**Abstract**

A diabetes diagnosis entails important consequences for its recipients. They obtain important health information but also face the challenge of having to manage the condition via lifestyle adjustments, with potential consequences for – among other things – their economic activity. We investigate the causal eﬀect of a diabetes diagnosis on health behaviours as well as on employment chances, two potentially intertwined factors. We used lon-gitudinal data from the China Health and Nutrition Survey (CHNS), covering the years 1997 to 2011. Two complementary statistical techniques—marginal structural models and fixed eﬀects panel estimation—were used to estimate the causal eﬀect of a diabetes diagnosis on alcohol consumption, smoking, BMI, waist circumference and daily calorie consumption as well as employment probabilities. Both models suggest female employment chances to decline significantly as a result of a diabetes diagnosis (over 11 percentage points), but no adverse effects for men. Chinese men appear to respond to a diagnosis by significantly reducing their body mass index (BMI), waist circumference and calorie consumption, in ways that are sustained over time. The eﬀects on behavioural outcomes for women are smaller and less consistent. In light of the results it may be worthwhile for the Chinese healthcare system to focus efforts on addressing the needs of women with diabetes as it is them who have particular difficulties in improving their behavioural risk factors and who – perhaps as a consequence – pay the biggest price in terms of loss of employment chances. Future research is needed to unravel the mechanism behind these sex diﬀerences.

**0.1 Introduction**

While the risk factors for type 2 diabetes, post diagnosis blood glucose management and the resulting complications of poor management have received considerable research attention —including in the context of China (Batis et al.,  [2014;](#page33) Chan et al.,  [2014](#page33); Ma et al.,  [2014;](#page34) Pan,  [2015;](#page34) Yang, Zhao, et al.,  [2012;](#page35) Zhao, Zhu, et al.,  [2012](#page35))—, the impact of health information received at diabetes diagnosis on health behaviours and economic outcomes is less well known. This is despite research suggesting that behaviour changes after a diabetes diagnosis can have substantial positive eﬀects and reduce the risk of subsequent cardiovascular events(Long et al.,  [2014)](#page34) and may help in eﬀectively managing blood glucose levels and achieving further treatment goals (Zhou et al.,  [2016](#page35)).

Such information may be particularly important for low- and middle-income countries (LMICs) such as China, where diabetes prevalence has surged from 1% in the early 1980s to about 10% in recent years (Hu,  [2011;](#page33) NCD Risk Factor Collaboration,  [2016](#page34)). Confronting this diabetes epidemic puts a strain on healthcare systems (Seuring, Archangelidi, et al.,  [2015),](#page35) increasing the need to find highly cost-eﬀective prevention and treatment options in very resource constraint settings. However, to do this it is important to assess how successful the current system is in promoting positive health behaviours that are known to reduce the burden of diabetes.

1

So far, population-level research on the eﬀects of health information on post-diagnosis behaviour change is scarce and has been limited to high-income countries. The sole study on the eﬀects of recently diagnosed diabetes found positive behaviour changes shortly after diagnosis in a USA population. However, the eﬀects were mostly short lived and tended to dissipate over time, particularly considering weight loss (Slade,  [2012](#page35)). Slade created an "at risk" control group without diabetes that intended to be similar to the treatment group with diabetes, apart from not having received a diagnosis. He used information on diabetes biomarkers to estimate the propensity score of those without a diabetes diagnosis to be above a specific at risk threshold, so that everybody above a certain propensity score was used to form the control group. He then estimates dynamic population averaged as well as fixed eﬀects (FE) models for identification. While this allowed for the construction of a potentially better control group, the study was not able to account for the possibility of selection into treatment based on pre-treatment values of the lagged dependent variables.

Another study investigated the eﬀect of a hypertension diagnosis on nutritional be-haviours in China using a regression-discontinuity design and biomarker information on blood pressure (Zhao, Konishi, et al.,  [2013](#page35)). A crucial assumption in the study was that people above the hypertension threshold were indeed informed about that outcome while those just below the threshold were not. These two groups were then compared to isolate the particular ef-fect of the additional health information on food consumption. The results indicated that a diagnosis leads to reductions in fat consumption of the rich, at the consecutive wave 3–4 years later. However, an important caveat of the study is that the above assumption about who gets (not) informed about the outcome of the examination might not have been fully met. It remains somewhat unclear whether the participants received just the actual blood pressure measurement information and had to interpret these data them-selves, or whether they were made explicitly aware of their hypertension (or also pre-hypertension) status (Zhao, Konishi, et al.,  [2013](#page35)). Further, the results may have limited generalisability, since the measured treatment effect is a very local one, applying only to the population around the hypertension threshold. The eﬀects might be diﬀerent for people receiving their ‘diagnosis’ and having very severe hypertension already.

This study adds in several ways to the existing literature. First, it shows the impact of diabetes diagnosis on labour outcomes in China, not only over the short term, but for a period covering the entire decade of the 2000s, allowing for a more long term investigation of the eﬀects. This both confirms and extends earlier evidence for other settings and using different methods. Second, it provides information on the eﬀect of a diabetes diagnosis on health behaviours. Third, by considering the effects over time on both employment and health behaviour simultaneously, the results shed light on potential pathways through which the impact on employment may work. Fourth, the study provides a methodological innovation by using both MSM and FE estimation methods, offering insights not only on the robustness of MSM results, but also on the validity of some of its assumptions

2

We used panel data from six waves of the China Health and Nutrition Survey (CHNS), spanning the period from 1997 to 2011. The survey provides one of the most compre-hensive sources of data on nutrition and health in China and has been used widely in epidemiological and economic studies, particularly to investigate the causes of nutrition and health changes over the last two decades (Zhang et al.,  [2014](#page35)). We hypothesized that a diagnosis of diabetes leads to positive behaviour change, at least initially after diagnosis, but at the same time reduces employment probabilities. We investigated both, the overall average change in outcomes after a diagnosis as well as the time trends.

**0.2 Methods**

**Study sample**

The CHNS is an international collaborative project led by the Carolina Population Center at the University of North Carolina at Chapel Hill investigating nutrition and health behaviours in nine provinces of China (Zhang et al.,  [2014](#page35)). We use data from 1997 onwards, which was the first time survey participants provided diabetes information. In total we use six waves (1997, 2000, 2004, 2006, 2009 and 2011) obtained from the longitudinal dataset released in 2015. The data provide extensive information on nutrition and health, including anthropometric measures of weight and height, reducing potential measurement issues. It further provides socioeconomic information, most importantly for this study about employment. The sample is limited to the adult population from age 18–64. The sample is not nationally representative and as such does not provide sampling weights (Popkin et al.,  [2010](#page34)).

