030		
Lin et al. (2004)	1998–1999	TWN	20757185 (in 1998), 21089859 (in 1999)	People with DM in National Health Insurance	Healthc. system	SDS	TWD				62617 (1998), 60775 (1999)	3499 (1998), 3396 (1999)		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Chang (2010)	2006–2007	TWN	498	Diabetes patients in outpatient clinics in northern Taiwan	Societal	WTP	TWD			4003			68118	4004
Chi et al. (2011)		TWN	16094	Elderly with DM in Taiwan	Healthc. system	SAM			51		111982	6338		
Chatterjee et al. (2011)	2008	THA	475	Diabetes patients treated in district hospital	Societal	Survey	TWD				17638	1082	10569	649
Barceló et al. (2003)	2000	TTO	71300	Pop. from all countries in Latin America and Caribbean	Societal	SAM	TTD	540	72	468[b]	3358[a]	1011[a]	21780[a]	6560[a]
Abdulkadri et al. (2009)	2001	TTO	135093	General pop.	Societal	SDS	TTD	852	227	625	5722	1677	15797	4628

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Al-Maskari et al. (2010)	2004	ARE	150	Diabetes patients in Al-Ain District	Healthc. system	Survey	AED				no complication: 5906, with complications: 20774, overall: 16115	no complications: 2047, with complications: 7199, overall: 5585		
Jönsson (2002)	1999	GBR	756 patients	General pop.	Healthc. system	SAM	GBP		244		1558	3065		
Dall, Zhang, et al. (2010)	2007	USA	Diabetes prevalence of 16.5 million	General pop.	Societal	SDS	USD	167862	111257	56604	6414	6751	3263	3434
Buescher et al. (2010)	1998	USA	127991	Medicaid pop.	Healthc. system	SDS	USD		540		4098	4221		
Dall, Nikolov, et al. (2003)	2002	USA	Diagnosed DM prevalence of 12.1 million	General pop.	Societal	SDS	USD	161896	112947	48948	7601[a]	9346[a]	3294[a]	4050[a]
Druss et al. (2001)	1996	USA	23200	General pop.	Societal	Survey	USD	78518	13768	4771	1097	1495	380[ac]	518[ac]

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Durden et al. (2009)	2000, 2005	USA	21592 (2000), 127254 (2005)	Employees of large, privately-insured companies	Healthc. system	RB/M	USD				7365 (2000), 7327 (2005)	8349 (2000), 8306 (2005)		
Trogdon and Hylands (2008)	2000–2004	USA	3790 (diabetes), 42413 (no diabetes)	General pop.	Healthc. system	RB/M	USD				5035[i]	5708[i]		
Brandle et al. (2003)	2000	USA	1364	People with T2D enrolled in managed care programs	Healthc. system	SAM	USD				3715 (median)	4747 (median)		
O'Connell et al. (2012)	2005	USA	32052	American Indians in and around Phoenix, Arizona	Healthc. system	RB/M	USD				5542	6282		
Peele et al. (2002)	1996	USA	20937 with diabetes	Employed DM patients	Healthc. system	SAM	USD		126		4430 (17.9% OOP)	6039 (17.9% OOP)		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Rodbard et al. (2010)	2006	USA	3551 (diabetes), 8686 (no diabetes)	General pop.	Patient	RB/M	USD				233	264		
Honeycutt et al. (2009)	1998–2003	USA	96873 (5289 had diabetes)	General pop.	Healthc. system	SDS and RB/M	USD		61958 (regression), 43452 (attributable fraction)		4240 (regression), 2980 (attributable fraction)	4966(regression), 3490 (attributable fraction)		
Maciejewski and Maynard (2004)	1998	USA	429918	USA veterans	Healthc. system	SAM	USD		2214		3888[a]	5150[a]		

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Birnbaum et al. (2003)	1997–1998	USA	3759 (diabetes), 3759 (without diabetes)	Employed and retired women	Healthc. system	RB/M	USD				5.500 for women <age 65 per year, 25000 for women >= age 65 per year, 233000 lifetime costs	6680[f]or women <age 65 per year, 30362 for women >= age 65 per year, 282973 lifetime costs		
Zhou et al. (2005)	10 year follow up	USA	1223 with T2D	People with DM in Michigan	Healthc. system	SAM	USD				7100 (undiscounted per year over 10 year period)	9072 (undiscounted per year over 10 year period)		
Dall, Mann, et al. (2008)	2007	USA	Diagnosed DM prevalence of 17.5 million	General pop.	Societal	SDS	USD	185682	123788	62108	6649	7095	3328	3552

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Tunceli, Wade, et al. (2010)	2007	USA	256245 (T2D), 256223 (controls)	Non-institutionalized adults	Healthc. system	SDS and RB/M	USD				Matching: 4217, Disease attributable: 3002	Matching: 4500, Disease attributable: 3204		
Condliffe et al. (2013)	2007	USA	7514 with diabetes	USA pop. with positive healthcare expenditures in survey	Healthc. system	SAM	USD				11167[g]	11917[g]		
Ramsey et al. (2002)	1998	USA	8748 diabetes patients, 8748 matched controls	Employees of large, privately-insured companies	Employer	RB/M	USD				3842	5021	568	743

Ref.	Horizon	Country	Sample size	Population	Perspective	Approach	LCU	Aggregate costs (mill. $)			Per capita costs			
								Total	Direct	Indirect	Direct (LCU)	Direct ($)	Indirect (LCU)	Indirect ($)
Lee et al. (2006)	2000	USA	984 with DM (540 white, 210 African American, 234 Hispanic)	White, African Americans and Hispanics in the USA	Healthc. system	SAM	USD				6616 (6887 if white, 6162 if African American, 5647 if Hispanic)	8453 (8799 if white, 7873 if African American, 7215 if Hispanic)		
Barceló et al. (2003)	2000	URY	119000	General pop.	Societal	SAM	UYU	1202	147	1055[b]	9619[a]	1233[a]	69171[a]	8867[a]
Barceló et al. (2003)	2000	VEN	610800	General pop.	Societal	SAM	VEF	4820	317	4503[b]	342[a]	518[a]	2100[a]	7373[a]
Kirigia et al. (2009)	2005	WHO African region	7020000	General pop.	societal	SAM	USD	28610	9090	19520	876	983	10556	11845

DM Diabetes Mellitus Healthc. System Healthcare system LCU Local currency unit Pop. Population Prev. Prevalence Ref. Reference RB/M regression based/matching SAM Sum-all medical SDS Sum-diagnosis specific.

[b] [a] Own calculation dividing presented aggregate cost estimate by number of people with diabetes in study. Total and direct cost estimates were presented in paper and indirect costs calculated, but not explicitly stated. We calculated indirect costs by deducting the presented direct costs estimate from the presented total costs estimate to arrive at an indirect costs estimate.

[c] Calculated the number of people with diabetes by dividing the aggregated direct costs and the per capita direct costs estimate as presented in the study.

[d] Calculated total costs of diabetes for papers summing up direct and indirect costs.

[e] Calculated per capita indirect costs deducting direct from total cost estimate presented in study.

[f] Costs originally presented per visit, to arrive at yearly costs had to multiply costs per visit by number of visits per year.

[g] Per capita direct costs were presented for different groups of diabetics, calculated average costs for person with diabetes by summing up and weighting costs people with diabetes + hypertension, people with diabetes + obesity, people with diabetes + obesity + hypertension.

[h] The study assumes sample would be nationally representative.

[i] Study only reported the adjusted incremental cost ratio of 2.39 compared to the average healthcare expenditures of people without diabetes of USA$3630. To calculate the incremental costs of a person with diabetes we multiplied the average healthcare expenditures of people without diabetes by the given cost ratio .

Table 2.9: COI study costing components

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Smith-Spangler et al. (2012)	LMIC (2002-2003)	No breakdown of costs provided								
Kirigia et al. (2009)	AFR (2000-2005)	x	x	x	x	x	x	x	x	No exact information on share in expenditures is available
Davis et al. (2006)	AUS (1993-1996)	x	x	x	x	x	x			No exact information on share in expenditures is available
Lau et al. (2011)	CAN (1995-2007)	x	x	x						Hospital, physician
Pohar, Majumdar, et al. (2007)	CAN (1993-2001)	x	x	x	x	x	x			Hospital, medication
Ohinmaa et al. (2004)	CAN (1996)	x	x	x	x	x	x			Hospital, medication
Dawson et al. (2002)	CAN (1998)	x	x	x	x	x				No exact information on share in expenditures is available
Johnson et al. (2006)	CAN (1992-2001)	x	x	x	x					Hospital
Simpson et al. (2003)	CAN (1991-1996)	x	x	x	x					Hospital, prescription drugs
Pohar and Johnson (2007)	CAN (1991-2001)	x	x	x						Hospital

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Wang, Fu, Zhuo, et al. (2010)	CHN (2007)	x	x	x				x		Complications, insulin therapy
Wang, McGreevey, et al. (2009)	CHN (2007)	x	x					x		Hospital, outpatient visits
Yang et al. (2012)	CHN (2009-2010)	x	x	x	x	x	x			Hospital, medication
Wang, Fu, Pan, et al. (2009)	CHN (2007)	x	x	x	x	x	x	x		No exact information on share in expenditures available
Camilo González et al. (2009)	COL (2007)	No breakdown of costs provided								
Horak (2009)	CZE (2007)	x	x	x	x	x	x			Hospital, medication
Honkasalo et al. (2014)	FIN (2005-2010)	x	x	x	x	x	x			
Ricordeau et al. (2003)	FRA (1998,2000)	x	x	x				x		Hospital, medication
Köster, Ferber, et al. (2006)	DEU (2001)	x	x	x	x	x	x	x		Hospital, medication
Köster, Huppertz, et al. (2011)	DEU (2000-2007)	x	x	x	x	x	x	x	x	Hospital, other services (medical devices, remedies, professional home nursing, transportation)
Martin et al. (2007)	DEU (1995-2003)	x	x	x	x	x	x			No exact information on share in expenditures available
Köster, Schubert, et al. (2012)	DEU (2000-2009)	x	x	x	x	x	x	x	x	Hospital, medication

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Jönsson (2002)	EUR (1999)	x	x	x	x	x	x	x		Hospital, medication
Chan et al. (2007)	HKG (2004)	x	x	x	x	x	x	x	x	Hospital, outpatient clinic visits
Ramachandran et al. (2007)	IND (2005)	x	x	x	x	x	x			Hospital/surgery, medication
Tharkar et al. (2010)	IND (2009)	x	x	x				x		Hospital, medication
Javanbakht et al. (2011)	IRN (2009)	x	x	x	x	x	x	x	x	Complications, medication
Esteghamati et al. (2009)	IRN (2004;2005)	x	x	x	x	x	x	x		Hospital, medication and devices
Nolan et al. (2006)	IRL (1999-2000)	x	x	x	x	x				Hospital, ambulatory/drug costs
Chodick et al. (2005)	ISR (1999-2001)	x	x	x	x					Medication and lab/diagnostics
Lucioni et al. (2003)	ITA (1999)	x	x	x	x	x				Hospital, drugs
Bruno et al. (2012)	ITA (August 2003-July 2004)	x	x		x	x				Hospital, drugs
Morsanutto et al. (2006)	ITA (Jan 2001-Aug 2002)	x		x	x	x				Hospital, drugs
Marchesini et al. (2011)	ITA (1997-2006)	x		x	x	x	x			Hospital, drugs
Nakamura et al. (2008)	JPN (1990-2001)					No breakdown of costs provided				
Barceló et al. (2003)	LAC (2000)	x	x	x	x					Medication, complications

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Arredondo, Zúñiga, and Parada (2005)	MEX (1989-2003)	x	x	x	x	x				No exact information on share in expenditures available
Arredondo and De Icaza (2011a)	MEX (1990-2008)	x	x	x	x	x				Medication, complications
Arredondo and Zúñiga (2004)	MEX (1989-2002)	x	x	x	x	x				Drugs, complications
Arredondo and Barcelo (2007)	MEX (2002-2004)	x	x	x	x	x				Drugs, complications
Rodríguez Bolaños et al. (2010)	MEX (2002-2004)	x	x	x	x	x	x		x	Hospital, administrative costs
Redekop et al. (2002)	NLD (1998)	x	x	x	x	x	x	x		Hospital, medication
Linden et al. (2009)	NLD (2000-2004)	x			x					Hospital, medication
Suleiman et al. (2006)	NGA (2003-2004)		x		x	x	x	x	x	Drugs, diagnostic tests
Solli et al. (2010)	NOR (2005)	x	x	x	x		x		x	Drugs, medical devices
Khowaja et al. (2007)	PAK (2006)		x		x	x		x		Medicine cost, laboratory costs
Lesniowska et al. (2014)	POL (2005-2009)	x	x	x	x	x	x			Medication, primary care
Biorac et al. (2009)	SRB (2007)	x	x	x	x	x	x			Medication, medical services (incl. ambulatory and hospital costs)

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Bjegovic et al. (2007)	SRB (2002)		x	x	x	x	x			No exact information on share in expenditures available
Mata et al. (2002)	ESP (1998-1999)	x	x	x	x	x	x			Drugs, hospital
Ballesta et al. (2006)	ESP (1999)	x	x	x	x		x		x	Medication, hospital
Oliva et al. (2004)	ESP (2002)	x	x	x						Hospital, medication
Bastida, Aguilar, et al. (2002)	ESP (1998)	x	x	x	x	x				Hospital, medication
Elrayah-Eliadarous et al. (2010)	SDN (2005)		x		x	x				Outpatient clinic, drugs
Bolin et al. (2009)	SWE (1987 and 2005)	x	x		x					Hospital, drugs
Norlund et al. (2001)	SWE (1992-1993)	x	x	x				x		Hospital, home help hours
Wiréhn et al. (2008)	SWE (2005)	x	x	x						Hospital, medication
Ringborg et al. (2008)	SWE (2000-2004)	x	x		x	x	x			Hospital, outpatient visits
Schmitt-Koopmann et al. (2004)	CHE (1998-1999)	x	x	x						Hospital, medication
Lin et al. (2004)	TWN (1998-1999)	x	x	x	x	x				No exact information on share in expenditures available
Chi et al. (2011)	TWN (2000)	x	x							Outpatient visits

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
Chatterjee et al. (2011)	THA (2007-2008)	x	x		x	x		x	x	Informal care, hospitalizations
Abdulkadri et al. (2009)	CARICOM (2001)	x	x	x	x	x				Medication and lab/diagnostics
Al-Maskari et al. (2010)	ARE (2004-2005)	x	x	x	x	x				Hospital (information on other cost components not presented)
Dall, Zhang, et al. (2010)	USA (2007)	x	x	x	x	x	x	x	x	No exact information on share in expenditures available
Ramsey et al. (2002)	USA (1998)	x	x	x	x	x	x		x	Inpatient, outpatient
Buescher et al. (2010)	USA (1998)	x	x	x	x	x	x	x	x	Physician visits, hospital
Dall, Nikolov, et al. (2003)	USA (1998-2000)	x	x	x	x	x	x			Institutional care (nursing home stays, hospital), outpatient care
Druss et al. (2001)	USA (1996)	No breakdown of costs provided. Only self reported healthcare cost estimate.								
Durden et al. (2009)	USA (2000, 2005)	x	x	x	x	x	x			Hospital, outpatient services
Trogdon and Hylands (2008)	USA (2000-2004)	No breakdown of costs provided. Only self reported healthcare cost estimate.								
Brandle et al. (2003)	USA (2000-2001)	x	x		x	x				No exact information on share in expenditures is available

Ref.	Country (year of cost data)	Hospital Visits	Outpatient Visits	Physician Visits	Drugs	Laboratory	Equipment	Non-medical	Other Costs	Two main cost contributors
O'Connell et al. (2012)	USA (2004-2005)	x	x	x						Hospital, medication
Peele et al. (2002)	USA (1996)	x	x	x		x				No exact information on share in expenditures available
Rodbard et al. (2010)	USA (2006)				No breakdown of costs provided.					
Honeycutt et al. (2009)	USA (1998-2003)	x	x	x	x	x	x			No exact information on share in expenditures available
Maciejewski and Maynard (2004)	USA (1998)	x	x							Hospital
Birnbaum et al. (2003)	USA (1997-1998)			No breakdown of costs provided. Only self reported healthcare cost estimate.						
Zhou et al. (2005)	USA (2000)	x	x	x	x	x	x			No exact information on share in expenditures available
Dall, Mann, et al. (2008)	USA (2006)	x	x	x						Hospital, medication
Tunceli, Wade, et al. (2010)	USA (2006-2007)	x	x	x						Hospital, medication
Condliffe et al. (2013)	USA (2004-2007)				No breakdown of costs provided.					
Lee et al. (2006)	USA (2000)		x	x				x	x	Medication, ambulatory

3 The Impact of Diabetes on Employment in Mexico

Abstract

This study explores the impact of diabetes on employment in Mexico using data from the Mexican Family Life Survey (MxFLS) (2005), taking into account the possible endogeneity of diabetes via an instrumental variable estimation strategy. We find that diabetes significantly decreases employment probabilities for men by about 10 percentage points (p<0.01) and somewhat less so for women—4.5 percentage points (p<0.1)—without any indication of diabetes being endogenous. Further analysis shows that diabetes mainly affects the employment probabilities of men and women above the age of 44 and also has stronger effects on the poor than on the rich, particularly for men. We also find some indication for more adverse effects of diabetes on those in the large informal labour market compared to those in formal employment. Our results highlight—for the first time—the detrimental employment impact of diabetes in a developing country.

3.1 Introduction

Diabetes, similar to other conditions that have been coined ”diseases of affluence”, has traditionally been seen as mostly a problem of the developed, more affluent countries. Only in recent years the awareness has been growing of the sheer size of the problem in health terms (Hu, 2011b; Yach et al., 2006). Mexico is one example of a middle-income country that has seen diabetes rates increase sharply over the last years, from about 7.5 percent in 2000 (Barquera, Campos-Nonato, et al., 2013) to 12.6 percent in 2013 (International Diabetes Federation, 2014). The high prevalence of diabetes in Mexico reflects an epidemiological transition from a disease pattern previously characterized by high mortality and infectious diseases to low-mortality rates and non-communicable diseases (NCDs) affecting predominantly adults (Stevens et al., 2008). This transition has likely been reinforced by nutritional changes away from a traditional diet towards an energy dense, but nutritionally poor diet with an increasing amount of processed foods and sugars (Barquera, Hernandez-Barrera, et al., 2008; Basu et al., 2013; Rivera, Barquera, González-Cossío, et al., 2004), a reduction in physical activity, as well as what appears to be a particular genetic predisposition of many Mexicans to develop type 2 diabetes (Williams et al., 2014). While many of the high-income countries may be in a position to cope resource-wise with the health care consequences of diabetes, this will be less so the case for Mexico and other LMICs. The most recent ”cost-of-illness” estimates put the costs of diabetes to the Mexican society at more than US$778 million in 2010, with a large part of these costs being paid out-of-pocket (Arredondo and De Icaza, 2011b). While the above includes some estimate of indirect costs, meant to capture the cost burden attributable to foregone productivity resulting from diabetes, there exists no rigorous, econometric assessment of the effect of diabetes on employment chances for Mexico, as the research has thus far focused on high-income countries (Bastida and Pagán, 2002b; Brown, Pagán, et al., 2005; Latif, 2009b; Lin, 2011; Minor, 2011b; Vijan et al., 2004b; Zhang et al., 2009b).

There are several reasons to expect a significant adverse effect of diabetes on employment chances in Mexico and that this effect might be stronger than in high-income countries. In Mexico type 2 diabetes is increasingly affecting people in their productive age, raising the possibility that a larger share of people with diabetes will have to cope with debilitating complications already relatively early in life (Barquera, Campos-Nonato, et al., 2013; Villalpando et al., 2010). Further, only a minority of Mexicans appears to successfully manage their diabetes condition, with as much as 70 percent of the people with diabetes having poor control over their disease (Villalpando et al., 2010). In addition, many Mexicans are working in the large informal economy[1], possibly limiting their access to quality health care and hence to appropriate treatment options. All these factors are likely to both increase the risk of developing debilitating diabetes complications as well as to reduce productivity as a result. Against this background, the aim of this study is to investigate how diabetes affects employment probabilities in a middle-income country such as Mexico. To the best of our knowledge this is the first such paper on Mexico and indeed on any LMIC. We also investigate if the impact of diabetes on employment chances differs across age groups and—again for the first time in this field—by wealth, as well as between those formally and informally employed.