Overall, between 84% to 90% are followed up in the following wave, with attrition being highest after 2006. Attrition in the CHNS due to mortality is around 1 percent in the CHNS (Popkin et al.,  [2010](#page34)). Other reasons mentioned by Popkin et al.  [(2010)](#page34) are loss in follow up due to migration, natural disasters and redevelopment of housing in the urban centres leading to relocations. We analysed if any of our variables of interest was significantly related to attrition and did only find lower calorie consumption to predict attrition. Further, attrition was related to urbanization, level of education and being of

3

younger age, suggesting that mostly younger, more urbanized participants tended to leave the survey. Given the generally older population of people with diabetes we are interested in, we do not think that attrition will aﬀect our estimates to any major extend. IS THIS SUFFICIENT REGARDING ATTRITION OR SHOULD I DO ANY FURTHER ANALYSIS. SO FAR I ONLY REGRESSED ATTRITION ON THE VARIABLES WE USE IN THIS PAPER.

**Assessment of diabetes**

We used self-reported information on a diabetes diagnosis to construct our diabetes in-dicator. We only relied on incident cases of self-reported diabetes, excluding individuals with self-reported diabetes at baseline. Given the chronic nature of diabetes, we assumed that after the initial diagnosis diabetes persisted for the rest of one’s life. To construct a measure of diabetes duration for incidence cases we used self-reported information on the year of diagnosis. If we found that the year of diagnosis was reported to be before the last wave without a reported diagnosis, we used the midpoint between the last wave without diagnosis and the first wave with a diagnosis as the year of diagnosis.

**Assessment of outcomes**

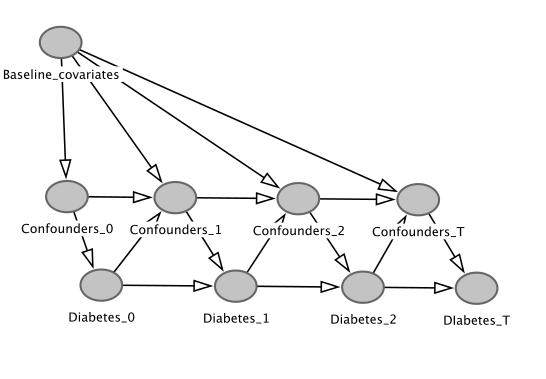
The economic outcome we focus on is employment status, based on a self-reported measure of if the person was currently working. People who were not working because they were students were excluded. We did include those that were not working due to any other reason such as doing housework, being disabled or being retired. The behavioural outcomes we estimated were current smoking status, any alcohol con-sumption over the last year, BMI, waist circumference in centimetres and daily calorie consumption. Smoking status and alcohol consumption were self-reported, while BMI and waist circumference were based on anthropometric measurements, minimizing potential reporting errors. Waist circumference is reported in centimetres. Finally, daily calorie consumption is a constructed variable available in the CHNS, based on the average daily consumption of carbohydrates, protein and fat of every individual in the survey, measured on three consecutive days.

**Statistical analysis**

We used two statistical approaches to account for potential confounding: marginal struc-tural models (MSMs) and fixed eﬀects panel estimation (FE).

4

Figure 0.1: Direct acyclic graph (DAG) representing the relations between con-founders/outcomes and a diabetes diagnosis.



Unobserved

time invariant

confounders

*Notes* The confounders not only include control variables but also pre-treatment values of our outcomesof interest to account for any predetermination of the treatment.

**Marginal structural models**

MSMs apply inverse probability weights to adjust for confounding and selection bias as a result of time-varying confounders being aﬀected by prior exposure (Robins et al.,  [2000](#page34)). Under the assumption of the MSM(Robins et al.,  [2000](#page34))—the reported treatment is the treatment that has actually been received (consistency), there are no unmeasured con-founders (exchangeability) and every person in the sample has a non-zero chance of re-ceiving the treatment (positivity) (see Section  [0.4](#page17) for a discussion of the validity of these assumptions in our case)—the causal DAG shown in Figure  [0.1](#page5) displays the association between confounders and outcomes and a diabetes diagnosis. This causal graph was drawn using DAGitty program version 2.3.(Textor et al.,  [2011](#page35)).

We accounted for two sets of confounders: time-varying variables at baseline as well as lagged time-varying variables in the waves following the baseline, to capture the eﬀects of changes in these variables over time. In our context it seemed possible that, for example, BMI could aﬀect the probability of being diagnosed with diabetes which then itself may aﬀected subsequent BMI levels, confounding the relationship between a diabetes diagnosis and BMI due to non-random selection. Similarly employment history and current em-ployment could aﬀect the probability of self-reporting a diabetes diagnosis through their impact on lifestyle, potentially due to an increase in disposable income or a reduction in leisure time, potentially confounding the relationship between a diabetes diagnosis and

5

employment status. We used time-invariant variables lagged by one period to prevent reverse causality in the prediction of diabetes risk. Because current diabetes status as reported in the survey was determined at some point within the current and the previous wave, we had to use confounders from the previous wave determined before the current diabetes status, to prevent reverse causality.

Using MSM and creating inverse probability of treatment weights, we were able to con-struct a pseudo population, based on the person’s potential exposure at each time point, that allowed us to adjust the estimates for confounding due to selection. We first con-structed unstabilized weights using baseline values of the time-invariant and time-variant confounders as well as time-variant confounders lagged by one period to predict the prob-ability of developing diabetes at each wave. The used predictors were age and age squared to account for changes in risk with increasing age, an index of urbanization pre-constructed within the CHNS data, ranging from 1 to 120 as the level of urbanization increases (Zhang et al.,  [2014),](#page35) to account for the impact of urbanization on diabetes risk (Attard et al.,  [2012](#page33)). We also used secondary and university education, being married, having any med-ical insurance, being of Han ethnicity, living in a rural area, dummies for the diﬀerent Chinese regions and the respective survey waves as predictors. Further we used inflation adjusted per-capita household expenditures to predict any eﬀects of household wealth on diabetes. Finally, all outcome variables (employment status, alcohol consumption, smok-ing status, BMI, waist circumference and average daily calorie consumption) were used as predictors.

Because unstabilized weights can be highly variable it is recommended to stabilize the weights (Cole and Hernan,  [2008](#page33)). Using the unstabilized weights as the denominator, stabilized weights were calculated by dividing the denominator by the predicted treatment propensity from a model using only time-invariant confounders and baseline information of the time-variant confounders as predictors. Because our analysis was stratified by males and females, we created weights separately for both groups.

The MSMs were estimated using ordinary least squares (OLS) for the continuous and a logistic model for the binary outcomes. For the logistic model we calculated marginal eﬀects and present them instead of odds ratios for greater comparability with the results of the FE models. All models were weighted by the stabilized weights constructed beforehand while adjusting for all baseline and time-invariant covariates used in the calculation of the stabilized weights, except for the respective outcome of interest. Robust standard errors to account for intra-class correlation of repeated outcome measurements in individuals were used throughout. In our primary analysis, we present the results of the MSM with untruncated stabilized weights, as these present unbiased estimates, albeit they likely are

6

less eﬃcient than truncated weights (Cole and Hernan,  [2008](#page33)).

**Fixed eﬀects estimation**

While the MSM can account for pre-treatment selection on observable and time-variant confounders, it assumes that time-invariant unobserved confounders such as family background, cognitive abilities, motivation and other personal characteristics play no role. This is a strong assumption that may be violated. The individual level FE model can help remedy this problem as it is able to account for both observed time-variant variables and time-invariant unobserved variables. It does so by demeaning the confounding variables by the mean of the confounder over all observed time points for an individual. It then uses solely the within-person variation for identification, thereby accounting for any time-invariant observed or unobserved as well as observed time-variant eﬀects.

This comes at a price: due to the demeaning, time-invariant variables such as Han ethnicity, were dropped from the model and could not be estimated. Further, because the FE model is not able to account for any treatment eﬀects on other time-variant confounders, only a more limited set of confounders could be included compared to the MSM. Otherwise our estimate of the eﬀect of a diabetes diagnosis would likely be biased due to the inclusion of ‘bad controls’ (Angrist and Pischke,  [2008](#page33)). Bad controls are those that have been aﬀected by the treatment itself—such as BMI or smoking status after a diabetes diagnosis—and therefore likely capture part of the causal eﬀect of diabetes on our outcome of interest, biasing the diabetes coeﬃcient (Angrist and Pischke,  [2008](#page33)). The FE models thus only included controls for age, age squared, the level of urbanization, education, being married, having any medical insurance, living in a rural area, and dummies for the diﬀerent Chinese regions and the respective survey waves. For the estimation of the eﬀect of time since diagnosis, the linear age variable was dropped. In FE models, two or more variables that change at the same rate between waves cannot be separately identified. Here this was the case with age and time-dummies, as both variables increased by one each additional year (Wooldridge,  [2012](#page35)). To identify the eﬀect of diabetes duration we had to rely on the presence of people without diabetes in the sample, for which diabetes duration did not increase at the same rate as time.

Because it is not possible to retrieve marginal eﬀects from a logistic FE model, we used a FE linear probability model instead. It generally produces very similar estimates compared to non-linear models (Angrist and Pischke,  [2008](#page33)). COMMENT TILL: CAN ALSO ADDITIONALLY PRESENT RESULTS FROM LOGIT MODELS IF NEEDED TO SHOW THAT THEY INDICATE SIMILAR EFFECTS. NOT SURE IF NEEDED AS THIS MEANS A LOT MORE TABLES

7

**Multiple imputation**

To deal with missing data, we used chained multiple imputation to impute the missing values in Stata 13 using the user written ICE command (Royston and White,  [2009](#page35)). Imputation models included all variables used in the MSMs. We did not impute missing diabetes information and instead assumed that once a diabetes diagnosis was reported, the individual had diabetes in every ensuing wave, even when the observation was missing. If diabetes was never reported in any wave, we assumed that the individual never had diabetes. Consequently, we only imputed missing values for those observations that had a non-missing diabetes status. For the calculation of the marginal eﬀects in the MSM logit models, Rubin’s rules were applied using the user written Stata command mimrgns (Klein,  [2014](#page33)).

**Number of observations**

Because we used lagged variables to construct the stabilized weights for the MSMs, the number of observations used in the MSMs was lower than those used in the FE models, where we did not use lagged variables. The summary statistics shown in Table  [0.1](#page10) are based on the observations used in the FE models.

**Sensitivity analyses**

We conducted three additional sensitivity analyses in order to test the robustness of our results to diﬀerent assumptions and estimation strategies. First, we truncated weights at the 1st and 99th percentile to investigate the sensitivity of the MSMs to the most extreme weights. While untruncated weights provide unbiased estimates under the assumptions of the MSM, they may not be the most eﬃcient and tend to have larger standard errors (Cole and Hernan,  [2008](#page33)). Second, we estimated all models using only covariate adjustment to investigate in how far this ’naive’ approach diverts from the "causal" estimates of the FE and MSMs. Third, we estimated the FE and MSMs using the original non-imputed data to ascertain the extent to which multiple imputation aﬀected the results.

**0.3 Results**

From the descriptive statistics, we can observe that people with self-reported diabetes were less likely to be employed. Looking at health behaviours, it is mainly men who smoke and report alcohol consumption while very few women do so. The prevalence of smoking and drinking was lower for men with diabetes; they also consumed fewer calories

8

compared to men without diabetes. Further, the self-reported diabetes group had both higher BMI and waist circumference levels. They were also older, lived in more urbanized areas, were more likely to have insurance and men were somewhat better educated while women were less educated compared to their counterparts without diabetes. Both men and women reported an average time since diagnosis of around 4 years. Interestingly, while there is no significant diﬀerence in household expenditures for men, women with diabetes come from families with significantly lower household expenditures, suggesting that they live in lower resource settings than women without diabetes. Overall it appears that in China it is less educated and poorer women that self-report a diagnosis, while self-reporting men are better educated and are not poorer than their counterparts without diabetes.