The majority of the more recent studies on the labour market impact of diabetes tried to account for the possible endogeneity of diabetes using family history of diabetes as an instrument. Endogeneity might arise due to reverse causality: employment status and its effect on a person's lifestyle may also influence the odds of developing diabetes. A job with long office working hours might push a person's diet or exercise pattern towards a more unhealthy and sedentary lifestyle due to reduced leisure time, increasing the person's risk for diabetes. In addition, unobserved factors, such as personal traits, could simultaneously influence a person's employment as well as his or her diabetes status and introduce an omitted variable bias. A less ambitious person could be less productive in a job, increasing the risk of being laid off, and he or she could simultaneously have only modest, if any, exercise goals or healthy eating habits, thereby increasing the chances of developing diabetes.

Brown, Pagán, et al. (2005) estimated the impact of the disease on employment in 1996–1997 in an older population of Mexican Americans in the USA close to the Mexican border, using a recursive bivariate probit model. They found diabetes to be endogenous for women but not for men. The results of the IV estimation suggested no significant effect on women which, compared to the adverse effect found in the probit model, indicated an overestimation of the effect for women when endogeneity was not accounted for. For men, the probit estimates showed a significant adverse effect of about 7 percentage points. Latif (2009b) estimated the effect of the disease on employment probabilities in Canada in 1998. Contrary to Brown, Pagán, et al.

[1]In 2005 around 58 percent of the working population in Mexico were employed in the informal sector (Aguila et al., 2011).

(2005), he found diabetes to be exogenous for females and endogenous for males; taking this into account he obtained a significant negative impact on the employment probabilities for women, but not for men. Because the simple probit model showed a significant negative effect for males, Latif (2009b) concluded that not accounting for endogeneity resulted in an overestimation of the effect on male employment chances. Minor (2011b) investigated the effect of diabetes on female employment, among other outcomes, in the USA in 2006. This particular study differed from earlier work in that it not only analysed the effects of diabetes in general, but also of type 1 and type 2 diabetes separately. The study found diabetes to be endogenous and underestimated if exogeneity was assumed. In the IV estimates, type 2 diabetes had a significant negative effect on female employment chances. For Taiwan, Lin (2011) found diabetes to be endogenous, with the IV results showing significant changes in the employment effect of diabetes. The impact was found to be significantly negative for men in the IV model indicating an underestimation in the standard probit model, where the diabetes coefficient was also significant but much smaller in size. For women, no significant effect was found in the IV estimation after the probit model had indicated a significant and negative impact of diabetes.

Accordingly, at least in some cases, there seems to be the risk of biased estimates of the impact of diabetes on employment, when exogeneity is assumed, with an a priori ambiguous bias. Hence, our decision in this study to also assess if diabetes is endogenous and how precisely taking account of endogeneity might affect the estimates. In order to account for this possible endogeneity we use data from the second wave of the Mexican Family Life Survey (MxFLS) from 2005, which not only provides information on people's diabetes status and socioeconomic background, but also on parental diabetes, enabling us to construct an instrumental variable similar to what has been used in the previous literature on high-income countries.[2] The data also allows the extension of the analysis to test if the inclusion of information on parental education as an additional control variable affects the IV parameter estimates.

The remainder of the paper is structured as follows. Section 3.2 provides details about the used dataset and the econometric specification; and section 4.5 presents and discusses the empirical results. Section 4.6 concludes.

3.2 Methodology

3.2.1 Dataset and descriptive statistics

The dataset used for the empirical analysis is the Mexican Family Life Survey (MxFLS). It is a nationally representative household survey which was conducted in 2002 and 2005. We use data from the second wave in 2005, which includes almost 40,000 individuals. Interviews

[2]Studies that have used the family history of diabetes as an instrument for diabetes are Brown, Pagán, et al. (2005) for a Mexican-American community, Latif (2009b) for Canada, Minor (2011b) for females in the USA and Lin (2011) for Taiwan.

were conducted with all household members aged 15+, and information on a wide range of social, demographic, economic and health related topics was collected (Rubalcava and Teruel, 2008). While there are more recent datasets available on Mexico, none of these provide as extensive information on parental characteristics as does the MxFLS which includes information on parental diabetes and education status, even if parents were not alive anymore or were living in a non-surveyed household at the time of the survey. Diabetes is self-reported and 3.7 percent of males and 5.1 percent of females report a diagnosis by a doctor.[3] Unfortunately we cannot—with the data at hand—distinguish between the different types of diabetes. It can be assumed, however, that about 90 percent of the reported diagnoses are due to type 2 diabetes, which is by far the most common type of diabetes (Sicree et al., 2011). The sub-sample used for analysis is limited to the age group of 15 to 64 years, which represents the majority of the working population. To allow for heterogeneity in the coefficients across gender, the sample has been split to estimate the male and female groups separately.

The descriptive statistics presented in Table 3.1 suggest that the groups of respondents with and without diabetes differ significantly in various aspects. Both males and females with diabetes have a lower employment rate than their counterparts. This would suggest that diabetes has a negative impact on the employment chances of both males and females with diabetes. However, since the groups with diabetes are also significantly older and differ in terms of education, this may be a spurious relationship. As a result, only a multivariate analysis will provide more reliable information on how diabetes truly affects employment probabilities.

3.2.2 Econometric specification

We first estimate a probit model with the following specification

$$Employed_i = \beta_0 + \beta_1 Diabetes_i + \beta_2 X_i + u_i \tag{3.1}$$

where diabetes is assumed to be exogenous. $Employed_i$ takes the value of 1 if person i is employed and 0 if unemployed. Employment status is defined as having worked or carried out an activity that helped with the household expenses for at least ten hours over the last week. This explicitly includes those employed informally, for instance people working in a family business. $Diabetes_i$ denotes the main independent variable of interest, taking the value of 1 if individual

[3] This is well below the estimated prevalence rate for 2013 of almost 12 percent. This is likely due to the fact that, according to the International Diabetes Federation (IDF), more than half of the people with diabetes in Mexico are undiagnosed and consequently did not report it (International Diabetes Federation, 2014). Further, the sample in the survey at hand is restricted to people between the age of 15 to 64, which does not match exactly with the population the IDF used for the diabetes prevalence estimates (20 – 79). Hence, our used sample includes a greater share of young people with a very low diabetes prevalence and excludes people above 64 years of age, which likely have a higher than average prevalence rate. Taken together, this—as well as a further increase in prevalence since 2005—should explain the difference between the diabetes prevalence in our sample and the one estimated by the IDF.

Table 3.1: Summary statistics for males and females with and without diabetes

	Males			Females		
	Mean with diabetes	Mean without diabetes	p (t-test)	Mean with diabetes	Mean without diabetes	p (t-test)
Employed	0.714	0.804	0.000	0.229	0.313	0.000
Age	50.945	35.016	0.000	48.955	34.717	0.000
Age 15–24	0.008	0.294	0.000	0.036	0.282	0.000
Age 25–34	0.043	0.232	0.000	0.076	0.250	0.000
Age 35–44	0.161	0.196	0.162	0.180	0.221	0.042
Age 45–54	0.392	0.166	0.000	0.366	0.159	0.000
Age 55–64	0.396	0.111	0.000	0.342	0.089	0.000
Rural	0.337	0.399	0.047	0.391	0.399	0.723
Small city	0.082	0.126	0.038	0.144	0.123	0.204
City	0.145	0.102	0.028	0.103	0.098	0.737
Big city	0.435	0.372	0.042	0.362	0.379	0.475
Southsoutheast	0.208	0.203	0.864	0.184	0.206	0.270
Central	0.243	0.184	0.017	0.231	0.195	0.062
Westcentral	0.173	0.213	0.124	0.191	0.210	0.343
Northeastcentral	0.196	0.177	0.446	0.209	0.186	0.236
Northwestcentral	0.180	0.223	0.112	0.184	0.202	0.355
No education	0.090	0.062	0.070	0.151	0.081	0.000
Primary	0.518	0.352	0.000	0.607	0.368	0.000
Secondary	0.231	0.308	0.009	0.171	0.314	0.000
Highschool	0.059	0.158	0.000	0.043	0.138	0.000
College or university	0.102	0.120	0.379	0.029	0.098	0.000
Indigenous	0.137	0.121	0.448	0.133	0.118	0.341
Married	0.812	0.535	0.000	0.663	0.539	0.000
Children (under 15)	1.118	1.510	0.000	1.207	1.600	0.000
Wealth	0.179	-0.010	0.003	0.004	-0.003	0.885
Diabetes	1.000	0.000	.	1.000	0.000	.
Diabetes father	0.180	0.071	0.000	0.146	0.079	0.000
Diabetes mother	0.251	0.107	0.000	0.236	0.113	0.000
Education parents	0.596	0.697	0.001	0.528	0.699	0.000
Formal employment	0.286	0.306	0.508	0.083	0.140	0.001
Informal employment	0.529	0.560	0.342	0.191	0.220	0.155
N	255	6031		7798	445	

i has reported a diagnosis of diabetes and 0 otherwise. X_i contains various control variables. Because no information on job history is available in the data to adequately account for work experience, we need to rely on the combination of age and education to proxy for work experience (Aaronson, 2010). The effect of age is captured through dummy variables for age intervals. Education is taken into account by dummy variables indicating if the highest level of schooling attained was either primary school, secondary school, high school, university or some other form of higher education with no education serving as the reference category, to control for the impact of education on employment and to account for the relationship between diabetes and education (Agardh et al., 2011). Since Mexico is a large and diverse country with regional socioeconomic differences we also include dummies for five different Mexican regions[4]. Apart from the more obvious effects economic differences between regions can have on employment chances and diabetes through their impact on employment opportunities and lifestyles, the dummies should

[4]The region variables have been constructed after recommendations on the MxFLS-Homepage. Southsoutheastern Mexico: Oaxaca, Veracruz, Yucatan; Central Mexico: Federal District of Mexico, State of Mexico, Morelos, Puebla; Central northeast Mexico: Coahuila, Durango, Nuevo Leon; Central western Mexico: Guanajuato, Jalisco, Michoacan; Northwest Mexico: Baja California Sur, Sinaloa, Sonora.

also account for less obvious effects that macroeconomic problems, such as a high unemployment rate, could have on employment chances and diabetes by affecting psychological well-being and sleeping patterns (Antillón et al., 2014). Because differences in economic opportunities and lifestyles should also be expected between rural and urban areas, three dummy variables are included to capture the effects these factors might have on employment chances and diabetes, with living in a rural area being the reference category[5] (Villalpando et al., 2010). Further, to control for labour market discrimination and possible differences in genetic susceptibility to diabetes of indigenous populations (Yu and Zinman, 2007), a dummy for being a member of an indigenous group is included. We also account for for the marital status to control for the impact of marriage on employment chances and lifestyle habits. Further a variable capturing the number of children residing in the household below the age of 15 is inlcuded, to control for their impact on employment chances and for the effect of childbearing and related gestational diabetes on the probabilities of women to develop type 2 diabetes (Bellamy et al., 2009). To account for the effect that household wealth might have on diabetes and employment chances, we use the well established method of principal component analysis of multiple indicators of household assets and housing conditions to create an indicator for household wealth (Filmer and Pritchett, 2001). Our composite wealth index consists of owning a vehicle, owning a house or other real estate, owning another house, owning a washing machine, dryer, stove, refrigerator or furniture, owning any electric appliances, owning any domestic appliances, owning a bicycle and owning farm animals. It further accounts for the physical condition of the house, proxied by the floor material of the house, and the type of water access.

The error term is denoted as u_i. We do not control for the general health status and other diabetes related chronic diseases as they are likely determined by diabetes itself and, hence, could bias the estimates and compromise a causal interpretation of the effect of diabetes on employment (Angrist and Pischke, 2008).

As diabetes could be endogenous, the probit model might deliver biased estimates. Therefore we employ an IV strategy, using a bivariate probit model to estimate the following two equations simultaneously:

$$Diabetes_i = \delta_0 + \delta_1 X_i + \delta_2 diabetesmother_i + \delta_3 diabetesfather_i + \eta_i \tag{3.2}$$

$$Employed_i = \beta_0 + \beta_1 Diabetes_i + \beta_2 X_i + u_i \tag{3.3}$$

In equation 3.2, $Diabetes_i$ is a dummy variable and is modelled as a function of the same socioeconomic and demographic factors X_i as in equation 3.1 and of the instrumental dummy variables $diabetesmother_i$ and $diabetesfather_i$, indicating if the father or the mother had been diagnosed with diabetes. The error term is denoted as η_i. Equation 4.2 is identical to the probit

[5]Rural: < 2,500 inhabitants; Small city: 2,500 to 15,000 inhabitants; City: 15,000 to 100,000 inhabitants; Big city: > 100,000 inhabitants.

specification (equation 3.1) and estimates the effect of diabetes on employment, now taking into account the possible endogeneity of diabetes. Diabetes is exogenous if the error terms of both equations are independent of each other ($Cov(u_i\eta_i) = 0$). Endogeneity is tested using a likelihood ratio test based on the idea that if $Cov(u_i\eta_i) = 0$, the log-likelihood for the bivariate probit will be equal to the sum of the log-likelihoods from the two univariate probit models (Knapp and Seaks, 1998). If u_i and η_i are correlated, the estimation of equation 3.1 using a probit model will not provide consistent estimates of the impact of diabetes on employment. In this case the simultaneous estimation of both equations using the bivariate probit should be preferred. For the estimation of the bivariate probit model it is assumed that u_i and η_i are distributed randomly and bivariate normal. To test the assumption of normality, we use Murphey's goodness-of-fit score test with the null-hypothesis of bivariate normally distributed errors, as suggested by Chiburis et al. (2012).[6]

We choose the bivariate probit model over the linear IV model to account for endogeneity, as there is evidence that it performs better if the sample is relatively small (<5,000) and—more important in our case—when treatment probabilities are low. In such cases the linear IV can produce uninformative estimates while the bivariate probit model has been shown to provide much more reasonable results (Chiburis et al., 2012). Because only 4 percent of males and 5.4 percent of females report a diagnosis of diabetes, treatment probabilities are indeed low in the given case, providing good justification for the use of the bivariate probit model.

In order to fulfil the conditions of a valid instrument, parental diabetes needs to impact the diabetes risk of the offspring while at the same time being unrelated to the offspring's employment chances. It has been shown that there is a strong hereditary component of type 2 diabetes which predisposes the offspring of people with diabetes to develop the condition as well (Herder and Roden, 2011; The Interact Consortium, 2013). This is supported by the notion that genes seem to play a crucial role, besides the recent epidemiological transition and the migration from rural to urban areas, in explaining Mexico's high diabetes prevalence according to a recent study by Williams et al. (2014). The authors identified a specific gene particularly prevalent in Mexican and other Latin American populations with native American ancestry, which is associated with a 20 percent increase in the risk of developing type 2 diabetes. Furthermore, research has shown that parental lifestyle factors, socioeconomic background as well as parental BMI can explain but a very small fraction of the increased risk of type 2 diabetes in the offspring, which is why we assume that the increased risk is mainly due to genetic factors unrelated to lifestyle (Herder and Roden, 2011; The Interact Consortium, 2013). This is supported by Hemminki et al. (2010), who find that adoptees whose biological parents had type 2 diabetes, had an increased risk of developing type 2 diabetes even though they were living in a different household, while if their

[6] Murphey's score test "... embeds the bivariate normal distribution within a larger family of distributions by adding more parameters to the model and checks whether the additional parameters are all zeros using the score for the additional parameters at the bivariate probit estimate." (Chiburis et al., 2012, p. 19).

adopted parents had the disease, they had no elevated risk.

Nonetheless, there might still be the chance that parental diabetes decreases the offspring's employment chances. The additional financial burden of diabetes or an early death due to diabetes could have prevented the parents from investing in their children's education the way they would have liked to or it could have led to the child dropping out of school in order to support the family. However, controlling for education should account for these effects if they exist. Therefore parental diabetes should be a valid instrument which predicts diabetes while not affecting employment probabilities through other unobserved pathways. To further improve instrument validity we also account for the possibility that parental education is simultaneously correlated with the parental diabetes status as well as their children's employment chances, by including a dummy variable indicating if any of the parents had attained more than primary education.

A possible limitation of using parental diabetes as our instrument is that it might directly affect the offspring's employment decision through other pathways than education. Conceivably, diabetes might deteriorate parental health in such a way that the offspring has or had to give up its own employment in order to care for its parents or is forced to take up work to financially provide for the parents. With the data at hand we are unable to account for this, but if this effect exists it should be picked up by the overidentification test.

We also estimate the linear probability model (LPM) and the linear IV model as they are consistent even under non-normality (Angrist and Pischke, 2008). The linear IV model takes the following form of a first (Equation 3.4) and a second stage (Equation 3.5).

$$Diabetes_i = \pi_0 + \pi_1 X_i + \pi_2 diabetesmother_i + \pi_3 diabetesfather_i + \eta_i \tag{3.4}$$

$$Employed_i = \beta_0 + \beta_1 Diabetes_i + \beta_2 X_i + u_i \tag{3.5}$$

In the second stage, the potentially endogenous actual diabetes values are replaced with the predicted values from the first stage. The covariates are the same as in the bivariate probit case described in equations 3.2 and 4.2. In the linear IV model the Hausman test is used to identify endogeneity. Validity of the instruments is tested using first stage diagnostics of the linear IV model, as similar tests are not available for the bivariate probit model. The results of the LPM are available on request as they do not differ meaningfully from the presented probit estimates.

3.3 Results

This section presents the estimation results using 1) a probit model model that assumes diabetes to be exogenous and 2) IV models with parental diabetes as an instrument for diabetes, to determine if diabetes is endogenous or if instead the results from the probit model can be used.

3.3.1 Probit results

Table 3.2: Impact of diabetes on employment probabilities (probit)

	(1) Males		(2) Females	
Age 25–34	0.124***	(.011)	0.121***	(.017)
Age 35–44	0.133***	(.012)	0.232***	(.018)
Age 45–54	0.085***	(.014)	0.170***	(.022)
Age 55–64	−.034	(.020)	0.039	(.026)
Small city	−.013	(.017)	0.043**	(.020)
City	−.036*	(.019)	0.042**	(.021)
Big city	0.029**	(.013)	0.101***	(.014)
Central	0.027	(.015)	−.032*	(.018)
Westcentral	0.020	(.015)	−.008	(.018)
Northeastcentral	0.003	(.016)	−.053***	(.017)
Northwestcentral	−.037**	(.016)	−.100***	(.016)
Primary	0.056***	(.020)	−.006	(.022)
Secondary	0.051**	(.021)	0.058**	(.025)
Highschool	0.040*	(.023)	0.126***	(.029)
College or university	0.047**	(.023)	0.297***	(.033)
Indigenous	0.005	(.016)	−.005	(.020)
Married	0.092***	(.012)	−.231***	(.012)
Children (under 15)	0.010**	(.004)	−.018***	(.004)
Wealth	0.002	(.006)	0.037***	(.007)
Education parents	−.007	(.013)	0.000	(.013)
Diabetes	−.100***	(.029)	−.045*	(.023)
Log likelihood	−2897.807		−4508.573	
N	6286		8243	

Marginal effects; Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2 indicates that the effect of diabetes is negative for both sexes. For males, it reduces the probability of being employed by 10 percentage points (p<0.01).

For females, the effect is also negative but smaller, and shows a reduction in employment probabilities of about 4.5 percentage points (p<0.1).

The other covariates largely show the expected relationships. Employability increases with age and is highest for the 35–44 years age group. Especially for women, living in a more urban environment increases employment chances compared to women living in rural areas. Also, women seem to benefit substantially from higher education in terms of employment chances. For men the effects of education are also positive, though, not as marked as for women. Perhaps surprisingly, being part of an indigenous population does not affect employment probabilities, neither for males or females.

The probit results suggest a significant negative effect of diabetes on the employment probabilities of males and likely also females in Mexico. In light of the concern that diabetes could be endogenous the following section presents the results of the IV estimations.