9

|  |
| --- |
| 10 |

Table 0.1: Sample means for males and females, by diabetes status

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Males |  |  |  | Females |  |  |
|  |  |  |  |  |  |  |  |  |
|  | No diabetes | Diabetes | p-value (t-test) |  | No diabetes | Diabetes | p-value (t-test) |  |
|  |  |  |  |  | |  |  |  |
| Employed | 81% | 69% | 0.000 | 66% | | 28% | 0.000 |  |
| Smokes | 59% | 46% | 0.000 | 3% | | 4% | 0.000 |  |
| Any alcohol consumption | 64% | 53% | 0.000 | 9% | | 4% | 0.000 |  |
| Daily Kcal eaten (3-day average) | 2438 | 2149 | 0.000 | 2065 | | 1919 | 0.001 |  |
| BMI | 23.04 | 25.22 | 0.000 | 23.15 | | 25.89 | 0.000 |  |
| Waist circ. (cm) | 82.28 | 89.70 | 0.000 | 78.92 | | 88.04 | 0.000 |  |
| Age | 43.33 | 52.90 | 0.000 | 43.70 | | 55.56 | 0.000 |  |
| Han ethnicity | 87% | 89% | 0.443 | 87% | | 91% | 0.079 |  |
| Rural area | 68% | 51% | 0.000 | 68% | | 48% | 0.000 |  |
| Married | 84% | 95% | 0.000 | 88% | | 85% | 0.108 |  |
| Secondary education | 64% | 67% | 0.401 | 49% | | 40% | 0.006 |  |
| University education | 6% | 14% | 0.000 | 4% | | 1% | 0.017 |  |
| Any health insurance | 52% | 82% | 0.000 | 50% | | 67% | 0.000 |  |
| Urbanization Index | 61.27 | 75.62 | 0.000 | 61.92 | | 69.25 | 0.000 |  |
| Household expenditures (Yuan (2011)) | 5013.40 | 4178.76 | 0.927 | 4610.57 | | 1829.39 | 0.020 |  |
| Years since diabetes diagnosis | − | 3.93 | − |  | − | 4.22 | − | |
| Observations | 23413 | 284 | 23697 | 23577 | | 336 | 23913 |  |
|  |  |  |  |  |  |  |  |  |

Predicting the denominator for the stabilized weights we find that for men a higher baseline BMI, increasing age and increasing household expenditures have some predictive value for a diabetes diagnosis, as does employment (Table  [0.2](#page12)). For women, higher age at baseline, increases in BMI and waist circumference as well as living in a non-rural environment predict a diabetes diagnosis. Further, especially living in certain provinces of the country appears to increase the likelihood to receive a diabetes diagnosis for men, but not for women.

The results of our regression analysis are presented in Table  [0.3.](#page13) Both the FE model and the MSM indicate that women self-reporting a diabetes diagnosis have lower prob-abilities of being employed than their counterparts without diabetes, with a reduction of 11 percentage points in the FE model and 12 percentage points in the MSM. For men no such eﬀect is observed (even though it is significant in Table 0.2).

There is a more ambiguous picture for the eﬀect of a diabetes diagnosis on behavioural outcomes. There is no consistent – and only marginally significant evidence in the MSM model – that men reduced their smoking rate, however, a diabetes diagnosis led to a reduction in alcohol consumption as according to either model, though particularly so for the MSM. For waist circumference, BMI and calorie consump-tion, the FE and MSM both indicate reductions in BMI of close to 0.7, of about 2 cm in waist circumference and of up to 170 calories per day for men. Results for women look different, in that while the point estimates indicate a reduction in all outcomes, these tend to be smaller than those for men and only statistically significant in the FE model for BMI and waist circumference.

Exploring the eﬀect of a diabetes diagnosis over time, we first estimated a specification using time since diagnosis as a continuous variable. The results of the FE model (Table  [0.4](#page14)) indicate a steady reduction of female employment probabilities of almost two percentage points per year and of male alcohol consumption, BMI, waist circumference and consumed calories. The MSM again supports the finding of the FE model, finding very similar eﬀects in terms of size and statistical significance. For women, the MSM also support the finding of a reduction in employment probabilities, with the point estimate being somewhat larger compared to the FE model. The evidence for changes in health behaviours is less consistent across models and outcomes, with the FE indicating a reduction in waist circumference but not in BMI, and the MSM suggesting the opposite. The eﬀect sizes for changes in health behaviours in women are about half the size to those found in men.

11

Table 0.2: Time variant and invariant predictors of diabetes self-report (denominator of stabilized weights)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Males |  |  | Females |  |  |  |
|  |  |  |  |  |  |  |  |
|  | (1) | (2) |  | (3) | (4) |  |  |
|  | *β* | SE |  | *β* | SE | |  |
|  |  |  |  |  |  |  |  |
| Age (bl) | −*.*001 | .001 |  | *.*004∗∗ | .002 |  |  |
| Age squared (bl) | *.*000 | .000 |  | −*.*000∗ | .000 |  |  |
| Urbanization index (bl) | *.*000 | .000 |  | −*.*000 | .000 |  |  |
| BMI (bl) | *.*001∗∗ | .001 | *.*001 | | .001 |  |  |
| Waist circumference (cm) (bl) | *.*000 | .000 | *.*000 | | .000 |  |  |
| 3-Day Ave: Energy (kcal) (bl) | −*.*000 | .000 |  | −*.*000 | .000 |  |  |
| Smoking (bl) | *.*001 | .002 | *.*002 | | .007 |  |  |
| Any alcohol (bl) | *.*000 | .002 | *.*000 | | .003 |  |  |
| Secondary educ. (bl) | −*.*002 | .004 | *.*002 | | .004 |  |  |
| University educ. (bl) | −*.*005 | .007 |  | − |  |  |  |
| Married (bl) | −*.*003 | .004 |  | −*.*000 | .005 |  |  |
| Any medical insurance (bl) | *.*002 | .002 |  | −*.*001 | .002 |  |  |
| Employed (bl) | *.*002 | .003 | *.*002 | | .002 |  |  |
| Han ethnicity | *.*002 | .003 |  | −*.*003 | .004 |  |  |
| Rural | −*.*001 | .002 |  | −*.*005∗∗∗ | .002 |  |  |
| Household expenditures (2011 Yuan) (bl) | −*.*000 | .000 | *.*000 | | .000 |  |  |
| Survey year |  |  |  | −*.*003 |  |  |  |
| 2004 | *.*003 | .003 |  | .003 |  |  |
| 2006 | *.*001 | .003 |  | −*.*004 | .003 |  |  |
| 2009 | *.*011∗∗∗ | .004 |  | −*.*001 | .004 |  |  |
| 2011 | *.*001 | .003 | *.*001 | | .004 |  |  |
| Age | *.*004∗∗ | .002 |  | −*.*002 | .002 |  |  |
| Age squared | −*.*000∗∗∗ | .000 | *.*000 | | .000 |  |  |
| BMI | −*.*001 | .001 |  | *.*001∗ | .001 |  |  |
| Urbanization index | *.*000 | .000 | *.*000 | | .000 |  |  |
| Waist circumference (cm) | *.*000∗ | .000 |  | −*.*000 | .000 |  |  |
| 3-Day Ave: Energy (kcal) | −*.*000 | .000 | *.*000 | | .000 |  |  |
| Smoking | −*.*004∗ | .002 |  | −*.*001 | 007 |  |  |
| Any alcohol | −*.*003 | .002 |  | −*.*006 | .004 |  |  |
| Secondary education | *.*003 | .004 | *.*000 | | .004 |  |  |
| University education | *.*003 | .007 |  | − |  |  |  |
| Married | *.*000 | .005 |  | −*.*002 | .004 |  |  |
| Any medical insurance | *.*002 | .002 |  | −*.*000 | .002 |  |  |
| Employed | −*.*004∗ | .002 |  | −*.*003 | .002 |  |  |
| Household expenditures (2011 Yuan) | *.*000∗∗ | .000 |  | −*.*000 | .000 |  |  |

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01

Results for province dummies omitted to preserve space.