3.3.2 IV results

Using the bivariate probit model, the diabetes coefficient for males increases in size and remains negative whereas for females it decreases but also remains negative. However, standard errors increase in both models and the results turn insignificant, suggesting considerable loss of efficiency (see Table 3.3). The likelihood-ratio test does not reject the null hypothesis of no correlation between the disturbance terms of equations 3.2 and 4.2 for males and females, suggesting exogeneity of diabetes. The test for normality of the error term does not reject the null hypothesis of normality for the male and the female model, increasing our confidence in the estimates. Nonetheless we also consider the results of the linear IV model: the test statistics indicate sufficiently strong and valid instruments, as shown by the Kleibergen-Paap Wald F statistic for weak instruments of 20.48 for men and 27.71 for women, being above the critical value of 19.93 for ten percent IV size and well above the rule of thumb of 10 for weak identification not to be considered a problem (Baum et al., 2007; Staiger and Stock, 1997). The Sargan test does not reject the null hypothesis of instruments uncorrelated with the error term and instruments correctly excluded from the estimated equation. The coefficients of the linear IV model are very different from the bivariate probit model, turning positive for males and females, but also very imprecise as indicated by the large standard errors (see Table 3.4 displaying the main results and Table 3.8 in the appendix presenting the complete first and second stage estimates). As mentioned before, Chiburis et al. (2012) show that the estimates of the linear IV model are likely to be imprecise when low treatment probabilities exist and can differ substantially from the bivariate probit model, which seems to be the case here.[7] Since the linear IV models fail to reject exogeneity of diabetes as well, we are confident that the standard probit model provides unbiased and efficient estimates of the effect of diabetes on employment chances in Mexico and should therefore be used for inference.

[7] It could also be the case that the difference in estimates is due to the fact that while the bivariate probit model estimates the average treatment effect (ATE) of the variable of interest for the whole sample, the linear IV model estimates the local average treatment effect (LATE), which estimates the effect of diabetes on employment only for those that have diabetes and whose parents have or have had diabetes as well. Therefore, the estimates of both models can be different (Angrist and Pischke, 2008; Chiburis et al., 2012).

Table 3.3: Impact of diabetes on employment probabilities (bivariate probit)

	(1) Males		(2) Females	
Age 25–34	0.125***	(.012)	0.109***	(.015)
Age 35–44	0.134***	(.012)	0.207***	(.016)
Age 45–54	0.089***	(.016)	0.149***	(.021)
Age 55–64	−.025	(.025)	0.032	(.029)
Small city	−.014	(.017)	0.039**	(.018)
City	−.035**	(.018)	0.038**	(.019)
Big city	0.030**	(.013)	0.093***	(.013)
Central	0.027	(.018)	−.030*	(.015)
Westcentral	0.019	(.018)	−.007	(.016)
Northeastcentral	0.002	(.018)	−.049***	(.017)
Northwestcentral	−.038**	(.017)	−.091***	(.015)
Primary	0.057***	(.020)	−.006	(.021)
Secondary	0.052**	(.023)	0.052**	(.022)
Highschool	0.040	(.025)	0.113***	(.027)
College or university	0.046*	(.025)	0.273***	(.032)
Indigenous	0.006	(.017)	−.005	(.016)
Married	0.093***	(.012)	−.215***	(.011)
Children (under 15)	0.010**	(.004)	−.016***	(.004)
Wealth	0.002	(.006)	0.033***	(.007)
Parental education	−.006	(.013)	0.000	(.012)
Diabetes	−.185	(.143)	−.021	(.108)
Instruments				
Diabetes father	0.048***	(.011)	0.041***	(.010)
Diabetes mother	0.037***	(.008)	0.054***	(.008)
Log likelihood	−3737.766		−5939.588	
Score goodness-of-fit (H0=normality of errors)	12.32		8.85	
p value	0.196		0.451	
Endogeneity (H0: Diabetes exogeneous)	0.443		0.039	
p value	0.506		0.844	
N	6286		8243	

Marginal effects; Robust standard errors in parentheses.

The presented coefficients and standard errors for the instruments result from the estimation of the model specified in Equation II, indicating the effect of parental diabetes on a person's diabetes risk.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The next section investigates the effects of diabetes for two different age groups, 15–44 and 45–64, to explore whether, and if so, how the effect of diabetes on employment chances differs between older and younger people. There might be reason to believe that diabetes has a more adverse effect in older age groups, when those suffering from diabetes are likely to have accumulated more years lived with diabetes, and hence are more likely to develop complications.

Table 3.4: Impact of diabetes on employment probabilities (linear IV)

	(1) Males		(2) Females	
Diabetes	0.098	(.215)	0.239	(.214)
R2	0.067		0.120	
F stat (H0: weak instruments)	20.483		27.706	
Sargan test (H0: valid instruments)	0.862		0.295	
p value	0.353		0.587	
Endogeneity (H0: Diabetes exogenous)	0.864		1.796	
p value	0.353		0.180	
N	6286		8243	

Robust standard errors in parentheses.

Instruments: diabetes of mother, diabetes of father.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

Critical values for weak identification test F statistic: 10 percent maximal IV size 19.93, 15 percent maximal IV size 11.59, 20 percent maximal IV size 8.75, 25 percent maximal IV size 7.25.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.3.3 Differences by age groups

When divided into an older and younger age group using the cut-off point of 45 years,the negative effect of diabetes is mainly found in the older age group, for males and females alike (see Table 3.5), where 12.5 percent report having diabetes, compared to only 1.7 percent in the younger age group. The probability of being employed is reduced by about 10 percentage points for men between 45 and 64 years at the one percent significance level, while there is no significant effect on younger men. For women, the employment probability is reduced by about 6 percentage points, with the effect being significant at the five percent level. Similar to men, there is no effect of diabetes on younger women. To investigate in more detail for which age group the effect is strongest, we run separate regressions for both age groups above 44 years. The results (Table 3.9 in the appendix) show that for men the strongest effect appears in the oldest age group (i.e. 55–64 years), where employment chances are reduced by almost 13 percentage points. For females, a significant effect is found solely for those between 45 and 54 years, where employment chances are reduced by 7.6 percentage points. Hence, there appear to be relevant differences between males and females in the age at which the biggest adverse effect of diabetes on employment chances occurs.

Table 3.5: Impact of diabetes on employment probabilities by age group (probit)

	15-44		45-64	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	$-.009$	$-.004$	$-.110^{***}$	$-.057^{**}$
	(.062)	(.042)	(.034)	(.025)
Log likelihood	-1987.285	-3354.003	-925.409	-1167.491
N	4415	5997	1871	2246

Marginal effects; Robust standard errors in parentheses.
For the younger age group, the model contains the age categories 25–34 and 35–44 with 15–24 as the reference category. For the older age group, the model contains the age category 55–64 with 45–54 as the reference category.
Other control variables: region, urban, education, indigenous, marital status, children, wealth, parental education.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The use of IV methods in the age stratified samples is compromised due to a reduction in instrument power, sample size and particularly treatment probabilities. Especially for the younger age group, where treatment probabilities are close to zero, a meaningful interpretation of the IV results is difficult. Further, because no endogeneity was found in the pooled samples for males and females presented in section 3.3.2, we would not expect endogeneity of diabetes in the age stratified samples. We nonetheless test for the possibility of diabetes being endogenous using the bivariate probit model and an approach suggested by Lewbel (2012), to improve instrument strength. The results and interpretation of this analysis are available in the appendix (Section D) and support our reliance on the standard probit estimates for inference (see Table 3.11 and Table 3.12).

3.3.4 Differences by wealth

To explore the heterogeneity of the effect of diabetes on employment across different levels of wealth, we divide the sample into two wealth groups at the 50^{th} percentile of our constructed wealth index.

We run separate regressions for both groups stratified by gender, finding the strongest negative effect for less wealthy males, where employment chances are reduced by 15 percentage points, and a smaller and less significant effect for less wealthy females (see Table 3.6). Whereas the coefficients for wealthier males and females have a negative sign, they are not significant at the ten percent significance level. This indicates that mainly the less wealthy experience an adverse effect from diabetes. To further explore this, we stratified the sample into wealth quartiles (see Table 3.10 in the appendix), finding that significant adverse effects for males appear in the first and second wealth quartile, where employment chances are reduced by about 14 percentage points. For females a highly significant and strong effect is only found in the poorest quartile, were employment chances are reduced by 10 percentage points. Together these results indicate that the impact of diabetes on employment chances varies with wealth, with men and women being more affected when being in the lower wealth quartiles.

To consider the possible endogeneity of diabetes in the upper and lower wealth half, we again present the results of the IV models. The stratification into wealth groups significantly reduces instrument power as well as sample size. For none of the wealth groups the bivariate probit model indicates endogeneity (see Table 3.13 in section E of the appendix*)*. This does not change even when using the Lewbel approach to increase instrument strength and we therefore rely on the probit results for inference.

Table 3.6: Impact of diabetes on employment probabilities by wealth group (probit)

	Poor		Rich	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	$-.150^{***}$	$-.047^{*}$	$-.060$	$-.038$
	(.047)	(.027)	(.038)	(.035)
Log likelihood	−1459.235	−2040.517	−1408.746	−2421.910
N	3140	4091	3106	4117

Marginal effects; Robust standard errors in parentheses.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.3.5 Differences by employment type

To investigate the effect of diabetes on the employment chances in the formal and informal labour market, respectively, we estimate separate models with being employed in the formal and informal sector as the respective dependent variables. We define formal employment on the basis of having a written labour contract. Informal employment is defined as working without a written contract or being self-employed.

For this investigation we use two restricted samples: for the estimation of the effect of diabetes on informal employment we exclude those currently in formal employment and for the effect of diabetes on formal employment we exclude those in informal employment from our sample. We further assume that those who have worked previously and are currently unemployed are looking for employment in the same sector, i.e. if they were previously employed in the informal (formal) labour market they are again looking for an informal (formal) employment. We therefore exclude those previously working in the informal (formal) labour market from our estimation of the effect of diabetes on employment in the formal (informal) labour market. The respective sample thus only contains those currently working in the informal (formal) labour market, those previously employed in the informal (formal) labour market and those that have never worked before. Using this assumption allows the use of a normal probit model and the investigation of a possible endogeneity bias using IV techniques.

Admittedly, the assumption that the currently unemployed look for work in the same labour market they had previously worked in is quite strong and is likely not true for everybody. We therefore additionally estimate a multinomial logit model which is most useful if the decision to work is not binary but there are more than two choices, such as the choice of being either unemployed, employed in the informal or employed in the formal labour market (Wooldridge,

Table 3.7: Impact of diabetes on employment probabilities by employment status (probit)

	Males		Females	
	(1) Informal	(2) Formal	(3) Informal	(4) Formal
Diabetes	$-.063^{**}$	$-.041$	$-.051^{**}$	0.019
	(.031)	(.043)	(.022)	(.022)
Log likelihood	-1780.023	-1021.771	-3818.588	-1859.048
N	4604	2204	6983	5652

Marginal effects; Robust standard errors in parentheses

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2002). Being unemployed is used as the reference category.

All estimated models (see Tables 3.7 and 3.15), regardless of the estimation approach, indicate that diabetes significantly reduces the chances of being in informal employment, while it has no effect on formal employment.[8] This applies to both males and females. This indicates that people with diabetes are less likely to be working in the informal labour market relative to being unemployed, while there is no difference for those working in the formal labour market. We further find no indication of endogeneity (see Tables 3.16 and 3.17 in the appendix). Overall, there seem to be strong differences in terms of the impact of diabetes on people in formal and informal employment, with diabetes having a stronger negative effect for those without a written contract.

3.4 Conclusion

The contribution of this paper has been to analyse—for the first time for a LMIC—the impact of diabetes on employment in Mexico, taking into account the potential endogeneity in the relationship between diabetes and employment chances. The presented results add to the growing literature on the adverse economic effects of diabetes. They indicate that having diabetes

[8]Please note, however, that the coefficients of the multinomial logit and the probit model cannot be directly compared as they are based on different assumptions. The former takes into account that a person can choose from more than two employment outcomes (i.e. being unemployed, being formally employed or being informally employed), while the latter only allows for a binary outcome without considering any other options (e.g. being unemployed or informally employed without considering the possibility of formal employment).

substantially reduces the chances to work for men and likely also for women. Hence, diabetes may contribute to a reduction in the pool of the productive workforce available to the Mexican economy.

We have also shown that diabetes reduces employment chances particularly in older people, likely because in this age group people are more common to already have developed diabetes-related complications which reduce their productivity and eventually force them into unemployment. Further, particularly for men the effects of diabetes on employment chances seem to be particularly strong when they belong to the poorer half of the population. While there might be some self-selection into the poorer group by those who lost their job due to diabetes and as a result descended into the lower wealth group, this finding is indicative of potentially substantial adverse equity impacts. This is also in line with our finding that diabetes reduces employment chances particularly for the informally employed, whereas those in formal employment seem to be less affected. Nonetheless, in order to establish causality more research in this area will be needed.

While in parts of the earlier literature diabetes was found to be exogenous only for either males or females (Brown, Pagán, et al., 2005; Latif, 2009b), our study found diabetes to be exogenous using the samples stratified into males and females, allowing the use of the more efficient probit model to arrive at a consistent estimate of the effect of diabetes on employment chances. Further, we found no endogeneity of diabetes for the sample comprised of the age group above the age of 44, for the samples stratified into an upper and lower wealth half and for the samples stratified by employment type. For the younger age group the bivariate probit model only indicated exogeneity of diabetes for males, while for females diabetes was shown to be endogenous and showing a significant positive effect of diabetes on employment. This result is rather counterintuitive because there is no obvious reason why diabetes should increase employment chances. Because all samples stratified into age,wealth and employment groups suffered from reduced instrument strength which could cause biased IV estimates, we used a method proposed by Lewbel (2012) to create additional instruments and increase instrument power. Using this method we no longer found a significant positive effect of diabetes on female employment chances in the younger age group and could not reject the assumption of exogeneity of diabetes in this sample. Also, for all other wealth, age and employment samples, the Lewbel IV method did not reject the assumption of exogeneity. We are therefore confident that we can rely on the probit estimates for inference.

Why was diabetes found to be exclusively exogenous in the Mexican case? We can only speculate on the potential reasons. Diabetes being exogenous seems to indicate that a person's employment status might not have such a strong effect on his or her diabetes risk through the potential pathways such as lifestyle changes. Rather, the rapid epidemiological transition experienced in Mexico over the last decades (Barquera, Hotz, et al., 2006; Barquera, Hernandez-Barrera, et al., 2008; Rivera, Barquera, Campirano, et al., 2002) together with the heightened genetic susceptibility of Mexicans to diabetes (Williams et al., 2014), seem to have increased the risk of developing diabetes in both employed and unemployed Mexicans.

Taking our results for the older age group and comparing them to those of Brown, Pagán, et al. (2005) for the USA, whose sample of Mexican Americans 45 years and older might be the best suited for a meaningful comparison, our findings indicate a stronger negative impact of diabetes on males and particularly females residing in Mexico.[9] This finding lends some support to our hypothesis that the adverse impact of diabetes on employment could be larger in LMICs than in high-income countries. Comparing the study to Lin (2011) for Taiwan, who also uses a sample of people between 45 and 64 years of age, our results are similar in that a larger effect is found for males than for females. We found a somewhat stronger effect for females while the effect for males was lower in our study. However, when compared to other studies in more developed countries, with more advanced health systems and very different populations, such as Latif (2009b) for Canada and Minor (2011b) for women in the US, our results differ in that they do not indicate very strong effects for women.

It is difficult to say precisely what might cause these differences. Potentially, they are related to the differences in the physical demands placed on males and females in their respective jobs. Men in Mexico might need to rely more on their physical fitness to perform well in their jobs than women, causing men to drop out of the labour market earlier due to diabetes complications. Due to the large informal and physically demanding labour market in Mexico compared to Canada or the US, men in Mexico possibly experience a greater reduction in their employment chances due to diabetes than men in higher-income countries. Further, the larger impact diabetes has on males in the poor to middle wealth quartiles and the informal sector could indicate that employers more rapidly replace workers with diabetes with healthy workers, especially if jobs are not particularly specialized or lack regulatory protection and other workers with a similar skill set can be easily found, which is likely the case in Mexico. Higher skilled male workers residing in the richer wealth quartile or in the formal sector might be able to prevent losing their job because of diabetes due to physically less demanding jobs, a more unique skill set which is harder to replace and possibly stronger regulatory job protection. The same seems to be true for women. In higher-income countries jobs are likely more similar between men and women and generally less physical demanding so that physical attributes are not as important and diabetes might not limit men to a greater extent than women. In these countries the stronger impact of diabetes on female employment chances might be explained by more severe health consequences of diabetes for women compared to men (Huxley, 2006). Nonetheless, explaining these differences remains speculative and more research is needed to investigate this.

A limitation of this study is the use of cross-sectional data, which does not allow for the use of fixed effects and hence for the control of unobserved time-invariant heterogeneity. Data

[9] This is based on comparing our estimates to the appropriate models in Brown, Pagán, et al. (2005) based on their test for endogeneity, which indicates the use of the bivariate probit results for women and the probit results for men.

spanning a longer time period of 10 to 15 years would be required to be able to observe changes in the diabetes and employment status which would allow the use of fixed effects. A further limitation is the somewhat old data from 2005, which precedes the main implementation period of the public health insurance scheme called Seguro Popular. This should be taken into account when interpreting our results as the effects might be different today, where most Mexicans have access to some sort of health insurance (Knaul et al., 2012). The presented results rather show the effects of diabetes on employment chances in 2005 in an environment were insufficient healthcare coverage was common for parts of the Mexican population. Further, the data only provided self-reported information on diabetes, which might have caused some attenuation bias in our estimated parameters, making them rather conservative (Lewbel, 2007). We nonetheless deliberately chose this particular data as it provided us with a sensible instrument in parental diabetes as well as an array of other socioeconomic information which—as far as we have been able to ascertain—is not provided by any other dataset in LMICs. Finally, due to data limitations, we were not able to investigate the relationship between diabetes duration and employment chances and how long it takes for an employment penalty to develop. Recent research by Minor (2013) on the US has shown that the effect of diabetes on employment chances changes with the duration of diabetes and is strongest in the first five years after diagnosis for males, whereas for females a strong effect appears only about 11–15 years after diagnosis.

Looking ahead, it would evidently be worthwhile to investigate the effects of diabetes on employment in Mexico using more recent data. In light of the recently completed implementation of Seguro Popular—which increased its coverage from about 10 million people in 2005 to over 50 million in 2012 and now provides almost all previously uninsured Mexicans with access to healthcare (Knaul et al., 2012) — the results of this paper might be used as a baseline to judge the success of Seguro Popular in reducing the adverse effects of diabetes on employment. In addition, the reasons for the differences between males and females in the estimated effects remain a matter of speculation and more research is needed to explore the underlying pathways. This information would be valuable in the design of more effective measures to reduce the negative effects of diabetes for both males and females.

In conclusion, this paper shows that diabetes represents a large burden for people in Mexico and likely in other LMICs, not only due to the associated disease and medical cost burden but also because of its effect on employment chances. This is particularly a problem for the poor who are more adversely affected by diabetes than the more affluent. To alleviate some of the negative effects of diabetes Seguro Popular may provide an opportunity to further improve the prevention and treatment of diabetes in the poor, especially if the health system adapts to the challenges presented by chronic diseases (Samb et al., 2010). Evidence of possible cost-effective interventions for secondary prevention in the context of Seguro Popular already exists (Salomon et al., 2012). There remains, however, an evidence gap on cost-effective strategies for the primary prevention of diabetes.