12

Table 0.3: Analysis of the eﬀect of a diabetes diagnosis on employment status and be-havioural outcomes using fixed eﬀects and marginal structural models

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Fixed eﬀects* | |  |  |  |
| Male sample |  | −.026 | −.087∗∗∗ | −.727∗∗∗ | −2.215∗∗∗ | −171.754∗∗∗ |  |
| Diabetes | 0.018 |  |
| Female sample | (.030) | (.032) | (.034) | (.187) | (.590) | (65.287) |  |
| −.110∗∗∗ | −.027∗∗ | −.023 | −.668∗∗ | −1.177∗ | −62.971 |  |
| Diabetes |  |
|  | (.035) | (.013) | (.018) | (.264) | (.622) | (55.960) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Marginal structural model* | |  |  |  |
| Male sample | −.008 | −.054∗ | −.132∗∗∗ | −.785∗∗∗ | −1.686∗∗∗ | −125.161∗ |  |
| Diabetes |  |
| Female sample | (.027) | (.030) | (.034) | (.188) | (.591) | (66.311) |  |
| −.122∗∗∗ | −.017∗ | −.076∗∗∗ | −.458∗ | −.559 | −60.878∗ |  |
| Diabetes |  |
|  | (.030) | (.009) | (.023) | (.251) | (.633) | (34.744) |  |

Notes: Standard errors in parentheses. Other control variables: age (only MSM), age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. Fixed eﬀects: N=23697 (male sample), N=23913 (female sample); MSM: N=13231 (male sample), N=14630 (female sample).

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01)

13

Table 0.4: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using fixed eﬀects and marginal structural models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Odds ratios |  |  |  | Beta coeﬃcients | | |  |
|  |  |  |  |  |  |  |  |  |  |
|  | (1) | (2) | (3) |  | (4) | (5) | (6) |  |  |
|  | Employment | Smoking | Any alcohol |  | BMI | Waist (cm) | Calories (kcal) | |  |
|  |  |  |  | | |  |  |  |  |
|  |  |  | *Fixed eﬀects* | | |  |  |  |  |
| Male sample | −*.*000 |  | −*.*014∗∗ |  |  | −*.*542∗∗∗ |  |  |  |
| Time since diagnosis | -.004 |  | -.155∗∗∗ | -25.595∗∗ | |  |
| Female sample | (*.*008) | (.007) | (*.*007) | (.036) | | (*.*111) | (10.297) |  |  |
| −*.*018∗∗ |  | −*.*002 |  |  | −*.*260∗∗ |  |  |  |
| Time since diagnosis | -.005∗∗ | -.087 | | -11.696 |  |  |
|  | (*.*007) | (.002) | (*.*003) | (.053) | | (*.*122) | (8.946) |  |  |
|  |  |  |  | | | |  |  |  |
|  |  |  | *Marginal structural model* | | | |  |  |  |
| Male sample | −*.*006 |  | −*.*021∗∗∗ |  |  | −*.*417∗∗∗ |  |  |  |
| Time since diagnosis | -.011∗ |  | -.151∗∗∗ | -28.974∗∗∗ | |  |
| Female sample | (*.*005) | (.006) | (*.*006) | (.037) | | (*.*103) | (10.199) |  |  |
| −*.*025∗∗∗ |  | −*.*015∗∗∗ |  |  | −*.*166 |  |  |  |
| Time since diagnosis | -.003∗ |  | -.117∗∗ | -10.991∗ | |  |
|  | (*.*006) | (.001) | (*.*006) | (.047) | | (*.*124) | (6.338) |  |  |

Notes: Standard errors in parentheses. Other control variables: age (only MSM), age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. Fixed eﬀects: N=23697 (male sample), N=23913 (female sample); MSM: N=13231 (male sample), N=14630 (female sample).

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01)

14

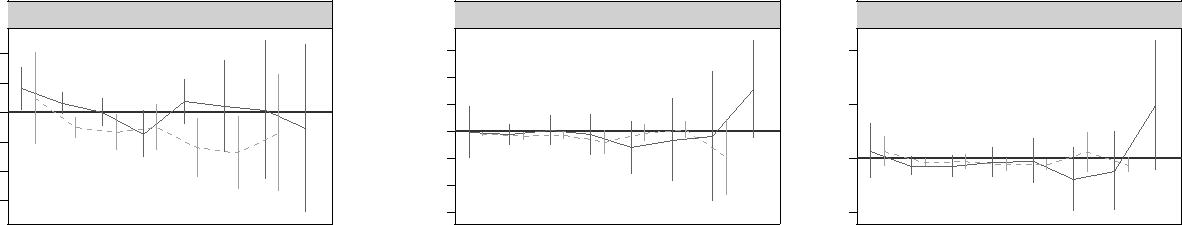
In a second step we estimated a specification using year dummies to capture the potential non-linearity in the relationship between time since diagnosis and our outcomes. The results for both estimation methods are visualized in Figure  [0.2](#page16) and presented in Tables  [0.6](#page21) and  [0.7](#page22) for the FE and MSM, respectively. Despite the smaller sample size in each group and hence lower precision, the FE model still indicates a reduction in BMI and waist circumference for men, especially in the first 8 to 10 years after diagnosis. A similar eﬀect is found for females, especially for years 3 to 8 after diagnosis. Interestingly, female employment already decreases rapidly in the first to second year after diagnosis and it does not appear that females are able to increase their employment chances later on. Using the MSM, all point estimates suggest similar eﬀects, but due to the lower sample size, we were not able to estimate the eﬀects for females on smoking and alcohol consumption.