Appendix

3.5 Linear IV estimates (1st and 2nd stage)

Table 3.8: Impact of diabetes on employment probabilities (linear IV, 1st and 2nd stage)

	linear IV male				linear IV female			
	(1) Diabetes		(2) Employed		(3) Diabetes		(4) Employed	
Age 25–34	−.001	(.005)	0.151***	(.015)	0.003	(.005)	0.111***	(.015)
Age 35–44	0.016*	(.009)	0.154***	(.019)	0.032***	(.008)	0.198***	(.017)
Age 45–54	0.081***	(.014)	0.098***	(.028)	0.108***	(.014)	0.122***	(.028)
Age 55–64	0.101***	(.016)	−.052	(.039)	0.198***	(.021)	0.001	(.040)
Small city	0.001	(.010)	−.010	(.019)	−.005	(.011)	0.034**	(.017)
City	0.014	(.014)	−.041**	(.020)	−.009	(.013)	0.032*	(.019)
Big city	0.008	(.008)	0.027*	(.014)	−.004	(.009)	0.093***	(.013)
Central	0.011	(.011)	0.024	(.017)	0.015	(.011)	−.035**	(.017)
Westcentral	−.002	(.010)	0.021	(.017)	−.002	(.010)	−.006	(.018)
Northeastcentral	0.007	(.012)	0.005	(.017)	0.009	(.012)	−.051***	(.017)
Northwestcentral	−.006	(.009)	−.033**	(.017)	0.007	(.011)	−.095***	(.017)
Primary	−.009	(.020)	0.060**	(.027)	0.017	(.018)	−.011	(.019)
Secondary	−.003	(.020)	0.056*	(.030)	−.005	(.018)	0.052**	(.021)
Highschool	−.027	(.020)	0.045	(.031)	−.008	(.020)	0.117***	(.026)
College or university	−.018	(.023)	0.057*	(.032)	−.028	(.020)	0.291***	(.025)
Indigenous	0.009	(.010)	0.005	(.017)	0.012	(.013)	−.006	(.018)
Married	0.015**	(.007)	0.086***	(.012)	−.002	(.007)	−.216***	(.011)
Children (under 15)	−.005**	(.002)	0.010**	(.004)	0.003	(.002)	−.016***	(.004)
Wealth	0.003	(.004)	−.001	(.007)	0.003	(.004)	0.030***	(.006)
Parental education	0.019**	(.009)	−.010	(.013)	0.014	(.009)	−.001	(.011)
Diabetes father	0.068***	(.020)			0.035**	(.014)		
Diabetes mother	0.043***	(.016)			0.055***	(.013)		
Diabetes			0.098	(.215)			0.239	(.214)
Constant	−.015	(.022)	0.607***	(.036)	−.020	(.021)	0.289***	(.027)
R2	0.075		0.067		0.090		0.120	
F stat (H0: weak instruements)			20.483				27.706	
Sargan test (H0: valid instruments)			0.862				0.295	
p value			0.353				0.587	
Endogeneity (H0: Diabetes exogenous)			0.864				1.796	
p value			0.353				0.180	
N	6228		6286		8186		8243	

Robust standard errors in parentheses.

Instruments: diabetes of mother, diabetes of father.

Other control variables: age, region, urban, education, indigenous marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.6 Results for older age groups

Table 3.9: Impact of diabetes on employment probabilities by age groups older than 44 (probit)

	45-54		55-64	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	$-.083^{*}$	$-.076^{**}$	$-.128^{**}$	$-.033$
	(.048)	(.034)	(.056)	(.039)
Log likelihood	-451.544	-764.722	-458.632	-392.174
N	1101	1399	770	847

Marginal effects; Robust standard errors in parentheses.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.7 Results for wealth quartiles

Table 3.10: Impact of diabetes on employment probabilities by wealth quartile (probit)

	1st		2nd		3rd		4th	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females	(7) Males	(8) Females
Diabetes	−.142*	−.101***	−.144**	0.028	−.082	−.026	−.040	−.053
	(.077)	(.029)	(.060)	(.048)	(.053)	(.044)	(.046)	(.048)
Log likelihood	−776.619	−937.144	−672.633	−1092.280	−689.910	−1266.304	−703.495	−1144.588
N	1577	2039	1563	2052	1516	2143	1590	1974

Marginal effects; Robust standard errors in parentheses.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.8 Instrumental variable analysis for age groups

The results of the bivariate probit models do not indicate endogeneity for the older age group and for males in the younger age group (see Tables 3.11 and 3.12), suggesting that particularly for males the results of the more efficient probit model (Table 3.5) show the true effect of diabetes on employment chances. Only for females in the younger age group the test for endogeneity rejects the assumption of exogeneity and the diabetes coefficient—surprisingly—shows a strong positive effect of diabetes on female employment chances. Instrument strength, however, is reduced significantly, which together with the very low treatment probabilities questions the validity of the IV results for the sample of the younger age group, as weak instruments possibly introduce a bias similar to or stronger than the potential bias in the probit estimates (Staiger and Stock, 1997). We therefore additionally apply a method proposed by Lewbel (2012), which uses heteroscedasticity in the estimated models to construct additional instruments. Instruments are generated by multiplying the heteroscedastic residuals from the first-stage regressions with a subset of the included exogenous variables. Lewbel (2012) recommends the use of this method when traditional instruments are not available or if it is suspected that the traditional instrument is too weak for identification, which is the issue at hand. The approach has been widely used over the last years both in health economics (Brown, 2014; Drichoutis et al., 2011; Kelly et al., 2014; Schroeter et al., 2012) and in other economic disciplines (Denny and Oppedisano, 2013; Emran and Shilpi, 2012; Huang et al., 2009). Using this method to construct additional instruments by using our age group dummies, we are able to increase instrument strength significantly in the younger age group and the overidentification test indicates validity of the instruments. The results of the linear IV model with the additional instruments show exogeneity of diabetes for males and females and do not indicate a significant positive effect of diabetes on employment chances.

Apart from the results of the Lewbel approach, we also think that there are theoretical reasons why diabetes is likely exogenous in the younger age group. While we cannot distinguish between the types of diabetes with the data at hand, it is likely that a relatively large proportion of the people reporting diabetes in this age group have type 1 diabetes, which people tend to get at a younger age (Maahs et al., 2010). The disease has a strong genetic component and it is very unlikely that there are unobserved factors that affect the chances to develop type 1 diabetes and being employed at the same time, nor that employment status would affect the development of type 1 diabetes. Therefore, for a large part of the people reporting diabetes in the younger age group, endogeneity should not present a problem because they have type 1 diabetes. Furthermore, it is also less likely that reverse causality is a problem for those having type 2 diabetes in this age group, because any effects of being employed on developing type 2 diabetes take time to develop. It would be reasonable to expect that if being employed affected a person's weight or any other diabetes risk factor, this would happen by changing the person's lifestyle due to changes in income or available leisure time, or by reducing or increasing a person's activity levels at work. Until these changes are expressed in changes in weight or any other risk factor for diabetes and finally cause a development of type 2 diabetes, a considerable time period of various years has likely passed and people have reached an advanced age. We therefore believe, that the risk of diabetes being affected by employment is much lower in the younger age group based on the nature of the disease, compared to the older age group. Hence we think that the assumption of exogeneity of diabetes in the younger age group is valid—which is also supported by the Lewbel estimates — and that the endogeneity indicated for younger females in the bivariate probit model is likely the result of the low prevalence rates, and consequently the very low treatment probabilities, together with weak instruments, making a meaningful IV analysis difficult (Chiburis et al., 2012). We are therefore confident that we can rely on our probit estimates for inference.

Table 3.11: IV estimates for the age group 15–44

	BP		Lewbel IV	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	0.171***	0.496***	0.007	0.051
	(.046)	(.080)	(.053)	(.071)
R2			0.093	0.143
Score goodness-of-fit (H0=normality of errors)	9.56	14.25		
p value	0.387	0.114		
F stat (H0: weak instruments)	4.288[a]	10.835[a]	366.480	65.872
Sargan test (H0: valid instruments)	0.008[a]	0.044[a]	1.817	3.487
p value	0.930[a]	0.834[a]	0.611	0.322
Endogeneity (H0: Diabetes exogenous)	1.422	12.948	1.065	1.429
p value	0.233	0.000	0.302	0.232
N	4415	5997	4415	5997

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.
Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.
The models contain the age categories 25–34 and 35–44 with 15–24 as the reference category.
Other control variables: region, urban, education, indigenous, marital status, children, wealth, parental education.
[a] The test statistics are taken from the linear IV model not presented here.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12: IV estimates for the age group 45–64

	BP		Lewbel IV	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	−.022	−.112	−.178	−.042
	(.138)	(.111)	(.160)	(.104)
R2			0.058	0.118
Score goodness-of-fit (H0=normality of errors)	7.00	11.10		
p value	0.637	0.269		
F stat. (H0: weak instruments)	15.408[a]	18.305[a]	12.534	18.897
Sargan test (H0: valid instruments)	2.717[a]	0.482[a]	4.397	1.688
p value	0.067[a]	0.487[a]	0.111	0.430
Endogeneity (H0: Diabetes exogenous)	0.688	0.574	0.082	0.024
p value	0.407	0.449	0.774	0.876
N	1871	2246	1871	2246

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.

Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.

The models contain the age category 55–64 with 45–54 as the reference category.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

[a] The test statistics are taken from the linear IV model not presented here.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.9 Instrumental variable analysis for wealth groups

To consider the possible endogeneity of diabetes in the upper and lower wealth half, we again present the results of the bivariate probit and the Lewbel model. The stratification into wealth groups significantly reduces instrument power as well as sample size. For none of the wealth groups the bivariate probit model indicates endogeneity (see Table 3.13 and Table 3.14*)*. This does not change even when using the Lewbel approach to increase instrument strength. Accordingly, we do not find any indication of endogeneity of diabetes in the wealth groups and rely on our probit estimates for inference.

Table 3.13: IV results for lower wealth half

	BP		Lewbel IV	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	−.354	−.064	−.142***	−.054*
	(.241)	(.139)	(.050)	(.032)
R2			0.071	0.099
Score goodness-of-fit (H0=normality of errors)	NA[a]	7.41		
p value	NA[a]	0.594		
F stat (H0: weak instruments)	6.322[b]	15.420[b]	2589.091	1311.647
Sargan test (H0: valid instruments)	0.342[b]	1.106[b]	4.169	2.804
p value	0.558[b]	0.293[b]	0.525	0.730
Endogeneity (H0: Diabetes exogenous)	1.190	0.016	0.005	0.156
p value	0.275	0.901	0.941	0.693
N	3169	4111	3169	4111

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.

Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

[a] The command SCOREGOF failed to produce the test statisitic for this subsample.

[b] The test statistics are taken from the linear IV model not presented here.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.14: IV results for upper wealth half

	BP		Lewbel IV	
	(1) Males	(2) Females	(3) Males	(4) Females
Diabetes	−.142	0.103	−.057	−.000
	(.199)	(.203)	(.037)	(.039)
R2			0.089	0.142
Score goodness-of-fit (H0=normality of errors)	11.40	12.92		
p value	0.249	0.166		
F stat (H0: weak instruments)	14.003[a]	13.215[a]	28673.088	1225.456
Sargan test (H0: valid instruments)	0.848[a]	0.019[a]	10.180	5.787
p value	0.357[a]	0.889[a]	0.070	0.327
Endogeneity (H0: Diabetes exogenous)	0.238	0.730	0.955	1.807
p value	0.626	0.393	0.329	0.179
N	3117	4132	3117	4132

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.

Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

[a] The test statistics are taken from the linear IV model not presented here.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.10 Multinomial logit and IV results for formal and informal employment

Table 3.15: Impact of diabetes on employment probabilities by employment status (multinomial logit)

	Males		Females	
	(1) Informal	(2) Formal	(3) Informal	(4) Formal
Diabetes	$-.073^{**}$	0.031	$-.044^{**}$	0.008
	(.031)	(.026)	(.019)	(.018)
Log likelihood	−4997.064	−4997.064	−6267.941	−6267.941
N	6286	6286	8243	8243

Marginal effects; Robust standard errors in parentheses.

Base category is being unemployed.

Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To consider the possible endogeneity of diabetes when estimating its effect on formal and informal employment, we again present the results of the bivariate probit and the Lewbel model. The stratification into formal and informal employment groups significantly reduces instrument power as well as sample size. For none of the employment groups the bivariate probit model indicates endogeneity (see Table 3.16 and Table 3.17). This does not change even when using the Lewbel approach to increase instrument strength. Accordingly, we do not find any indication of endogeneity of diabetes for the stratification into formal and informal employment and rely on our probit estimates for inference.

Table 3.16: IV results for informal employment

	BP		Lewbel IV	
	(1) Male	(2) Female	(3) Male	(4) Female
Diabetes	−.046	0.069	−.048	−.037
	(.123)	(.130)	(.030)	(.025)
R2			0.103	0.088
Score goodness-of-fit (H0=normality of errors)	13.84	17.37		
p value	0.128	0.043		
F stat (H0: weak instruments)	13.565[a]	25.123[a]	5349.118	2536.362
Sargan test (H0: valid instruments)	0.551[a]	1.684[a]	4.067	4.063
p value	0.458[a]	0.194[a]	0.540	0.540
Endogeneity (H0: Diabetes exogenous)	0.025	1.152	1.128	0.722
p value	0.873	0.283	0.288	0.395
N	4604	6983	4604	6983

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.
Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.
Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.
[a] The test statistics are taken from the linear IV model not presented here.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.17: IV results for formal employment

	BP		Lewbel IV	
	(1) Male	(2) Female	(3) Male	(4) Female
Diabetes	0.098	−.103	−.022	0.003
	(.195)	(.069)	(.049)	(.021)
R2			0.256	0.262
Score goodness-of-fit (H0=normality of errors)	12.95	8.03		
p value	0.165	0.531		
F stat (H0: weak instruments)	8.518[a]	19.996[a]	2764.273	1647.887
Sargan test (H0: valid instruments)	1.111[a]	1.075[a]	9.286	6.741
p value	0.292[a]	0.300[a]	0.098	0.241
Endogeneity (H0: Diabetes exogenous)	0.516	1.833	1.602	0.318
p value	0.473	0.176	0.206	0.573
N	2204	5652	2204	5652

Marginal effects for bivariate probit (BP); robust standard errors in parentheses.
Instruments: diabetes of mother, diabetes of father; for Lewbel additionally created age groups instruments.
Other control variables: age, region, urban, education, indigenous, marital status, children, wealth, parental education.
[a] The test statistics are taken from the linear IV model not presented here.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Acknowledgement

We are grateful to the participants in the 2013 Economics of Disease conference (Darmstadt, Germany), participants in the Health Economics Group Seminar at the University of East Anglia and Marcello Morciano, Veronica Toffolutti, Pieter Serneels, Ruth Hancock and three anonymous referees for helpful comments.

4 The Impact of Diabetes on Labour Market Outcomes in Mexico: a Panel Data and Biomarker Analysis

Abstract

There is limited evidence on the labor market impact of diabetes, and existing evidence tends to be weakly identified. Making use of Mexican panel data to estimate individual fixed effects models, we find evidence for adverse effects of self-reported diabetes on employment probabilities, but not on wages or hours worked. Complementary biomarker information for a cross section indicates that a large diabetes population is unaware of the disease. When accounting for this, the negative relationship of self-reported diabetes with employment remains, but does not extend to those unaware of their diabetes. Further analysis suggests that this difference stems from worse general health among the self-reports rather than more severe diabetes.

4.1 Introduction

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

The observed trend has been attributed to a deterioration in diet and a reduction in physical activity (Barquera, Hernandez-Barrera, et al., 2008; Basu et al., 2013), while genetic predisposition among Mexicans with pre-hispanic ancestry may also have played a role (Williams et al., 2014). Recent evidence indicates that the onset of diabetes has been occurring at an ever earlier age in Mexico (Villalpando et al., 2010). With treatment as ineffective as it currently is —only a minority achieves adequate blood glucose control (Barquera, Campos-Nonato, et al., 2013)— the earlier onset will increase the likelihood of complications during the productive lifespan.

Diabetes is a term used to describe various conditions characterized by high blood glucose values, with the predominant disease being type 2 diabetes accounting for about 90 percent of all diabetes cases (Sicree et al., 2011). The elevated blood glucose levels that are a result of the body's inability to use insulin properly to maintain blood glucose at normal levels, can entail a range of adverse health effects for the individual concerned. However, via effective self-management of the disease much if not all of the complications can be avoided (Gregg et al., 2012; Lim et al., 2011). In the absence of effective self-management—or in the case of inadequate treatment— diabetes has been documented to lead to conditions such as heart disease and stroke, blindness, kidney problems, and nerve problems which together with impaired wound healing can lead to the loss of limbs (Reynoso-Noverón et al., 2011). These conditions can be seriously debilitating and may therefore reduce an individual's economic activity, including its productivity and labor market participation.

The effect of diabetes on labor market outcomes has been studied predominantly in HICs— with the exception of a study on Mexico (Seuring et al., 2015) and one on China (Liu and Zhu, 2014) each. In the HIC studies diabetes has been found to be associated with reductions in employment probabilities as well as wages and labor supply (Brown, Pagán, et al., 2005; Brown, Perez, et al., 2011; Brown, 2014; Latif, 2009b; Minor, 2011c, 2013; Minor and MacEwan, 2016; Seuring, Archangelidi, et al., 2015).

While these studies have provided useful evidence on the potential labor market effects of diabetes, many of the complexities of the relationship have not been comprehensively addressed in any given study. First of all, unobserved heterogeneity presents a challenge to estimate the relationship between diabetes and labor outcomes. Especially time-invariant unobserved individual characteristics, e.g. health endowments—often related to health during uteru, infant and child years, and to low household income or adverse health shocks during these early years— as well as risk preferences have been shown to adversely affect health in general and the propensity to develop type 2 diabetes more specifically (Ewijk, 2011; Li et al., 2010; Sotomayor, 2013). These and other unobserved personal characteristics (e.g. ability) may also affect employment probabilities, wages or working hours directly through their effects on contemporaneous productivity (Currie and Vogl, 2013) and indirectly by limiting educational attainment and human capital accumulation (Ayyagari et al., 2011). Further, only focusing on the overall effect of a self-reported diabetes diagnosis does not reveal when potential labor market penalties appear, given the dynamic aspect of diabetes and the potential differences in its effects over time. Additionally, apart from its health impact diabetes might also affect labor market outcomes through other channels. For instance, people aware of their condition may be less inclined to continue working if this interferes with their disease management or be suffering from psychological consequences (depression, anxiety) of becoming aware of the disease; they may also use the diagnosis as a justification for decreasing their labor supply, leading to a potential justification bias in the estimated effect of diabetes (Kapteyn et al., 2009). Importantly, for these reasons the labor market effects may also be distinct for people with self-reported versus those unaware of their condition, potentially leading to biased estimates if the analysis is solely based on self-reports.

The objective of this study is to provide new evidence on the impact of diabetes on labor outcomes, while improving upon previous work by paying close attention to the above challenges. We use three waves of panel data from Mexico covering the period 2002–2012, provided by the MxFLS. The MxFLS is particularly useful for the analysis of diabetes as it allows us to account for the above complexities in a more refined way than has been the case so far. Using individual level fixed effects (FE) analysis for the first time in this literature, we take account of time-invariant heterogeneity when assessing the impact of self-reported diabetes and self-reported diabetes duration on labor market outcomes.[1] Further, we add to the current literature in exploring the role of undiagnosed diabetes, using novel and rich biomarker data - an issue of considerable importance in light of the large prevalence of undiagnosed diabetes (see Beagley et al. (2014)) that remained unaccounted for in most earlier studies which typically rely on self-reported information. Doing so sheds light on the issue of measurement error and the potentially differential effects of self-reported and undiagnosed diabetes.

Our results using self-reported diabetes suggest an economically important decrease in the employment probability of people aware of their disease. Wages and working hours, however, do not appear to be negatively associated with self-reported diabetes. We further find that employment probabilities are reduced with each additional year since diagnosis, with some evidence for an even larger effect per year after the initial 10 years.

The biomarker analysis indicates that self-reported diabetes entails a significant employment penalty, while biometrically measured diabetes does not. Overall, undiagnosed diabetes does not appear to affect any of the labor market outcomes examined here, suggesting that adverse effects mainly occur to those self-reporting a diagnosis. We argue that, nonetheless, the effects found for self-reported diabetes in this study are largely unbiased as long as inference is not extended to the unobserved undiagnosed population, and are economically important in light of the sheer size of the diagnosed population in Mexico.

4.2 Diabetes and labor outcomes – existing evidence

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

For the USA, Brown, Pagán, et al. (2005) estimate the impact on employment in 1996–1997 in an elderly population of Mexican Americans living close to the Mexican border, using a bivariate probit model. The study finds diabetes to be endogenous for women but not for men. For the latter, the estimates show a significant adverse effect of 7 percentage points (p.p.). For women, the negative effect becomes insignificant when using IV estimation. In another study, again for a cross-sectional sample of Mexican-Americans, Brown, Perez, et al. (2011) look at how diabetes management, inferred from measured HbA1c levels, is associated with employment chances and wages. The authors detect a linear negative association between HbA1c levels and both employment chances and wages for men.

Two further studies also examine the impact of diabetes on employment and productivity for the USA: Minor (2011c) focuses on the effect of diabetes on female employment, earnings, working hours and lost work days in 2006, finding diabetes to be endogenous and its effect underestimated if exogeneity is assumed. In the IV estimates, diabetes has a significant negative

[1] We are not aware of any other evidence on the effect on wages and working hours in a MIC.

effect on female employment as well as annual earnings but not on working hours. In a later study Minor (2013) investigates the relationship of diabetes duration and labor market outcomes using a cross-sectional analysis, providing evidence of a non-linear relationship, with employment probabilities declining shortly after diagnosis for men and after about 10 years for women; wages are not affected by duration. Finally, a recent study by Minor and MacEwan (2016) investigates the association of self-reported diabetes and undiagnosed diabetes with employment probabilities and working hours in an adult USA population, using cross-sectional data. This study indicates a reduction in the coefficient size of diabetes if undiagnosed diabetes cases are included in the diabetes indicator instead of only self-reported diabetes. Further, they find that there is no association of undiagnosed diabetes with employment probabilities itself. However, the results of the study, particularly those for undiagnosed diabetes, are based on a very small number of cases, warranting further investigation.

For Canada, Latif (2009b) estimate the effect of the disease on employment probabilities using an IV strategy similar to Brown, Pagán, et al. (2005). His results suggest diabetes to be exogenous for females, and both endogenous and overestimated for males in the univariate model, with the estimates of the bivariate model indicating a significant negative impact on the employment probabilities for women, but not for men. For Australia, Zhang et al. (2009b) analyze the effects of diabetes on labor force participation using a multivariate endogenous probit model. Their results demonstrate reduced labor market participation for males and females as a result of diabetes, with the effects appearing overstated if the endogeneity of diabetes is unaccounted for.

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

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

1. They rely exclusively on cross-sectional data, limiting the possibilities to account for unobserved individual characteristics.

2. The use of the family history of diabetes, which has been the sole instrumental variable employed so far, relies on the genetic and heritable component of type 2 diabetes that could theoretically provide valid identification of the true effect of diabetes. However, it remains unclear whether the variable fully satisfies the exclusion restriction, as it may also proxy for other genetically transferred traits, including unobserved abilities that im-

pact labor outcomes directly. This traditional identification strategy also abstracts from intrahousehold or intergenerational labor supply effects (Seuring et al., 2015).[2]

3. The use of self-reported diabetes can introduce non-classical measurement error due to systematic misreporting which has been shown to cause estimates of economic impacts to be potentially biased and overstated (Cawley, Maclean, et al., 2015; O'Neill and Sweetman, 2013; Perks, 2015).

4. A final potential limitation lies in the selection into diagnosis as a result of disease severity: those who are more severely ill are more likely to have visited a medical doctor and be diagnosed.

To overcome some of these limitations, this paper applies an individual level FE panel estimation strategy and makes use of biomarker data. We also estimate models for different types of employment, i.e. non-agricultural wage employment, agricultural employment and self-employment, as ill health may have distinct effects across these activities.

4.3 Data

We use the Mexican Family Life Survey (MxFLS), a nationally representative, longitudinal household survey, which has three waves conducted in 2002, 2005–2006 and 2009–2012. All household members aged 15 and above were interviewed, covering information on a wide range of social, demographic, economic and health characteristics of the individuals and their families (Rubalcava and Teruel, 2013). Apart from self-reported diabetes information that is available in all rounds, we also use information on the self-reported year of diagnosis as well as biomarker data including HbA1c levels for a subsample of respondents. Our main analysis uses all three waves taking advantage of the large amount of observations and the panel structure of the data. Our variable of interest is self-reported diabetes, which is based on the survey question: "Have you ever been diagnosed with diabetes?".

Because the response to this question may well suffer from measurement error due to recall bias, we investigate and try to increase the consistency of the self-reported diabetes variable, using disease information from earlier and ensuing waves to infer on the current, missing or inconsistent, diabetes status (see Appendix 4.7 for further details on our correction procedures). A further, and no less important, source of measurement error is the omission of those with undiagnosed diabetes. In order to investigate how this may affect estimates of the labor market impact of diabetes we use information from a subsample of the 2009-2012 wave containing over 6000 respondents (everybody aged 45+ and a random subsample of those aged 15–44 (Crimmins et al.,

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

Table 4.1: Descriptive statistics for panel and biomarker sample.

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

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

2015)) that have biometrically measured blood glucose values, allowing for the identification of those with undiagnosed diabetes. Throughout our analysis the samples we use are restricted to the working age population (15–64). To prevent pregnant women from biasing our results due to the increased diabetes risk during pregnancy and its effects on female employment status, we have dropped all observations of women reporting to be pregnant at the time of the survey (N=764). We further exclude everybody currently in school.

The detailed information in the MxFLS allows us to consider the following outcome variables of interest: employment[3], hourly wage and weekly working hours.[4] For the pooled data of all three waves (Table 4.1), diabetes was self-reported by 5% of men and 6% of women, respectively. This is consistent with other prevalence estimates of self-reported diabetes for this time period in Mexico.[5] About half of the respondents in the sample live in rural areas. Looking at our outcome variables, 86% of men report some form of employment compared to 37% of women. Interestingly, men do not report considerably higher hourly wages than women but work more hours per week. Also, men are working more often in agricultural jobs while women are more likely to be self-employed or in non-agricultural wage employment. Women also have lower educational attainment on average.

Turning to the biomarker subsample of the third wave (2009-2012), respondents are somewhat older on average than in the pooled sample, as it includes everybody above the age of 44 but only a random subsample of those aged 44 or below (Crimmins et al., 2015). Also, self-reported diabetes is higher than in the pooled sample[6]. Regarding the other control and outcome variables, the sample is fairly similar to the pooled sample. Remarkably, a relatively large share of people have an HbA1c indicative of diabetes, defined by the World Health Organization (WHO) as levels above or equal 6.5% (World Health Organization, 2011)[7]: 18% of males and females are

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

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

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

[6]As well as in the full sample of wave 3.

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

unaware of their diabetes. This suggests that relying on self-reported diabetes as a measure for diabetes in Mexico might considerably understate the true extent of diabetes, potentially leading to biased estimates of its economic impact.

4.4 Estimation strategy

Strauss and Thomas (1998) provide a useful framework to think about the relationship between health and labor outcomes:

$$L = L(H, pc, w(H; S, A, B, I, \alpha, e_w), S, A, B, V, \xi) \tag{4.1}$$

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

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

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

When estimating Eq. 4.1 empirically with observational data, unobserved heterogeneity may bias the results. As mentioned in section 4.1 unobserved factors captured in α such as early childhood investments, innate ability and risk preference could affect wages as well as the probability to develop diabetes. Further, changes in lifestyle due to changes in wages or employment status may also affect the probability to develop diabetes through changes in diet and physical activity. Finally, measurement error e_w may be an important issue due to the large undiagnosed population with diabetes, particularly if being diagnosed is related to employment or wages via better access to healthcare through employment benefits and higher income.

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

in the USA and China.

4.4.1 Panel data on self-reported diabetes

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

We estimate the following model:

$$Y_{it} = \beta_0 + \beta_1 Diabetes_{it} + \beta_2 X_{it} + c_i + \gamma_t + u_{it}. \tag{4.2}$$

where Y_{it} is a binary variable taking a value of 1 if respondent i reports being in employment at time t and 0 otherwise, $Diabetes_{it}$ is a binary variable taking a value of 1 at time t if the respondent reports having ever received a diagnosis of diabetes[10], X_{it} is a vector of control variables, c_i represents an individual fixed effect, γ_t represents a year dummies, and u_{it} is the error term.

For the relationship of self-reported diabetes with wages and working hours our empirical models are estimated conditional on having positive wages and being employed, respectively. In these models Y_{it} represents the log hourly wage of respondent i at time t or the weekly working hours over the last year.

The control variables in both FE specifications include dummy variables to capture the effects of the living environment, of living in a small, medium or large city with rural as the reference category, and state dummies. We also include a marital status dummy and the number of children residing in the household below the age of 6 to control for the impact of marriage and children on labor market outcomes and the effect of childbearing and related gestational diabetes on the probability of developing type 2 diabetes (Bellamy et al., 2009). To account for the effect of changes in household wealth on diabetes and employment probabilities, we use standard principal component analysis of multiple indicators of household assets and housing conditions to create an indicator for household wealth[11] (Filmer and Pritchett, 2001). Finally, a

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

[9] Results are available on request.

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

[11] 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.

quadratic age term and calendar year dummies are included to capture the non-linear effect of age and any trends over time, respectively.

Before moving on, it bears emphasizing that despite our efforts to reduce any bias in our estimates, the estimated coefficients do not reflect true causal effects since time-variant unobserved heterogeneity may still bias the estimates. With respect to employment status, one potential issue would be that job loss affects lifestyle choices that increase the probability to develop diabetes, which could then in turn negatively affect labor market outcomes. So far, no strong adverse effects of job loss as a result of diabetes self-reports have been reported in the literature (Bergemann et al., 2011; Schaller and Stevens, 2015), but this has so far only been researched in a high-income country context. Another example relates to stress at work, which has been linked to the development of type 2 diabetes (Eriksson et al., 2013; Heraclides et al., 2012). However, while stress levels may change over time, a person's coping mechanisms to deal with stress are likely time-invariant (Schneiderman, Ironson, et al., 2005). While we cannot exclude the role of these time variant unobserved factors, it seems that the role of time-invariant variables, e.g. genetic predisposition and relatively stable personality traits, is predominant. The applied FE approach should then limit the bias resulting from these time-invariant confounding factors.

4.4.2 Self-reported diabetes duration

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

$$Y_{it} = \beta_0 + \beta_1 Dyears_{it} + \beta_2 X_{it} + c_i + u_{it}, \tag{4.3}$$

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

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

$$Y_{it} = \delta_0 + g(Dyears_{it}) + \delta_2 X_{it} + c_i + u_{it}. \tag{4.4}$$

with $g(Dyears_{it}) = \sum_{n=1}^{N} \delta_n \cdot max\{Dyears_{it} - \eta_{n-1}\} I_{in}$ and $I_{in} = 1[\eta_{n-1} \leq Dyears_{it} < \eta_n]$, with η_n being the place of the n-th node for $n = 1, 2, \ldots, N$. We choose three nodes that —based on visual inspection (see Figures 4.1, 4.2 and 4.3 in Section 4.5.2)— best captured any possible non-linearity in the relationship between diabetes duration and labor outcomes. These are located at 4, 11 and 20 years after diagnosis. The first four years should capture any immediate effects of the diagnosis, the years five to eleven should capture any effects of adaptation to the disease. After 11 years it is conceivable that many of the debilitating complications of diabetes would appear that could deteriorate health and lead to adverse effects on labor market outcomes. The coefficient δ_n captures the effect of diabetes for the n-th interval. The effects are linear if $\delta_1 = \delta_2 =, \ldots, = \delta_n$.

Because the year of diagnosis was only reported in the third wave, duration of diabetes (or time since diagnosis) for the earlier waves was only calculated for those that had also been interviewed in the third wave, reducing the comparability of the results to those using the binary diabetes indicator.[12]

One caveat of using FE is that, when year dummies are included, any variable that varies by one unit in each time period, is not separately identified (Wooldridge, 2012). Because this is also the case for diabetes duration, in Eq. (4.3) and Eq. (4.4), identification of this variable relies on the presence of people without diabetes in the sample, for which diabetes duration does not increase at the same rate as time.[13] As a further robustness check, we also estimate two models that only use between-individuals variation, i.e. a 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.[14]

4.4.3 Cross-section: biomarker and self-reported data

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

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

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

3. Diagnosis is more likely for those who are more likely to have visited a doctor, for instance because they are more affected by the condition, wealthier, or hypochondriac.[15]

Overreporting may attenuate the effect of diabetes if those falsely reporting a diabetes diagnosis are in fact in good health; it may also lead to overestimation of the impact if some of those misreports reflect other factors that negatively affect labor outcomes (e.g. other illnesses

[12] To obtain the time passed since diagnosis, the year of diagnosis was subtracted from the year of the interview.

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

[14] Models excluding the calendar year dummies provide similar results.

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

or general ill health), or if they are used to justify other adverse events that may negatively affect labor outcomes. Similarly, underreporting may lead to overestimation if those with undiagnosed diabetes are generally healthier, hence more likely to have positive labor market outcomes than those with self-reported diabetes. However, if the undiagnosed and the diagnosed groups are similar in terms of health, then this would lead to an underestimation of the effect of diabetes.

The health information received at a diabetes diagnosis may also have an effect in itself. It may for instance affect an individual's psychology which in turn may influence economic behavior. Two studies found a diabetes diagnosis and subsequent treatment to increase the odds of psychological problems, including depression and anxiety (Paddison et al., 2011; Thoolen et al., 2006), while similar results have not been found for people with undiagnosed diabetes (Nouwen et al., 2011). Looking at behavioral change, health information has been shown to affect behavior after the diagnosis of not only diabetes (Slade, 2012) but also of other chronic diseases (see Baird et al. (2014), Gong (2015), Thornton (2008), and Zhao, Konishi, et al. (2013)). However, little is known about the effects of health information on labor market outcomes. For diabetes, only Liu and Zhu (2014) investigate the effect of receiving a diabetes diagnosis on labor income in Chinese employees. This study finds a reduction in labor income which was attributed to the psychological effects of the diagnosis.[16]

The use of biomarker data allows to explore the relationship of measured diabetes with labor outcomes which can then be compared to the estimated effect of self-reported diabetes. The biomarker data also enables us to look at diabetes severity, as measured by HbA1c values. Since this data is only available for a subsample of one wave —the most recent one— our analysis here is limited to cross-sectional data no longer directly comparable to the panel-based results in this paper. Nonetheless, it allows for a first exploration of the relationships of measured diabetes and disease severity with labor market outcomes.

Our analysis of the biomarker sample consists of three steps. We first estimate Eq. 4.5 to assess the association of self reported diabetes with labor outcomes, as before, but this time for the biomarker sample only, using the following specification:

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 X_i + c_i + u_i \tag{4.5}$$

We then estimate the relations between diabetes, as defined by our biomarker, and labor outcomes, via the following equation:

$$Y_i = \beta_0 + \beta_1 Dbio_i + \beta_2 X_i + c_i + u_i \tag{4.6}$$

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

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

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

$$Y_i = \beta_0 + \beta_1 Dsr_i + \beta_2 Dbio_i + \beta_3 X_i + v_i + u_i. \tag{4.7}$$

For the biomarker analysis we rely on within-household variation v_i for identification to account for unobserved community characteristics, such as the access to healthcare and the quality of healthcare in the community, poverty and unemployment levels in the community or the amount of public green space and recreational possibilities available. These factors potentially affect both the propensity to develop diabetes and to receive a diagnosis; they may also be related to labor market outcomes.[17]

4.5 Results

4.5.1 Incidence of self-reported diabetes

Table 4.2 presents the estimation results of the FE model using Eq. 4.2, which indicate significant and substantial reductions in the probability of employment for men and women with self-reported diabetes. The effects are surprisingly similar across both sexes, showing a reduction in employment probabilities of over 5 p.p..

Table 4.2: Self-reported diabetes and labor market outcomes.

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

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

The results in Columns 3–6 show no significant relationship between self-reported diabetes and wages or working hours. One may expect this relationship to differ by the type of work,

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

as those with diabetes working in an agricultural job that requires strenuous, physical efforts may see their productivity more adversely affected than those engaged in more sedentary work. We therefore estimate a model including interaction terms between self-reported diabetes and agricultural employment and between self-reported diabetes and self-employment, respectively, using non-agricultural wage employment as the comparison group, and restricting our sample to those employed only.

Table 4.3: Effect of self-reported diabetes on wages and working hours, by type of work.

	Log hourly wage		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females
Agricultural worker	$-.078^{*}$	$-.280$	-3.577^{***}	-4.473^{*}
	(.044)	(.186)	(.800)	(2.702)
Self-employed	0.028	$-.144^{*}$	-1.452^{**}	-4.713^{***}
	(.043)	(.087)	(.704)	(1.388)
Self-reported diabetes	0.105	0.064	0.617	$-.524$
	(.076)	(.169)	(1.606)	(2.252)
Self-reported diabetes x agricultural worker	$-.242$	$-.409$	-5.495^{*}	-3.535
	(.188)	(.373)	(2.833)	(22.300)
Self-reported diabetes x self-employed	$-.105$	0.125	0.306	-4.149
	(.192)	(.326)	(2.503)	(4.739)
N	13828	7068	17616	9112

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

The results in Table 4.3 show that while male agricultural workers have lower wages in general, the relationship with diabetes does not depend on the type of work, as none of the interaction terms show up as significant. In the working hours regression one interaction term is significant, suggesting that those with self-reported diabetes working in agriculture supply 5 hours less relative to non-agricultural workers and employees. However, because we have more than two work types we cannot draw conclusions solely on the basis of the t-statistic. We therefore perform a Wald test for the overall significance of the interaction term which does not reject the null of no interaction effects ($p = .15$), indicating that the effect of diabetes on working hours does not vary significantly by type of work.

In summary, we find no evidence for an association between self-reported diabetes and wages or working hours. This lack of effects may be explained by selection: potentially, only those with "mild" or asymptomatic diabetes are still in the same job continuing to earn similar wages. Only once complications become increasingly severe would they switch activity (or drop out of the labor market), without going through a notable phase of reduced productivity and labor supply.

To explore whether diabetes affects the selection into certain types of work we estimate FE models of the probability of being in non-agricultural wage employment, agricultural employment or self-employment using three dummy variables indicating the respective type of work as the left hand side variables. The results in Table 4.4 indicate a negative association with self-employment, though the estimates are quite imprecise. For women, those who self-report diabetes are less likely to work in agriculture and potentially self-employment. This may suggest that having diabetes drives people out of self-employment and agricultural jobs, for instance because these jobs are physically more demanding and possibly also because they provide less protection in terms of insurance and employment duration. We also estimated a pooled multinomial logit model augmented with the within-between approach (Bell and Jones, 2015), based on the work of Mundlak (1978), which allows interpreting the coefficients of all time-varying variables as within-effects by including individual means of all time-varying covariates[18]. The results indicate a very similar pattern both in size and significance (results available on request).[19]

Table 4.4: Relationship between self-reported diabetes and selection into types of work.

	Males			Females		
	(1) Non-agric.	(2) Agric.	(3) Self-employed	(4) Non-agric.	(5) Agric.	(6) Self-employed
Self-reported diabetes	−.006	−.008	−.043	−.001	−.022**	−.029
	(.029)	(.022)	(.026)	(.018)	(.009)	(.018)
N	20719	20719	20719	26577	26577	26577

Individual level fixed effects. Robust standard errors in parentheses. Other control variables: state dummies, urbanization dummies, education dummies, married dummy, number of children < 6, wealth, health insurance status, age squared and calender year dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.5.2 Duration of self-reported diabetes

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

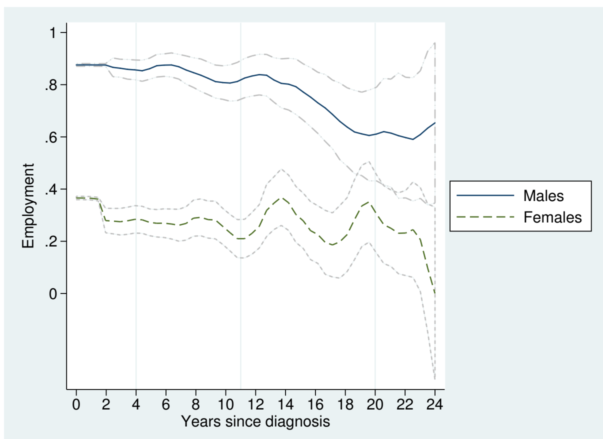
[18]Several other studies in economics have used this approach recently, e.g., Boll et al. (2016), Geishecker and Siedler (2011), and Wunder and Riphahn (2014)

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

period. It is also possible that after some time complications start to appear, again reducing health and leading to reductions in labor supply and productivity.