We conducted three sensitivity analyses. First, we truncated weights at the 1st and 99th percentile to investigate the sensitivity of the MSMs to the most extreme weights. The estimated eﬀects are very similar to those using the untruncated weights (Table  [0.](#page23)8 and  [0.9),](#page24) suggesting no important loss in eﬃciency and supporting the decision to use untruncated weights. Second, we estimated all models using only covariate adjustment to investigate in how far this ’naive’ approach diverts from the "causal" estimates of the FE and MSMs. The results show that the bias is particularly strong for BMI and waist circumference, where the naive regression indicates a positive association with a diabetes diagnosis (Tables  [0.10](#page25) and  [0.11](#page26)). For the other outcomes, the results are close to or at least point into the same direction as the FE and MSMs. This suggests that for these outcomes the risk of introducing bias while using a naive regression method may be lower. Third, we estimated the FE and MSMs using the original non-imputed data. The results are broadly similar (Tables  [0.12,](#page27)  [0.13,](#page28)  [0.14](#page29) and  [0.15),](#page30) in particular for the FE model, still indicating a reduction in female employment chances and male alcohol consumption, BMI and waist circumference. The coeﬃcients of the MSM still point into the same direction as those using the imputed data, but the estimated eﬀects are smaller in size and confidence intervals are relatively large.

15

Figure 0.2: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using fixed eﬀects and marginal structural models (duration groups)

Fixed eﬀects



|  |
| --- |
| Marginal effect |

Employed

.4

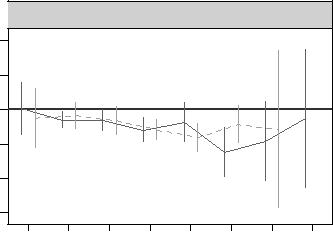
.2

0

−.2

−.4

−.6



BMI

4

2

0

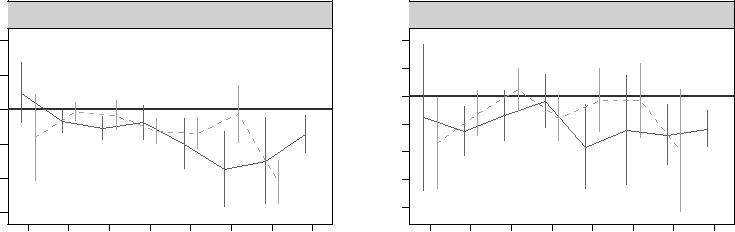
−2

−4

−6

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 |

|  |  |
| --- | --- |
| Smoking | Alcohol |
| .6 | 1 |
| .4 |  |
| .2 | .5 |
| 0 |  |
| −.2 | 0 |
| −.4 |  |
| −.6 | −.5 |
| Waist (cm) | Kcal |



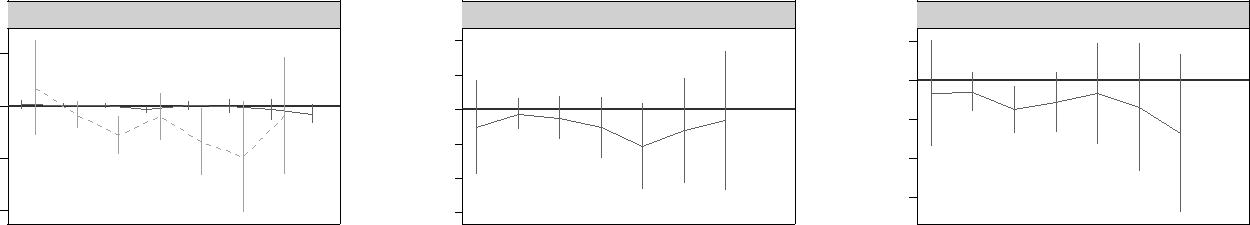
|  |  |  |
| --- | --- | --- |
| 10 | 400 |  |
| 5 | 200 |  |
|  |  |
| 0 | 0 |  |
| −200 |  |
|  |  |
| −5 | −400 |  |
|  |  |
| −10 | −600 |  |
| −15 | −800 |  |
|  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 | 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 |

Years after diagnosis

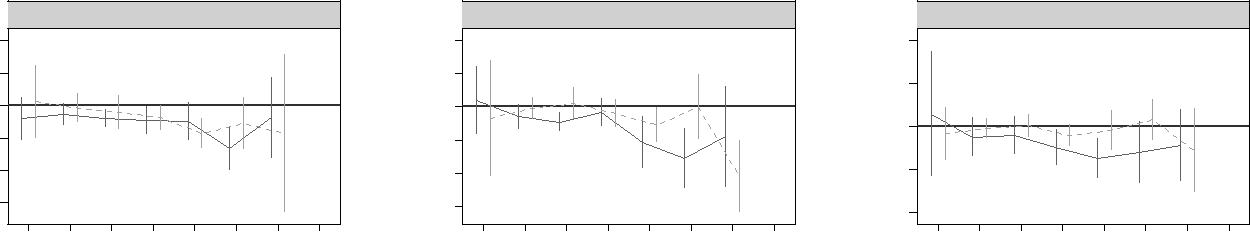
men women

Marginal structural models



|  |
| --- |
| Marginal effect |

|  |  |  |  |
| --- | --- | --- | --- |
| Employed | Smoking | Alcohol |  |
| 2 | .4 | .2 |  |
|  |  |  |
|  | .2 | 0 |  |
|  |  |  |
| 0 | 0 | −.2 |  |
|  |  |  |
| −2 | −.2 | −.4 |  |
|  |  |
|  | −.4 |  |  |
| −4 | −.6 | −.6 |  |
|  |  |
| BMI | Waist (cm) | Kcal |  |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4 |  |  |  |  |  |  |  | 10 |  |  |  |  |  |  |  | 1000 |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  | 5 |  |  |  |  |  |  |  | 500 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| −2 |  |  |  |  |  |  |  | −5 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| −4 |  |  |  |  |  |  |  | −10 |  |  |  |  |  |  |  | −500 |  |  |  |  |  |  |  |  |
| −6 |  |  |  |  |  |  |  | −15 |  |  |  |  |  |  |  | −1000 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 | 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 | 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 |  |
|  |  |  |  |  |  |  |  |  | Years after diagnosis | | | | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | men | |  |  |  | women | |  |  |  |  |  |  |  |  |  |

16

**0.4 Disussion**

Our results suggest that receiving a diabetes diagnosis in China led to a lasting reduction in male BMI and waist circumference as well as in risk behaviours such as alcohol and calorie consumption. For females, our primary results did not find as strong indications for behaviour change. However, we found a reduction in female employment probabilities, suggesting that women stopped working as a result of the diagnosis. Medical evidence suggest that sustained reductions in weight and body fat can lead to increasing insulin sensitivity, better blood glucose levels and consequently a reduced risk for diabetes related complications (Long et al.,  [2014;](#page34) Zhou et al.,  [2016](#page35)). Given our results, it appears that women in China may not have been as successful at making behaviour changes to reduce their risk.

Having both the FE models and MSM models indicating very similar results suggests that results are robust and that time invariant confounding factors may play limited role over and above baseline and time varying confounding factors.