To obtain an initial idea of the relationship between our outcome variables and diabetes duration we use a non-parametric kernel-weighted local polynomial regression. As Figure 4.1 shows, the relationship between diabetes duration and the probability of employment for men shows a more or less steady decline that becomes more pronounced as time progresses. For women, a first drop-off occurs right after diagnosis; thereafter no consistent pattern is observed.[20] A similar analysis for wages shows somewhat more erratic relationships, although there seems to be a long term negative trend for women but not for men (see figures 4.2 and 4.3). A similar negative trend can be observed for working hours for women, but not for men.

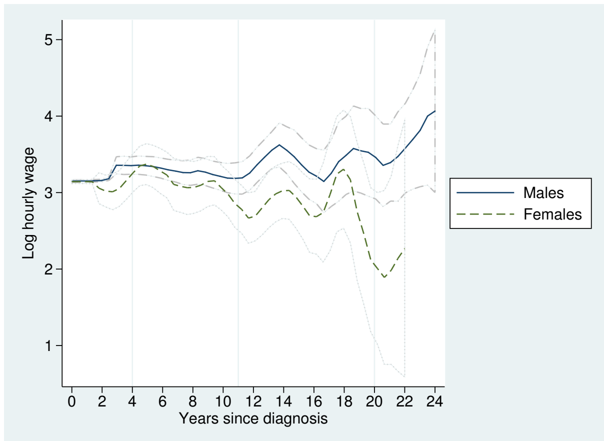
Figure 4.1: Kernel-weighted local polynomial regression of employment status on diabetes duration.



The dotted lines around the main line show 95% confidence intervals.

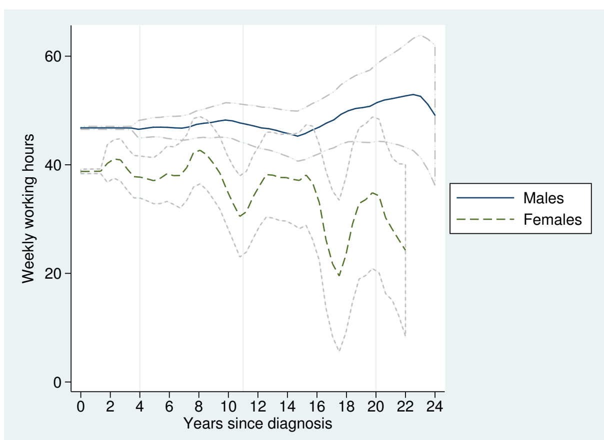
[20]Since long run estimations suffer from large standard errors —as the sample size is strongly reduced— this limits its interpretation and we therefore truncate the graphs at a disease duration of 24 years.

Figure 4.2: Kernel-weighted local polynomial regression of log hourly wages on diabetes duration.



The dotted lines around the main line show 95% confidence intervals.

Figure 4.3: Kernel-weighted local polynomial regression of working hours on diabetes duration.



The dotted lines around the main line show 95% confidence intervals.

Table 4.5 presents the results of the linear and non-linear duration models (for which we created the following splines to capture the immediate, intermediate and long-term relationships:0–4, 5–11, 12–19 and 20+), starting with the results of the cross-sectional LPM, followed by the pooled LPM and then the FE model as specified in Eq. (4.3) and Eq. (4.4).

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

Turning to wages, the FE model indicates a reduction in female wages of about 7% per year with diabetes. For men we find no consistent effect. The results of the non-linear specification indicate that there may be a reduction in wages 5–11 years after the initial diagnosis. We also find associations for women with more than 20 years of diabetes, but these estimates may be spurious due to the considerably reduced number of observations in this group.[21]. There appears to be no consistent relationship between working hours and time since being diagnosed with diabetes.

Overall these results suggest a fairly constant decrease in the probability of employment for both men and women and in earnings for women, which contrast with estimates for the USA (Minor, 2013), where no such linear relationship is observed. Minor (2013) finds a reduction in employment probabilities of 82 p.p. for females after 11 to 15 years and a reduction of 60 p.p. for males after 2-5 years, indicating very large employment penalties, in particular in comparison to our results for Mexico. However, our non-linear results are not directly comparable to these estimates as Minor used pooled cross-sectional data, constructed dummy variables instead of splines and used different duration groups.[22]

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

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

Table 4.5: Relationship between self-reported years since diagnosis and labor market outcomes using continuous duration and duration splines.

	Males			Females		
	(1) OLS (wave 3)	(2) Pooled OLS	(3) FE	(4) OLS (wave 3)	(5) Pooled OLS	(6) FE
			Employment probabilities			
Panel A:						
Diabetes duration (linear)	$-.008^{***}$	$-.007^{***}$	$-.017^{***}$	$-.005^{***}$	$-.004^{***}$	$-.009^{*}$
	(.002)	(.002)	(.006)	(.002)	(.001)	(.005)
Panel B:						
Diabetes duration (splines)						
0–4	$-.007$	$-.007$	$-.026^{*}$	$-.010$	$-.015^{**}$	$-.017$
	(.007)	(.006)	(.014)	(.007)	(.006)	(.016)
5–11	0.000	$-.003$	$-.003$	$-.004$	0.004	$-.003$
	(.009)	(.006)	(.009)	(.008)	(.006)	(.008)
12–20	$-.030^{**}$	$-.017^{*}$	$-.029^{*}$	0.005	$-.004$	$-.014$
	(.012)	(.010)	(.016)	(.008)	(.006)	(.011)
> 20	0.011	0.007	$-.046^{*}$	$-.010^{*}$	$-.003$	$-.015$
	(.016)	(.014)	(.028)	(.006)	(.003)	(.018)
N	8217	16292	16292	10467	22407	22407
			Log hourly wage			
Panel A:						
Diabetes duration (linear)	0.001	0.010^{**}	$-.019$	$-.014^{*}$	$-.009$	$-.073^{**}$
	(.006)	(.005)	(.018)	(.008)	(.008)	(.029)
Panel B:						
Diabetes duration (splines)						
0–4	0.034^{*}	0.046^{***}	0.033	0.027	0.030	0.015
	(.017)	(.016)	(.055)	(.031)	(.026)	(.138)
5–11	$-.041^{*}$	$-.037^{**}$	$-.055^{*}$	$-.039$	$-.034$	$-.101^{*}$
	(.021)	(.018)	(.033)	(.030)	(.024)	(.056)
12–20	0.015	0.044	0.062	$-.032$	$-.071^{*}$	$-.051$
	(.033)	(.029)	(.056)	(.042)	(.039)	(.047)
> 20	0.053	0.014	$-.111$	$-.007$	0.041^{***}	$-.204^{***}$
	(.054)	(.040)	(.104)	(.028)	(.015)	(.053)
N	5509	10767	10767	2874	5741	5741
			Weekly working hours			
Panel A:						
Diabetes duration (linear)	0.069	0.048	0.181	$-.020$	$-.124$	0.208
	(.124)	(.102)	(.330)	(.187)	(.127)	(.652)
Panel B:						
Diabetes duration (splines)						
0–4	$-.033$	$-.233$	0.709	0.739	0.470	2.014
	(.421)	(.325)	(.938)	(.645)	(.586)	(2.947)
5–11	0.269	0.338	$-.218$	$-.410$	$-.479$	$-.508$
	(.539)	(.399)	(.568)	(.728)	(.553)	(1.020)
12–20	0.209	0.137	0.698	$-.164$	$-.051$	$-.402$
	(.730)	(.538)	(.945)	(.995)	(.700)	(1.207)
> 20	-1.300	$-.768$	0.039	$-.499$	$-.418$	8.117^{***}
	(.944)	(.930)	(2.184)	(.930)	(.305)	(1.612)
N	6807	13579	13579	3591	7383	7383

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

4.5.3 Cross-sectional biomarker analysis

In this section we gain additional insights from using the biomarker data collected in the third wave of the MxFLS. As noted in section 4.3, these data enable us to identify respondents with HbA1c levels equal to or above the internationally recognized diabetes threshold of 6.5%. This will allow the investigation of the direction of bias introduced when relying on self-reported diabetes only and when it is not possible to identify those unaware as well.

We first present a cross tabulation of self-reported diabetes and the results of the biomarker analysis (Table 4.6). The table indicates that 27% of the sample have HbA1c levels indicative of diabetes and 81% of those self-reporting a diabetes diagnosis also have HbA1c levels equal to or above the diabetes threshold. Overall, of the people with diabetes according to biomarker analysis, 32% self-report a diagnosis, while 68% do not.

Table 4.6: 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	4544	1181	5725
	79%	21%	100%
	97%	68%	89%
Self-reported diabetes	129	554	683
	19%	81%	100%
	3%	32%	11%
Total	4673	1735	6408
	73%	27%	100%
	100%	100%	100%

The first row of each category presents absolute values, the second row row percentages and the third row column percentages.

To further investigate the relationship of self-reported and biomarker tested diabetes, we estimate the models presented in section 4.4.3. The results in columns 1 and 2 of Table 4.7 show that the earlier results are robust for the biomarker sample. The coefficients in column 3 and 4 indicate that the associations with employment probabilities are much weaker when using diabetes defined by the biomarker instead of self-reported diabetes.[23] In columns 5 and 6, obtained from estimating Eq. 4.7, the coefficient for the biomarker diabetes population $Dbio_i$ now

[23] We also created a dummy variable that additionally to measured diabetes accounted for those with a self-reported diabetes diagnosis but biomaker levels below the diabetes threshold. This allowed us to investigate the effect for the entire diabetes population. The coefficients and their statistical significance are only marginally different to those presented in columns 3 and 4 of Table 4.7, which is why we do not present them here.

reflects the effect of undiagnosed diabetes, as the regression includes a control for self-reported diabetes, revealing that undiagnosed diabetes is not associated with any of the labor outcomes. The coefficient for self-reported diabetes is marginally bigger in size for men and somewhat smaller for women compared to column 1 and 2, respectively. However, these differences are not statistically significant (p>0.1) using a Z-test, suggesting that not accounting for undiagnosed diabetes will likely leave the estimates of self-reported diabetes unbiased.

Table 4.7: Biomarker results

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

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

As discussed earlier, differences in effects between self-reported diabetes and those undiagnosed are likely to stem from selection into the diagnosed population, for instance those in worse health or higher HbA1c levels are more likely to go to the doctor and be diagnosed as well as to lose their job because of their diabetes. To further explore this, we first estimate models additionally controlling for self-reported health status to capture differences in subjective individual health. Secondly, we investigate in how far differences in measured HbA1c, as a proxy for diabetes severity, may explain differences in employment effects of self-reported and undiagnosed diabetes.

To this end we estimate Eq. 4.7 additionally controlling for HbA1c levels.

Table 4.8: Self-reported diabetes, biomarkers, diabetes severity and self-reported health and their association with labor market outcomes

	Employment		Log hourly wages		Weekly working hours	
	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Panel A (self-reported health)						
Self-reported diabetes	−.036	−.023	0.002	0.060	0.123	−2.191
	(.026)	(.027)	(.079)	(.121)	(1.433)	(2.386)
Hba1c $\geq$ 6.5%	0.003	−.023	−.004	−.051	−.066	1.829
	(.017)	(.019)	(.049)	(.075)	(.926)	(1.569)
Self-reported health status						
good	0.023	0.057*	0.061	−.115	−1.131	3.521
	(.025)	(.034)	(.074)	(.124)	(1.376)	(2.499)
fair	−.007	0.006	0.025	−.157	−1.606	4.646*
	(.026)	(.034)	(.076)	(.128)	(1.424)	(2.607)
bad	−.127***	−.024	−.016	−.371*	−6.190**	6.918*
	(.043)	(.046)	(.135)	(.189)	(2.521)	(3.858)
very bad	−.165	0.117	−.331	0.316	−1.869	−17.400*
	(.110)	(.116)	(.300)	(.439)	(6.433)	(9.005)
N	2785	3621	1803	883	2302	1143
Panel B (HbA1c levels)						
Self-reported diabetes	−.056*	−.027	−.007	0.002	0.076	−1.440
	(.031)	(.025)	(.068)	(.114)	(1.362)	(2.382)
HbA1c $\geq$ 6.5%	−.005	−.005	−.010	−.019	1.032	1.887
	(.023)	(.026)	(.060)	(.099)	(1.279)	(2.490)
HbA1c	0.003	−.006	0.001	−.012	−.364	−.122
	(.005)	(.006)	(.013)	(.021)	(.279)	(.514)
N	2785	3623	1803	884	2302	1144

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

When additionally controlling for subjective health status, we find that for men and women the difference between self-reported diabetes and undiagnosed diabetes is reduced due to a smaller coefficient for self-reported diabetes (Table 4.8, Panel A). Especially for females, the point estimates for self-reported diabetes and undiagnosed diabetes are now virtually the same size, suggesting that differences can be almost exclusively explained by self-reported health. For men, factors not captured by self-reported health may still play a role. Additionally accounting for measures of overweight and obesity, self-reported hypertension, heart disease and depression does not further affect the interpretation of the diabetes coefficient (results available on request).

Turning to Panel B, we do not find an indication that differences in HbA1c levels are driving the different employment effects of diabetes for the aware and unaware. We therefore conclude that current diabetes severity is likely not associated with any labor outcome and does not explain the difference in effects between diagnosed and undiagnosed diabetes.

To the best of our knowledge only one study has previously used biomarkers to analyze the relationship with labor market outcomes in a comparable population. Brown, Perez, et al. (2011) use data for a Mexican American population in a broadly comparable way to this paper, though stopping short of investigating the labor market impact of undiagnosed diabetes. In concordance with our results this study also finds that once diabetes is diagnosed, current management plays a minor role in determining labor market outcomes. This is not surprising given that HbA1c levels only provide a picture of blood glucose levels over the last three months. They therefore may not be representative of blood glucose levels in the years before and after the diabetes diagnosis which ultimately determine how soon complications appear and how severe they will be.

4.6 Conclusion

Diabetes has become one of the most common chronic diseases in middle- and high-income countries, with the potential to severely impact the health and economic well-being of those directly (and possibly indirectly) affected. Yet there remains only limited 'hard' evidence on the economic consequences, especially for these countries. Moreover, what evidence does exist at best partially tackles the econometric challenges involved.

This paper improves on existing work by addressing several methodological challenges that arise due to the nature of the disease and types of data available, using rich longitudinal panel data from Mexico, a MIC for which the biomarker data used in this paper indicates that diabetes, including undiagnosed diabetes, has reached alarming levels.

Apart from providing unique evidence for a developing country, the paper makes methodological contributions for the estimation of labor market effects of diabetes. By estimating individual fixed effects the analysis provides an improved accounting for the endogeneity of self-reported diabetes, as this allows canceling out the potential role of unobserved individual traits that may affect both labor market outcomes and propensity to self-report (or suffer from) diabetes. Using further information on the year of diagnosis enables us to investigate the potential heterogeneity in the effect of self-reported diabetes on labor market outcomes over time. Finally, taking advantage of biomarker data to identify the entire diabetes population, i.e. including those with undiagnosed diabetes, allows for an assessment of the potential bias in estimates relying on self-reported diabetes (which is still the most frequent measure in the previous literature).

The first part of our results confirms a considerable gap in employment probabilities for both men and women reporting a diabetes diagnosis, compared to those that do not report the condition. We also find some evidence that diabetes is more likely to reduce the probability of employment in the agricultural and self-employment sector, characterized predominantly by informal arrangements, compared to the rest of the workforce. Those who remain employed do not suffer any wage or labor supply effects, possibly because they are still relatively healthy or are able to resort to a type of work that does not entail their diabetes status limiting their work-related performance. More research will be needed to confirm and further investigate this finding as well as its interpretation.

Regarding the heterogeneity in the effects of diabetes over time, our results indicate an adverse impact of self-reported diabetes on employment chances, with the impact growing in magnitude especially after the first 10 years post-diagnosis. This is plausible in that as time lived with diabetes evolves, complications associated with diabetes tend to become more frequent and more severe (Adler et al., 2003). Looking at wages as our labor market outcome, we uncover some adverse effects for females, indicating a sizable reduction with time since diagnosis. These findings may bode ill for countries were diabetes has started appearing at an increasingly younger age, causing people to live with the disease for larger parts of their productive lifespan, possibly exacerbating the economic effects of reduced employment due to diabetes (Hu, 2011a; Villalpando et al., 2010).

The second part of our results indicates that only relying on self-reported diabetes can lead to an overestimation of the relationship between diabetes and labor market outcomes. We find that a negative relationship only exists for those with self-reported, but not for those with undiagnosed diabetes. This perhaps surprising, notable difference, is at least mediated by the subjective health status being worse for those self-reporting compared to the undiagnosed. Current disease severity, as proxied by HbA1c levels, does not appear to play an important role in this context.

Our findings bear several implications. First, when interpreting labor market impact estimates relying on self-reported diabetes, one cannot assume that the results extend to those with undiagnosed diabetes. However, the strategy of simply merging self-reported and undiagnosed in one diabetes category may not be ideal, as doing so will fail to account for the heterogeneity between the groups in the amount of health information they possess, the time they have already been exposed to elevated blood glucose levels and consequently their subjective as well as true health status, leading to a potentially important loss of information. If, by contrast, both groups are separately accounted for in the model, thereby acknowledging their inherent differences, this allows us to gain information about the distribution of the economic burden across the two groups.

Further, the results of the biomarker analysis also reveal that the coefficient of self-reported diabetes is not strongly affected when accounting for biomarker diagnosed diabetes, suggesting that using self-reported diabetes still provides largely unbiased estimates. The latter estimates should then of course only be used to draw conclusions about the effect of self-reported diabetes, not of diabetes overall. In the case of Mexico, given that more than 7% of the Mexican population have been diagnosed with diabetes, the identified reduction in employment probabilities still amounts to a significant overall economic burden being associated with (diagnosed) diabetes.

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 potential gains to be had in terms of increasing the productive lifespan of people. This is of particular importance in LMICs, 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 deterioration of the household's economic situation. This can lead to forgone investments into child education, showcasing the potential for adverse long-term effects of health shocks due to diabetes (Bratti and Mendola, 2014). Moreover, the large proportion of undiagnosed people indicates that diagnosis —at least in Mexico— happens too late or not at all, thereby significantly reducing the possibility to prevent complications via appropriate treatment and self-management, which has repercussions by increasing the risk of severe complications appearing early. Hence, much of the health and economic burden may be prevented by earlier diagnosis and, given the generally limited success in achieving good control in Mexico, better treatment of those already diagnosed with diabetes. Ultimately of course, there will be a need to invest in the prevention of diabetes cases in the first place. Taxation of sugar sweetened beverages may be one promising way forward (Colchero et al., 2016), though the long-term effects in terms of diabetes prevention remain to be demonstrated.

Appendix

4.7 Strategies to deal with inconsistent self-reporting over time

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

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

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

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

We used this information to infer the "true" diabetes status for those with inconsistent reports. For those with 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 then it was not reported. For people where we had data from all three waves, we used that additional information to make a decision on how to deal with inconsistencies using the rules outlined in Table 4.9

This approach should add more consistency to the self-reported diabetes information by using all available information. We tested if this approach was supported by the HbA1c values provided in wave 3. Of those with inconsistencies in their diabetes elf-reports 95 were present in the

Table 4.9: Inconsistencies in diabetes self-report in MxFLS.

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

biomarker sample (46 with two and 49 with one self-report of diabetes). We therefore Using a t-test we compared the mean HbA1c for the two groups and found a significantly ($p<0.001$) higher mean HbA1c (9.7%) for those with two self-reports compared to for those with only one self-report of diabetes (7.0%). Further, of those with one self-report, for only 30% the HbA1c$\geq$ 6.5% compared to 87% of those with two self-reports. Based on these results we are reassured that the way we have dealt with the inconsistencies in the data minimizes misclassification of people into diabetes or no-diabetes and has reduced some of the measurement error in the diabetes data. Unfortunately we cannot use a similar method for dealing with inconsistencies in the self-reported year of diabetes diagnosis, as it has only been reported once. Hence, the results from duration analysis should be interpreted with care.