The MSM results suggest that in particular high baseline BMI and waist circumfer-ence levels help explain selection into a diabetes diagnosis, and thus need to be taken into account. As the regression results using a naive approach have shown, not accounting for this will lead to results suggesting an increase of BMI and waist circumference after diagnosis.

**Limitations**

While we used two estimation methods to reduce the influence of selection bias due to unobserved confounding, one limitation of the used approaches is their inability to account for all forms of selection simultaneously. Therefore a causal interpretation is only possible under restrictive assumptions, namely no unobserved time-variant confounding for the FE model and positivity, exchangeability and consistency for the MSM. The assumption of positivity is likely to hold, given that every person should have at least a small chance of receiving a diabetes diagnosis. This is also supported by the relatively small range of stabilized weights and no zero-weights. Exchangeability, or no unmeasured confounding, is not testable and could potentially be violated if not all time-invariant or time-variant confounders are accounted for. This has potentially been the case which is why we also estimated an FE model. Consistency would have been violated if a diabetes diagnosis had been reported but the person had actually not been diagnosed with diabetes. This is likely only violated in very rare cases of misreporting, given that specificity of diabetes self-report is very high in China (Yuan et al.,  [2015](#page35)). Because we were interested in the eﬀect of a diabetes diagnosis, unobserved diabetes did not violate the consistency assumption.

A limitation of the FE model is the possibility of time-variant confounding causing se-

17

lection into a diabetes diagnosis based on changes in pre-treatment values of our outcomes of interest. Further, the FE model is unable to account for the eﬀect of confounders that are causally related with diabetes such as BMI or waist circumference but that may also have an eﬀect on the outcome. This may lead to an over- or underestimation of the eﬀect of diabetes if the diabetes variable captures parts of the eﬀect of very high BMI levels. This may be the reason for some of the, albeit small, diﬀerences in point estimates between the FE model and MSM. On the other hand this diﬀerence may also be due to the inability of the MSM to account for unobserved characteristics.

**Potential mechanisms**

The permanent reduction in male BMI and waist circumference we have found has also been observed in a cohort of Danish patients (De Fine Olivarius et al.,  [2015),](#page33) where weight increased the years preceding diagnosis, while after diagnosis weight decreased. The exact reasons for this decrease were unknown but attributed to motivation changes as a result of the diagnosis, concluding that time around the diagnosis may represent a window of opportunity to obtain long lasting weight change. Nonetheless, reductions in weight may also be the result of treatment initiation with metformin or other diabetes drugs that have been shown to lead to weight reductions (Yang and Weng,  [2014](#page35)). Importantly, the reduc-tion in BMI in our study was accompanied by a reduction in waist circumference. Given that in China diabetes incidence has been especially attributed to a high accumulation of visceral fat and central obesity (Ma et al.,  [2014),](#page34) reductions in waist circumference therefore may have a particular positive eﬀect on diabetes control and the prevention of comorbidities. This also allows for the interpretation that the changes in BMI are due to reductions in fat and not lean body mass(Klein et al.,  [2007](#page33)).

For women, however, we did not find similar strong evidence for reductions in BMI and waist circumference. The relatively smaller eﬀects for women could indicate a lower ability to change behaviours supportive of weight loss. A potential mechanisms may be their lower educational attainment, which has been indicated as a factor in preventing better glucose control (Luo et al.,  [2015](#page34)). Lower income levels for females may also negatively aﬀect the ability to receive adequate treatment following a diagnosis, limiting their ability to change health behaviours (Luo et al.,  [2015](#page34)). We actually found that women with diabetes tended to live in households with substantially lower expenditure levels, suggesting that lower financial resources may be a potential driver of their limited change in behaviour and lead to lower employment chances. Further, there are likely biological factors that lead to worse health outcomes for women compared to men. There is some evidence that, due to diﬀerent ways of fat storage between men and women,

18

men are more likely to cross the diabetes threshold at an earlier point in time and at a comparatively healthier metabolic state then women (Peters, Huxley, Sattar, et al.,  [2015](#page34); Peters, Huxley, and Woodward,  [2014a,b](#page34)). Women are more likely to have spend more time in a pre-diabetes stadium (Bertram and Vos,  [2010),](#page33) and the diabetes threshold is only crossed once the metabolic state of women has significantly deteriorated, leading to a greater risk of cardiovascular disease and stroke (Peters, Huxley, Sattar, et al.,  [2015](#page34)). Supporting this, a study for China found a greater prevalence of diabetes comorbidities in Chinese women than men (Liu et al.,  [2010](#page34)). In this light it might not be surprising that we find more conclusive evidence of worsening employment probabilities for women than for men. If women are less likely to receive proper treatment and to change their health behaviours and at the same time have a greater risk for complications then men, the long term eﬀects of diabetes on their health are likely more severe than for men and consequently aﬀect their employment status.

Compared to the only other study that used population level observational data to investigate the eﬀect of a diabetes diagnosis on health behaviours in the USA by Slade  [(2012),](#page35) our results paint a somewhat diﬀerent picture. While Slade finds a more lasting eﬀect on smoking cessation and alcohol consumption, but not on overweight or obesity, our results indicate lasting eﬀects on male alcohol consumption but not on smoking, and we find evidence for a sustained reduction in male body weight. Nonetheless, one has to keep in mind that we used continuous weight indicators in BMI and waist circumference, Slade investigated the eﬀect on the probability of being obese or overweight, making a direct comparison diﬃcult. Importantly—and in concordance with our findings—he finds that simple covariate adjustment leads to estimates, indicating an increase in obesity and overweight and underlining the importance of accounting for unobserved heterogeneity and treatment selection. I ALSO DID RUN REGRESSIONS USING OBESITY AND OVER-WEIGHT INDICATORS SUGGESTING THAT MEN AFTER A DIAGNOSIS HAVE A LOWER PROBABILITY OF BEING OBESE. NO EFFECT ON OVERWEIGHT. WOULD IT BE WORTHWHILE TO INCLUDE IN THE RESULTS AS WELL? HAVE NOT DONE IT NOW BECAUSE OF SMALL MISTAKE IN CALCULATING THE WEIGHTS FOR THESE OUTCOMES, SO I HAVE TO REESTIMATE THE OBESITY MODELS WHICH TAKES SOME TIME. I HAVE ADDED RESULTS FOR THE FE MODELS FOR NOW THAT SHOW ESPECIALLY ,POTENTIALLY, LASTING EF-FECTS FOR MEN TO DROP OUT OF THE OBESITY CATEGORY. THIS IS THE CHINESE OBESITY AND OVERWEIGHT THRESHOLDS OF A BMI OF 28 AND 24 RESPECTIVELY. SEE END OF APPENDIX FOR THE RESULTS

The found adverse eﬀect of self-reported diabetes on employment is in line with other

19

studies on the labour market impact of diabetes and in particular with a study from Mexico that, using FE estimates for a similar time period, found significant reductions for both males and females of about 5 percentage points (Seuring, Serneels, et al.,  [2016),](#page35) with the relative impact being much stronger for women due to their lower overall employment rates.