Acknowledgements

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

5 The effects of receiving a diabetes diagnosis on health behaviour and economic outcomes in China

6 Discussion and Conclusions

Bibliography

Aaronson, S. (2010). "Comment on "Measuring Labor Composition. A Comparison of Alternate Methodologies" Chapter." *Labor in the New Economy*. Ed. by K. G. Abraham, J. R. Spletzer, and M. Harper. University of Chicago Press. Chap. Comment on, 485–491.

Abdulkadri, A. O., Cunningham-Myrie, C., and Forrester, T. (2009). "Economic Burden of Diabetes and Hypertension in CARICOM States." *Social and Economic Studies* 58 (3-4), 175–197.

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

Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T., and Sidorchuk, A. (2011). "Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis." *International journal of epidemiology* 40 (3), 804–18.

Aguila, E., Diaz, C., Fu, M. M., Kapteyn, A., and Pierson, A. (2011). *Living longer in Mexico: Income security and health*. RAND Corporation.

Akobundu, E., Ju, J., Blatt, L., and Mullins, C. D. (2006). "Cost-of-illness studies : a review of current methods." *PharmacoEconomics* 24 (9), 869–90.

Alavinia, S. M. and Burdorf, A. (2008). "Unemployment and retirement and ill-health: a cross-sectional analysis across European countries." *International Archives of Occupational and Environmental Health* 82 (1), 39–45.

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Antillón, M., Lauderdale, D. S., and Mullahy, J. (2014). "Sleep behavior and unemployment conditions." *Economics and human biology* 14, 22–32.

Antonakis, J., Bendahan, S., Jacquart, P., Lalive, R., and Day, D. V. (2012). "Causality and endogeneity: problems and solutions." *The Oxford Handbook of Leadership and Organizations*. January. Oxford.

Arredondo, A. and Barcelo, A. (2007). “The economic burden of out-of-pocket medical expenditures for patients seeking diabetes care in Mexico.” *Diabetologia* 50 (11), 2408–2409.

Arredondo, A., Zúñiga, A., and Parada, I. (2005). “Health care costs and financial consequences of epidemiological changes in chronic diseases in Latin America: evidence from Mexico.” *Public health* 119 (8), 711–720.

Arredondo, A. and De Icaza, E. (2011a). “Costs of diabetes in Latin America: Evidences from the Mexican case.” *Value in Health.* Costos de la Diabetes en America Latina: Evidencias del Caso Mexicano 14 (5 Suppl 1), S85–88.

– (2011b). “The cost of diabetes in Latin America: evidence from Mexico.” *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research.* Costos de la Diabetes en America Latina: Evidencias del Caso Mexicano 14 (5 Suppl 1), S85–8.

Arredondo, A. and Zúñiga, A. (2004). “Economic consequences of epidemiological changes in diabetes in middle-income countries: The Mexican case.” *Diabetes Care* 27 (1), 104–109.

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

Baird, S., Gong, E., McIntosh, C., and Ozler, B. (2014). “The heterogeneous effects of HIV testing.” *Journal of health economics* 37C, 98–112.

Ballesta, M., Carral, F., Olveira, G., Girón, J. A., and Aguilar, M. (2006). “Economic cost associated with type II diabetes in Spanish patients.” *The European journal of health economics* 7 (4), 270–5.

Barceló, A., Aedo, C., Rajpathak, S., and Robles, S. (2003). “The cost of diabetes in Latin America and the Caribbean.” *Bulletin of the World Health Organization* 81 (1), 19–27.

Barquera, S., Hotz, C., Rivera, J., Tolentino, L., Espinoza, J., and Campos, I. (2006). “Food consumption, food expenditure, anthropometric status and nutrition-related diseases in Mexico. The double burden of malnutrition. Case studies from six developing countries.” Rome.

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

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

Bastida, E. and Pagán, J. A. (2002a). "The impact of diabetes on adult employment and earnings of Mexican Americans: findings from a community based study." *Health economics* 11 (5), 403–413.

Bastida, E. and Pagán, J. A. (2002b). "The impact of diabetes on adult employment and earnings of Mexican Americans: findings from a community based study." *Health economics* 11 (5), 403–13.

Bastida, J. L., Aguilar, P. S., and Gonzalez, B. D. (2002). "The social and economic cost of diabetes mellitus." *Atencion primaria / Sociedad Española de Medicina de Familia y Comunitaria*. Los costes socioeconomicos de la diabetes mellitus 29 (3), 145–150.

Basu, S., Yoffe, P., Hills, N., and Lustig, R. H. (2013). "The relationship of sugar to population-level diabetes prevalence: an econometric analysis of repeated cross-sectional data." *PloS ONE* 8 (2), e57873.

Baum, C., Schaffer, M., and Stillman, S. (2007). "Enhanced routines for instrumental variables/generalized method of moments estimation and testing." *Stata Journal* 7 (4), 465–506.

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

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

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

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

Biorac, N., Jakovljević, M., Stefanović, D., Perović, S., and Janković, S. (2009). "[Assessment of diabetes mellitus type 2 treatment costs in the Republic of Serbia]." *Vojnosanitetski pregled. Military-medical and pharmaceutical review* 66 (4), 271–276.

Birnbaum, H., Leong, S., and Kabra, A. (2003). "Lifetime medical costs for women: Cardiovascular disease, diabetes, and stress urinary incontinence." *Women's Health Issues* 13 (6), 204–213. URL: `http://www.ncbi.nlm.nih.gov/pubmed/14675789%20http:`

//ovidsp.ovid.com/ovidweb.cgi?T=JS%7B%5C&%7DPAGE=reference%7B%5C&%7DD=emed6%7B%5C&%7DNEWS=N%7B%5C&%7DAN=2003507923.

Bjegovic, V., Terzic, Z., Marinkovic, J., Lalic, N., Sipetic, S., and Laaser, U. (2007). "The burden of type 2 diabetes in Serbia and the cost-effectiveness of its management." *The European journal of health economics: HEPAC: health economics in prevention and care* 8 (2), 97–103.

Bolin, K., Gip, C., Mörk, A.-C., and Lindgren, B. (2009). "Diabetes, healthcare cost and loss of productivity in Sweden 1987 and 2005—a register-based approach." *Diabetic Medicine* 26 (9), 928–934.

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

Boutayeb, A. and Boutayeb, W. (2014). "Estimation of the direct cost of diabetes in the Arab region." *Mediterranean Journal of Nutrition and Metabolism* 7 (1), 21–32.

Brandle, M., Zhou, H., Smith, B. R. K., Marriott, D., Burke, R., Tabaei, B. P., Brown, M. B., and Herman, W. H. (2003). "The direct medical cost of type 2 diabetes." *Diabetes care* 26 (8), 2300–2304.

Bratti, M. and Mendola, M. (2014). "Parental health and child schooling." *Journal of Health Economics* 35 (1), 94–108.

Breton, M.-C., Guénette, L., Amiche, M. A., Kayibanda, J.-F., Grégoire, J.-P., and Moisan, J. (2013). "Burden of diabetes on the ability to work: a systematic review." *Diabetes care* 36 (3), 740–9.

Brown, H. S., Estrada, J. K., Hazarika, G., and Bastida, E. (2005). "Diabetes and the Labor Market: The community-wide economic cost in the Lower Rio Grande Valley." *Diabetes Care* 28 (12), 2945–2947.

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

Brown, H. S., Pagán, J. A., and Bastida, E. (2005). "The impact of diabetes on employment: genetic IVs in a bivariate probit." *Health economics* 14 (5), 537–44.

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

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

Bruno, G., Picariello, R., Petrelli, A., Panero, F., Costa, G., Cavallo-Perin, P., Demaria, M., and Gnavi, R. (2012). "Direct costs in diabetic and non diabetic people: the population-based Turin study, Italy." *Nutrition, metabolism, and cardiovascular diseases : NMCD* 22 (8), 684–90.

Buescher, P. A., Whitmire, J. T., and Pullen-Smith, B. (2010). "Medical care costs for diabetes associated with health disparities among adult Medicaid enrollees in North Carolina." *North Carolina medical journal* 71 (4), 319–324.

Cameron, A., Ewen, M., Ross-Degnan, D., Ball, D., and Laing, R. (2009). "Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis." *The Lancet* 373 (9659), 240–249.

Cameron, A., Roubos, I., Ewen, M., Mantel-Teeuwisse, A. K., Leufkens, H. G. M., and Laing, R. O. (2011). "Differences in the availability of medicines for chronic and acute conditions in the public and private sectors of developing countries." *Bulletin of the World Health Organization* 89 (6), 412–421.

Camilo González, J., Walker, J. H., Einarson, T. R., and González, J. C. (2009). "Cost-of-illness study of type 2 diabetes mellitus in Colombia." *Revista panamericana de salud pública = Pan American journal of public health* 26 (1), 55–63.

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

Cawley, J. and Meyerhoefer, C. (2012). "The medical care costs of obesity: An instrumental variables approach." *Journal of Health Economics* 31 (1), 219–230. arXiv: `arXiv:1011.1669v3`.

Chan, B. S. W., Tsang, M. W., Lee, V. W. Y., and Lee, K. K. C. (2007). "Cost of Type 2 Diabetes mellitus in Hong Kong Chinese." *International journal of clinical pharmacology and therapeutics* 45 (8), 455–468.

Chang, K. (2010). "Comorbidities, quality of life and patients' willingness to pay for a cure for type 2 diabetes in Taiwan." *Public health* 124 (5), 284–294.

Chatterjee, S., Riewpaiboon, A., Piyauthakit, P., Riewpaiboon, W., Boupaijit, K., Panpuwong, N., and Archavanuntagul, V. (2011). "Cost of diabetes and its complications in Thailand: a complete picture of economic burden." *Health & social care in the community* 19 (3), 289–298.

Chi, M.-j., Lee, C.-y., and Wu, S.-c. (2011). "The prevalence of chronic conditions and medical expenditures of the elderly by chronic condition indicator (CCI)." *Archives of gerontology and geriatrics* 52 (3), 284–289.

Chiburis, R. C., Das, J., and Lokshin, M. (2012). "A practical comparison of the bivariate probit and linear IV estimators." *Economics Letters* 117 (3), 762–766.

Chodick, G., Heymann, A. D., Wood, F., and Kokia, E. (2005). “The direct medical cost of diabetes in Israel.” *The European journal of health economics : HEPAC : health economics in prevention and care* 6 (2), 166–71.

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

Collins, J. J., Baase, C. M., Sharda, C. E., Ozminkowski, R. J., Nicholson, S., Billotti, G. M., Turpin, R. S., Olson, M., and Berger, M. L. (2005). “The Assessment of Chronic Health Conditions on Work Performance, Absence, and Total Economic Impact for Employers.” *Journal of Occupational and Environmental Medicine* 47 (6), 547–557.

Condliffe, S., Link, C. R., Parasuraman, S., and Pollack, M. F. (2013). “The effects of hypertension and obesity on total health-care expenditures of diabetes patients in the United States.” *Applied Economics Letters* 20 (7), 649–652. URL: `http://www.tandfonline.com/doi/abs/10.1080/13504851.2012.727966`.

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., Ogilvie, D., Petticrew, M., Reeves, B., Sutton, M., and Thompson, S. (2012). “Using natural experiments to evaluate population health interventions: new Medical Research Council guidance.” *Journal of epidemiology and community health* 66 (12), 1182–6. URL: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3796763%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract`.

Crimmins, E., McDade, T., Rubalcava, L., Seeman, T., Teruel, G., and Thomas, D. (2015). “Health of the Mexican population: Results from the Mexican Family Life Survey (MxFLS).” URL: `http://gero.usc.edu/CBPH/files/4%7B%5C_%7D30%7B%5C_%7D2014%7B%5C_%7DPAA/Thomas%7B%5C_%7DHealth%7B%5C_%7Dof%7B%5C_%7Dthe%7B%5C_%7DMexican%7B%5C_%7Dpopulation%7B%5C_%7Din%7B%5C_%7DMxFLS.pdf`.

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

Dall, T., Mann, S., Zhang, Y., Martin, J., Chen, Y., Hogan, P., and Petersen, M. (2008). “Economic costs of diabetes in the U.S. In 2007.” *Diabetes care* 31 (3), 596–615.

Dall, T., Nikolov, P., and Hogan, P. (2003). “Economic costs of diabetes in the US in 2002.” *Diabetes care* 26 (3), 917–932.

Dall, T. M., Zhang, Y., Chen, Y. J., Quick, W. W., Yang, W. G., and Fogli, J. (2010). “The economic burden of diabetes.” *Health Affairs* 29 (2), 297–303.

Davis, W. A., Knuiman, M. W., Hendrie, D., and Davis, T. M. E. (2006). “The obesity-driven rising costs of type 2 diabetes in Australia: projections from the Fremantle Diabetes Study.” *Internal medicine journal* 36 (3), 155–161.

Dawson, K. G., Gomes, D., Gerstein, H., Blanchard, J. F., and Kahler, K. H. (2002). "The economic cost of diabetes in Canada, 1998." *Diabetes care* 25 (8), 1303–7.

Denny, K. and Oppedisano, V. (2013). "The surprising effect of larger class sizes: Evidence using two identification strategies." *Labour Economics* 23, 57–65. URL: `http://linkinghub.elsevier.com/retrieve/pii/S0927537113000468`.

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

Drichoutis, A. C., Nayga, R. M., and Lazaridis, P. (2011). "Food away from home expenditures and obesity among older Europeans: are there gender differences?" *Empirical Economics* 42 (3), 1051–1078.

Druss, B., Marcus, S., and Olfson, M. (2001). "Comparing the national economic burden of five chronic conditions." *Health Affairs* 20 (6), 233–241.

Durden, E. D., Alemayehu, B., Bouchard, J. R., Chu, B.-C., and Aagren, M. (2009). "Direct health care costs of patients with type 2 diabetes within a privately insured employed population, 2000 and 2005." *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 51 (12), 1460–1465.

Elrayah-Eliadarous, H., Yassin, K., Eltom, M., Abdelrahman, S., Wahlström, R., and Ostenson, C.-G. (2010). "Direct costs for care and glycaemic control in patients with type 2 diabetes in Sudan." *Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association* 118 (4), 220–5.

Emran, M. S. and Shilpi, F. (2012). "The extent of the market and stages of agricultural specialization." *Canadian Journal of Economics/Revue canadienne d'économique* 45 (3), 1125–1153.

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

Esteghamati, A., Khalilzadeh, O., Anvari, M., Meysamie, A., Abbasi, M., Forouzanfar, M., and Alaeddini, F. (2009). "The economic costs of diabetes: a population-based study in Tehran, Iran." *Diabetologia* 52 (8), 1520–1527.

Ettaro, L., Songer, T. J., Zhang, P., and Engelgau, M. M. (2004). "Cost-of-illness studies in diabetes mellitus." *PharmacoEconomics* 22 (3), 149–164.

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

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

Frankenberg, E., Ho, J. Y., and Thomas, D. (2015). "Biological Health Risks and Economic Development." NBER Working Paper 21277.

Geishecker, I. and Siedler, T. (2011). "Job Loss Fears and (Extreme) Party Identification: First Evidence from Panel Data." cege Discussion Paper 129.

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., Pownall, H. J., Johnson, K. C., Safford, M. M., Kitabchi, A. E., Pi-Sunyer, F. X., Wing, R. R., Bertoni, A. G., and Look AHEAD Research Group, for the (2012). "Association of an Intensive Lifestyle Intervention With Remission of Type 2 Diabetes." *Journal of the American Medical Association* 308 (23), 2489.

Gyldmark, M. and Morrison, G. C. (2001). "Demand for health care in Denmark: results of a national sample survey using contingent valuation." *Social Science and Medicine* 53 (8), 1023–1036.

Harris, A. (2009). "Diabetes, Cardiovascular Disease and Labour Force Participation in Australia: An Endogenous Multivariate Probit Analysis of Clinical Prevalence Data." *Economic Record* 85 (271), 472–484.

Hemminki, K., Li, X., Sundquist, K., and Sundquist, J. (2010). "Familial Risks for Type 2 Diabetes in Sweden." *Diabetes Care* 33 (2), 293–297.

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

Herder, C. and Roden, M. (2011). "Genetics of type 2 diabetes: pathophysiologic and clinical relevance." *European Journal of Clinical Investigation* 41 (6), 679–692.

Herquelot, E., Guéguen, A., Bonenfant, S., and Dray-Spira, R. (2011). "Impact of diabetes on work cessation: data from the GAZEL cohort study." *Diabetes care* 34 (6), 1344–9.

Holmes, J., Gear, E., Bottomley, J., Gillam, S., Murphy, M., and Williams, R. (2003). "Do people with type 2 diabetes and their carers lose income? (T2ARDIS-4)." *Health policy* 64 (3), 291–296.

Honeycutt, A. A., Segel, J. E., Hoerger, T. J., and Finkelstein, E. A. (2009). "Comparing cost-of-illness estimates from alternative approaches: an application to diabetes." *Health services research* 44 (1), 303–320.

Honkasalo, M. T., Linna, M., Sane, T., Honkasalo, A., and Elonheimo, O. (2014). "A comparative study of two various models of organising diabetes follow-up in public primary health care - the model influences the use of services, their quality and costs." *BMC health services research* 14, 26.

Horak, P. (2009). "[Pharmacoeconomy of diabetes mellitus–trends in the Czech Republic]." *Vnitr Lek* 55 (4), 331–340.

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

– (2011b). "Globalization of diabetes: the role of diet, lifestyle, and genes." *Diabetes care* 34 (6), 1249–57.

Huang, H.-C. (, Lin, Y.-C., and Yeh, C.-C. (2009). "Joint determinations of inequality and growth." *Economics Letters* 103 (3), 163–166.

Huxley, R. (2006). "Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies." *British Medical Journal* 332 (7533), 73–78.

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

Javanbakht, M., Baradaran, H. R., Mashayekhi, A., Haghdoost, A. A., Khamseh, M. E., Kharazmi, E., and Sadeghi, A. (2011). "Cost-of-illness analysis of type 2 diabetes mellitus in Iran." *PloS one* 6 (10), e26864.

Johnson, J. A., Pohar, S. L., and Majumdar, S. R. (2006). "Health care use and costs in the decade after identification of type 1 and type 2 diabetes: a population-based study." *Diabetes care* 29 (11), 2403–2408.

Jönsson, B. (2002). "Revealing the cost of Type II diabetes in Europe." *Diabetologia* 45 (7), S5–12.

Kahn, M. (1998). "Health and labor market performance: the case of diabetes." *Journal of Labor Economics* 16 (4), 878–899.

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

Kelly, I. R., Dave, D. M., Sindelar, J. L., and Gallo, W. T. (2014). "The impact of early occupational choice on health behaviors." *Review of Economics of the Household* 12 (4), 737–770.

Khowaja, L. A., Khuwaja, A. K., and Cosgrove, P. (2007). "Cost of diabetes care in outpatient clinics of Karachi, Pakistan." *BMC health services research* 7, 189.

Kirigia, J. M., Sambo, H. B., Sambo, L. G., and Barry, S. P. (2009). "Economic burden of diabetes mellitus in the WHO African region." *BMC international health and human rights* 9, 6.

Knapp, L. G. and Seaks, T. G. (1998). *A Hausman test for a dummy variable in probit.*

Knaul, F. M., González-Pier, E., Gómez-Dantés, O., García-Junco, D., Arreola-Ornelas, H., Barraza-Lloréns, M., Sandoval, R., Caballero, F., Hernández-Avila, M., Juan, M., Kershenobich, D., Nigenda, G., Ruelas, E., Sepúlveda, J., Tapia, R., Soberón, G., Chertorivski, S., and Frenk, J. (2012). "The quest for universal health coverage: achieving social protection for all in Mexico." *Lancet* 380 (9849), 1259–79.

Köster, I., Ferber, L. von, Ihle, P., Schubert, I., and Hauner, H. (2006). "The cost burden of diabetes mellitus: the evidence from Germany–the CoDiM study." *Diabetologia* 49 (7), 1498–1504.

Köster, I., Huppertz, E., Hauner, H., and Schubert, I. (2011). "Direct costs of diabetes mellitus in Germany - CoDiM 2000-2007." *Experimental and Clinical Endocrinology and Diabetes* 119 (6), 377–385.

Köster, I., Schubert, I., and Huppertz, E. (2012). "Follow up of the CoDiM-Study: Cost of diabetes mellitus 2000-2009." *Deutsche Medizinische Wochenschrift*. Fortschreibung der KoDiM-Studie: Kosten des Diabetes mellitus 2000-2009 137 (19), 1013–1016.

Kraut, A., Walld, R., Tate, R., and Mustard, C. (2001). "Impact of diabetes on employment and income in Manitoba, Canada." *Diabetes care* 24 (1), 64–68.

Larg, A. and Moss, J. R. (2011). "Cost-of-illness studies: a guide to critical evaluation." *PharmacoEconomics* 29 (8), 653–71. URL: `http://www.ncbi.nlm.nih.gov/pubmed/21604822`.

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

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

Lau, R. S., Ohinmaa, A., and Johnson, J. A. (2011). "Predicting the Future Burden of Diabetes in Alberta from 2008 to 2035." *Canadian Journal of Diabetes* 35 (3), 274–281.

Laugesen, M. J. and Glied, S. a. (2011). "Higher fees paid to US physicians drive higher spending for physician services compared to other countries." *Health affairs (Project Hope)* 30 (9), 1647–56.

Lee, J.-A., Liu, C.-F., and Sales, A. E. (2006). "Racial and ethnic differences in diabetes care and health care use and costs." *Preventing chronic disease* 3 (3), A85.

Leijten, F. R. M., Heuvel, S. G. van den, Ybema, J. F., Beek, A. J. van der, Robroek, S. J. W., and Burdorf, A. (2014). "The influence of chronic health problems on work ability and productivity at work: a longitudinal study among older employees." *Scandinavian journal of work, environment & health* 40 (5), 473–82.

Lenneman, J., Schwartz, S., Giuseffi, D. L., and Wang, C. (2011). "Productivity and health: an application of three perspectives to measuring productivity." *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 53 (1), 55–61.

Lesniowska, J., Schubert, A., Wojna, M. M., Skrzekowska-Baran, I., Fedyna, M., Leśniowska, J., Schubert, A., Wojna, M. M., Skrzekowska-Baran, I., and Fedyna, M. (2014). "Costs of Diabetes and Its Complications in Poland." *The European journal of health economics.* ISPOR 14th Annual European Congress Madrid Spain. Conference Start: 20111105 Conference End: 20111108 15 (6), 653–660.

Lewbel, A. (2007). "Estimation of Average Treatment Effects with Misclassification." *Econometrica* 75 (2), 537–551.

– (2012). "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business & Economic Statistics* 30 (1), 67–80.

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

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

Lin, S. (2011). "Estimating the impact of diabetes on employment in Taiwan." *Economics Bulletin* 31 (4), 3089–3102.

Lin, T., Chou, P., Tsai, S.-T., Lee, Y.-C., and Tai, T.-Y. (2004). "Predicting factors associated with costs of diabetic patients in Taiwan." *Diabetes research and clinical practice* 63 (2), 119–125.

Linden, M. W. van der, Plat, A. W., Erkens, J. A., Emneus, M., and Herings, R. M. C. (2009). "Large impact of antidiabetic drug treatment and hospitalizations on economic burden of diabetes mellitus in The Netherlands during 2000 to 2004." *Value in health* 12 (6), 909–14.

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

Lorenzoni, L., Belloni, A., and Sassi, F. (2014). "Health-care expenditure and health policy in the USA versus other high-spending OECD countries." *Lancet* 384 (9937), 83–92. URL: http://www.ncbi.nlm.nih.gov/pubmed/24993914.

Lucioni, C., Garancini, M. P., Massi-Benedetti, M., Mazzi, S., and Serra, G. (2003). "The costs of type 2 diabetes mellitus in Italy: A CODE-2 sub-study." *Treatments in Endocrinology* 2 (2), 121–133.

Maahs, D. M., West, N. A., Lawrence, J. M., and Mayer-Davis, E. J. (2010). "Epidemiology of Type 1 Diabetes." *Endocrinology and Metabolism Clinics of North America* 39 (3), 481–497.

Maciejewski, M. and Maynard, C. (2004). "Diabetes-related utilization and costs for inpatient and outpatient services in the Veterans Administration." *Diabetes Care* 27 (SUPPL.2), B69–B73.

Marchesini, G., Forlani, G., Rossi, E., Berti, A., and De Rosa, M. (2011). "The direct economic cost of pharmacologically-treated diabetes in Italy-2006. The ARNO observatory." *Nutrition, metabolism, and cardiovascular diseases : NMCD* 21 (5), 339–46.

Martin, S., Schramm, W., Schneider, B., Neeser, K., Weber, C., Lodwig, V., Heinemann, L., Scherbaum, W. A., and Kolb, H. (2007). "Epidemiology of complications and total treatment costs from diagnosis of Type 2 diabetes in Germany (ROSSO 4)." *Experimental and clinical endocrinology & diabetes* 115 (8), 495–501.

Al-Maskari, F., El-Sadig, M., and Nagelkerke, N. (2010). "Assessment of the direct medical costs of diabetes mellitus and its complications in the United Arab Emirates." *BMC public health* 10 (1), 679.

Mata, M., Antoñanzas, F., Tafalla, M., and Sanz, P. (2002). "[The cost of type 2 diabetes in Spain: the CODE-2 study]." *Gaceta sanitaria / S.E.S.P.A.S* 16 (6), 511–520.

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

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

– (2011b). "The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey." *Health economics.* 15th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research, ISPOR 2010 Atlanta, GA United States. Conference Start: 20100515 Conference End: 20100519 20 (12), 1468–86.

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

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

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

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *PLoS medicine* 6 (7), e1000097.

Morsanutto, A., Berto, P., Lopatriello, S., Gelisio, R., Voinovich, D., Cippo, P. P., and Mantovani, L. G. (2006). "Major complications have an impact on total annual medical cost of diabetes: results of a database analysis." *Journal of Diabetes and its Complications* 20 (3), 163–169.

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

Nakamura, K., Okamura, T., Kanda, H., Hayakawa, T., Murakami, Y., Okayama, A., and Ueshima, H. (2008). "Medical expenditure for diabetic patients: a 10-year follow-up study of National Health Insurance in Shiga, Japan." *Public health* 122 (11), 1226–1228.

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

Ng, Y. C., Philip, J., and Johnson, J. (2001). "Productivity Losses Associated With Diabetes in the U.S." *Diabetes Care* 24 (2), 257–261.

Nolan, J. J., O'Halloran, D., McKenna, T. J., Firth, R., and Redmond, S. (2006). "The cost of treating type 2 diabetes (CODEIRE)." *Irish medical journal* 99 (10), 307–310.

Norlund, A., Apelqvist, J., Bitzén, P. O., Nyberg, P., and Scherstén, B. (2001). "Cost of illness of adult diabetes mellitus underestimated if comorbidity is not considered." *Journal of Internal Medicine* 250 (1), 57–65.

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

O'Connell, J. M., Wilson, C., Manson, S. M., and Acton, K. J. (2012). "The costs of treating American Indian adults with diabetes within the Indian Health Service." *American journal of public health* 102 (2), 301–308.

Ohinmaa, A., Jacobs, P., Simpson, S., and Johnson, J. (2004). "The projection of prevalence and cost of diabetes in Canada: 2000 to 2016." *Canadian Journal of Diabetes* 28 (2), 116–123.

Oliva, J., Lobo, F., Molina, B., and Monereo, S. (2004). "Direct health care costs of diabetic patients in Spain." *Diabetes care* 27 (11), 2616–2621.

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

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

Peele, P. B., Lave, J. R., and Songer, T. J. (2002). "Diabetes in employer-sponsored health insurance." *Diabetes care* 25 (11), 1964–1968.

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

Pohar, S. L. and Johnson, J. A. (2007). "Health care utilization and costs in Saskatchewan's registered Indian population with diabetes." *BMC health services research* 7, 126.

Pohar, S. L., Majumdar, S. R., and Johnson, J. A. (2007). "Health Care Costs and Mortality for Canadian Urban and Rural Patients with Diabetes: Population-Based Trends from 1993-2001." *Clinical Therapeutics* 29 (6 PART 1), 1316–1324.

Ramachandran, A., Ramachandran, S., Snehalatha, C., Augustine, C., Murugesan, N., Viswanathan, V., Kapur, A., and Williams, R. (2007). "Increasing expenditure on health care incurred by diabetic subjects in a developing country: a study from India." *Diabetes care* 30 (2), 252–256.

Ramsey, S., Summers, K. H., Leong, S. A., Birnbaum, H. G., Kemner, J. E., and Greenberg, P. (2002). "Productivity and medical costs of diabetes in a large employer population." *Diabetes care* 25 (1), 23–29. URL: http://www.ncbi.nlm.nih.gov/pubmed/11772896%20http://ovidsp.ovid.com/ovidweb.cgi?T=JS%7B%5C&%7DPAGE=reference%7B%5C&%7DD=emed5%7B%5C&%7DNEWS=N%7B%5C&%7DAN=11772896%20http://ovidsp.ovid.com/ovidweb.cgi?T=JS%7B%5C&%7DPAGE=reference%7B%5C&%7DD=emed5%7B%5C&%7DNEWS=N%7B%5C&%7DAN=2005342883.

Redekop, W. K., Koopmanschap, M. A., Rutten, G. E. H. M., Wolffenbuttel, B. H. R., Stolk, R. P., and Niessen, L. W. (2002). "Resource consumption and costs in Dutch patients with type 2 diabetes mellitus. Results from 29 general practices." *Diabetic medicine: a journal of the British Diabetic Association* 19 (3), 246–253.

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

Ricordeau, P., Weill, A., Vallier, N., Bourrel, R., Schwartz, D., Guilhot, J., Fender, P., and Allemand, H. (2003). "The prevalence and cost of diabetes in metropolitan France: What trends between 1998 and 2000?" *Diabetes and Metabolism* 29 (5), 497–504.

Ringborg, A., Martinell, M., Stålhammar, J., Yin, D. D., and Lindgren, P. (2008). "Resource use and costs of type 2 diabetes in Sweden - estimates from population-based register data." *International journal of clinical practice* 62 (5), 708–716.

Rivera, J. a., Barquera, S., Campirano, F., Campos, I., Safdie, M., and Tovar, V. (2002). "Epidemiological and nutritional transition in Mexico: rapid increase of non-communicable chronic diseases and obesity." *Public Health Nutrition* 5 (1a), 113–122.

Rivera, J. A., Barquera, S., González-Cossío, T., Olaiz, G., and Sepúlveda, J. (2004). "Nutrition Transition in Mexico and in Other Latin American Countries." *Nutrition Reviews* 62 (July), S149–S157.

Rodbard, H. W., Green, A. J., Fox, K. M., and Grandy, S. (2010). "Impact of type 2 diabetes mellitus on prescription medication burden and out-of-pocket healthcare expenses." *Diabetes research and clinical practice* 87 (3), 360–365.

Rodríguez Bolaños, R. d. L. Á., Reynales Shigematsu, L. M., Jiménez Ruíz, J. A., Juárez Márquezy, S. A., Hernández Ávila, M., and R., D. l. A. R. B. (2010). "Direct costs of medical care for patients with type 2 diabetes mellitus in Mexico: Micro-costing analysis." *Revista Panamericana de Salud Publica/Pan American Journal of Public Health.* Costos directos de atencion medica en pacientes con diabetes mellitus tipo 2 en Mexico: analisis de microcosteo 28 (6), 412–420.

Rubalcava, L. and Teruel, G. (2008). "User's Guide for the Mexican Family Life Survey Second Wave."

– (2013). "User's Guide for the Mexican Family Life Survey Third Round."

Salomon, J. a., Carvalho, N., Gutierrez-Delgado, C., Orozco, R., Mancuso, A., Hogan, D. R., Lee, D., Murakami, Y., Sridharan, L., Medina-Mora, M. E., and Gonzalez-Pier, E. (2012). "Intervention strategies to reduce the burden of non-communicable diseases in Mexico: cost effectiveness analysis." *BMJ* 344, e355.

Samb, B., Desai, N., Nishtar, S., Mendis, S., Bekedam, H., Wright, A., Hsu, J., Martiniuk, A., Celletti, F., Patel, K., Adshead, F., McKee, M., Evans, T., Alwan, A., and Etienne, C. (2010). "Prevention and management of chronic disease: a litmus test for health-systems strengthening in low-income and middle-income countries." *The Lancet* 376 (9754), 1785–1797.

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

Schmitt-Koopmann, I., Schwenkglenks, M., Spinas, G. A., and Szucs, T. D. (2004). "Direct medical costs of type 2 diabetes and its complications in Switzerland." *European journal of public health* 14 (1), 3–9.

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

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

Schroeter, C., Anders, S., and Carlson, A. (2012). "The Economics of Health and Vitamin Consumption." *Applied Economic Perspectives and Policy* 35 (1), 125–149.

Segel, J. E. (2006). "Cost-of-Illness Studies — A Primer." *RTI-UNC Center of Excellence in Health Promotion Economics.*

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

Seuring, T., Goryakin, Y., and Suhrcke, M. (2014). "The impact of diabetes on employment in Mexico." *CHE Research Paper.* CHE Research Paper (100).

– (2015). "The impact of diabetes on employment in Mexico." *Economics & Human Biology* 18, 85–100.

Shemilt, I., Thomas, J., and Morciano, M. (2010). "A web-based tool for adjusting costs to a specific target currency and price year." *Evidence and Policy* 6 (1), 51–59.

Sicree, B. R., Shaw, J., and Zimmet, P. (2011). *The Global Burden: Diabetes and Impaired Glucose Tolerance.* Brussels, Belgium: International Diabetes Federation.

Simpson, S. H., Corabian, P., Jacobs, P., and Johnson, J. A. (2003). "The cost of major comorbidity in people with diabetes mellitus." *Canadian Medical Association Journal* 168 (13), 1661–1667.

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

Smith-Spangler, C. M., Bhattacharya, J., and Goldhaber-Fiebert, J. D. (2012). "Diabetes, its treatment, and catastrophic medical spending in 35 developing countries." *Diabetes care* 35 (2), 319–326.

Solli, O., Jenssen, T., and Kristiansen, I. S. (2010). "Diabetes: cost of illness in Norway." *BMC endocrine disorders* 10, 15.

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.

Squires, D. A. (2012). "Explaining High Health Care Spending in the United States: An International Comparison of Supply, Utilization, Prices, and Quality The." 10 (May), 1–14.

Staiger, D. and Stock, J. (1997). "Instrumental variables regression with weak instruments." *Econometrica* 65 (3), 557–586.

Stevens, G., Dias, R. H., Thomas, K. J. A., Rivera, J. A., Carvalho, N., Barquera, S., Hill, K., and Ezzati, M. (2008). "Characterizing the Epidemiological Transition in Mexico: National and Subnational Burden of Diseases, Injuries, and Risk Factors." *PLoS Medicine* 5 (6). Ed. by M. Tobias, e125.

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

Suleiman, I., Fadeke, O., and Okubanjo, O. (2006). "Pharmacoeconomic Evaluation of Anti-Diabetic Therapy in A Nigerian Tertiary Health Institution." *Annals of African Medicine* 5 (3), 132–137.

Tharkar, S., Devarajan, A., Kumpatla, S., and Viswanathan, V. (2010). "The socioeconomics of diabetes from a developing country: A population based cost of illness study." *Diabetes Research and Clinical Practice* 89 (3), 334–340.

The Interact Consortium (2013). "The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study." *Diabetologia* 56 (1), 60–9.

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

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

Trogdon, J. G. and Hylands, T. (2008). "Nationally representative medical costs of diabetes by time since diagnosis." *Diabetes care* 31 (12), 2307–2311.

Tunceli, K., Bradley, C. J., Nerenz, D., Williams, L. K., Pladevall, M., and Elston Lafata, J. (2005). "The impact of diabetes on employment and work productivity." *Diabetes care* 28 (11), 2662–2667.

Tunceli, K., Zeng, H., Habib, Z. A., and Williams, L. K. (2009). "Long-term projections for diabetes-related work loss and limitations among U.S. adults." *Diabetes research and clinical practice* 83 (1), e23–25.

Tunceli, O., Wade, R., Gu, T., Bouchard, J. R., Aagren, M., and Luo, W. (2010). "Cost of diabetes: comparison of disease-attributable and matched cohort cost estimation methods." *Current medical research and opinion* 26 (8), 1827–34.

Valdmanis, V., Smith, D. W., and Page, M. R. (2001). "Productivity and economic burden associated with diabetes." *American journal of public health* 91 (1), 129–30.

Vijan, S., Hayward, R. A., and Langa, K. M. (2004a). "The impact of diabetes on workforce participation: results from a national household sample." *Health services research* 39 (6 Pt 1), 1653–69.

– (2004b). "The impact of diabetes on workforce participation: results from a national household sample." *Health services research* 39 (6 Pt 1), 1653–69.

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

Wang, W., Fu, C. W., Pan, C. Y., Chen, W., Zhan, S., Luan, R., Tan, A., Liu, Z., and Xu, B. (2009). "How do type 2 diabetes mellitus-related chronic complications impact direct medical cost in four major cities of urban China?" *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 12 (6), 923–9.

Wang, W., Fu, C., Zhuo, H., Luo, J., and Xu, B. (2010). "Factors affecting costs and utilization of type 2 diabetes healthcare: a cross-sectional survey among 15 hospitals in urban China." *BMC health services research* 10.

Wang, W., McGreevey, W. P., Fu, C., Zhan, S., Luan, R., Chen, W., and Xu, B. (2009). "Type 2 diabetes mellitus in China: a preventable economic burden." *The American journal of managed care* 15 (9), 593–601.

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

Wiréhn, A.-B., Andersson, A., Ostgren, C. J., and Carstensen, J. (2008). "Age-specific direct healthcare costs attributable to diabetes in a Swedish population: a register-based analysis." *Diabetic medicine : a journal of the British Diabetic Association* 25 (6), 732–7.

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

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT press.

World Bank. *World Bank Analytical Classifications*. URL: `http://siteresources.worldbank.org/DATASTATISTICS/Resources/OGHIST.xls`.

World Health Organization (2011). "Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation."

Wunder, C. and Riphahn, R. T. (2014). "The dynamics of welfare entry and exit amongst natives and immigrants." *Oxford Economic Papers* 66 (2), 580–604.

Yach, D., Stuckler, D., and Brownell, K. D. (2006). "Epidemiologic and economic consequences of the global epidemics of obesity and diabetes." *Nature medicine* 12 (1), 62–6.

Yang, W., Zhao, W., Xiao, J., Li, R., Zhang, P., Kissimova-Skarbek, K., Schneider, E., Jia, W., Ji, L., Guo, X., Shan, Z., Liu, J., Tian, H., Chen, L., Zhou, Z., Ji, Q., Ge, J., Chen, G., and Brown, J. (2012). "Medical care and payment for diabetes in China: enormous threat and great opportunity." *PloS one* 7 (9), e39513. URL: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3458850%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract`.

Yu, C. H. and Zinman, B. (2007). "Type 2 diabetes and impaired glucose tolerance in aboriginal populations: A global perspective." *Diabetes Research and Clinical Practice* 78 (2), 159–170.

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

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

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

– (2009b). "Chronic diseases and labour force participation in Australia." *Journal of Health Economics* 28 (1), 91–108.

Zhao, F.-L., Xie, F., Hu, H., and Li, S.-C. (2013). “Transferability of indirect cost of chronic disease: a systematic review and meta-analysis.” *PharmacoEconomics* 31 (6), 501–8. URL: http://www.ncbi.nlm.nih.gov/pubmed/23620212.

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

Zhou, H., Isaman, D. J. M., Messinger, S., Brown, M. B., Klein, R., Brandle, M., and Herman, W. H. (2005). “A computer simulation model of diabetes progression, quality of life, and cost.” *Diabetes care* 28 (12), 2856–2863.