**0.5 Conclusion**

Our results indicate changes in male health behaviours after a diabetes diagnosis in China. These findings are robust to the application of two distinct econometric techniques. Fur-ther, women likely had to bear a larger diabetes burden also aﬀecting their economic well-being, as evidenced by their reduction in employment probabilities. Potentially, one of the causes of these adverse economic eﬀects is the lower ability of women to successfully change their behaviour as a result of the diagnosis. Further research should try to unravel the mechanisms behind these diﬀerential outcomes for men and women. Overall, given the large prevalence of undiagnosed diabetes, our results indicate that an early diagnosis can lead to early behaviour change that may lead to more positive health outcomes for people with diabetes over time. It appears, however, that more emphasis on the adequate treatment options for women may be needed to reduce their burden of diabetes.

**Appendix**

**Stabilized weights**

Table 0.5: Summary of stabilized weights

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Min | Max |
|  |  |  |  |
| Untruncated (men) | 1.001343 | .1740716 | 5.780513 |
| Untruncated (women) | 1.000773 | .1661002 | 8.754402 |
| Truncated 1 and 99 percentile (men) | .9997768 | .8906016 | 1.107517 |
| Truncated 1 and 99 percentile (women) | .9988097 | .8323757 | 1.119154 |

**0.5.1 Duration groups results**

20

Table 0.6: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using fixed eﬀects (duration groups)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Odds ratios |  |  |  | Beta coeﬃcients | | |  |
|  |  |  |  |  |  |  |  |  |  |
|  | (1) | (2) | (3) |  | (4) | (5) | (6) |  |  |
|  | Employment | Smoking | Any alcohol |  | BMI | Waist (cm) | Calories (kcal) | |  |
|  |  |  |  |  |  |  |  |  |  |
| Male sample |  |  |  |  |  |  |  |  |  |
| 0 | *.*163∗∗ | -.006 | *.*072 | .049 | | 2*.*397 | -155.323 |  |  |
|  | (*.*075) | (.098) | (*.*132) | (.792) | | (2*.*258) | (268.552) |  |  |
| 1-2 | *.*062 | -.023 | −*.*067 |  | -.605∗∗ | −1*.*782∗∗ | -254.427∗∗∗ | |  |
|  | (*.*039) | (.040) | (*.*045) | (.252) | | (*.*825) | (92.421) |  |  |
| 3-4 | *.*002 | .008 | −*.*068 |  | -.639∗∗ | −2*.*768∗∗∗ | -139.397 |  |  |
|  | (*.*050) | (.055) | (*.*052) | (.314) | | (*.*870) | (93.516) |  |  |
| 5-6 | −*.*145∗ | -.023 | −*.*030 |  | -1.183∗∗∗ | −1*.*901 | -39.053 |  |  |
|  | (*.*081) | (.077) | (*.*071) | (.362) | | (1*.*304) | (98.730) |  |  |
| 7-8 | *.*074 | -.121 | −*.*021 | -.751 | | −5*.*037∗∗∗ | -367.421∗∗ | |  |
|  | (*.*079) | (.100) | (*.*106) | (.588) | | (1*.*896) | (156.656) |  |  |
| 9-10 | *.*040 | -.062 | −*.*188 |  | -2.472∗∗∗ | −8*.*642∗∗∗ | -244.372 |  |  |
|  | (*.*159) | (.157) | (*.*152) | (.741) | | (2*.*812) | (201.960) |  |  |
| 11-12 | *.*017 | -.037 | −*.*114 | -1.835 | | −7*.*465∗∗ | -281.008∗∗ | |  |
|  | (*.*240) | (.245) | (*.*187) | (1.181) | | (3*.*238) | (112.461) |  |  |
| 13-14 | −*.*108 | .308∗ | *.*494 | -.524 | | −3*.*615∗∗∗ | -235.985∗∗∗ | |  |
|  | (*.*293) | (.184) | (*.*306) | (2.059) | | (1*.*383) | (67.961) |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Female sample |  |  |  |  |  | −4*.*090 |  |  |  |
| 0 | *.*098 | -.019∗∗ | *.*067 | -.505 | | -340.952∗∗ | |  |
|  | (*.*159) | (.008) | (*.*072) | (.896) | | (3*.*235) | (169.411) |  |  |
| 1-2 | −*.*103∗∗∗ | -.034∗∗ | −*.*036∗ | -.364 | | −*.*355 | -124.847 |  |  |
|  | (*.*037) | (.015) | (*.*020) | (.404) | | (*.*717) | (85.402) |  |  |
| 3-4 | −*.*134∗∗ | -.028∗ | −*.*025 | -.648 | | −*.*896 | 45.970 |  |  |
|  | (*.*062) | (.016) | (*.*036) | (.435) | | (1*.*120) | (79.063) |  |  |
| 5-6 | −*.*100 | -.079∗ | −*.*056∗ |  | -1.184∗∗∗ | −3*.*176∗∗∗ | -160.445∗ | |  |
|  | (*.*080) | (.045) | (*.*033) | (.323) | | (*.*990) | (83.164) |  |  |
| 7-8 | −*.*240∗∗ | -.013 | −*.*054∗∗ |  | -1.637∗∗∗ | −3*.*482∗∗∗ | -32.865 |  |  |
|  | (*.*103) | (.032) | (*.*026) | (.433) | | (1*.*182) | (117.750) |  |  |
| 9-10 | −*.*274∗∗ | .012 | *.*062 | -.876 | | −*.*639 | -30.676 |  |  |
|  | (*.*128) | (.031) | (*.*094) | (.568) | | (2*.*147) | (137.658) |  |  |
| 11-12 | −*.*140 | -.189 | −*.*059 | -1.163 | | −10*.*633∗∗∗ | -393.884∗ | |  |
|  | (*.*202) | (.143) | (*.*036) | (2.344) | | (1*.*618) | (227.502) |  |  |

Notes: Other control variables: age squared, region, urban, education, han, marital status, urbaniza-tion index, time dummies, health insurance status, household expenditures. N=13195 (male sample), N=14549 (female sample).

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01)

21

Table 0.7: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using marginal structural models (duration groups)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  |  |  |  |  |
| Male sample |  | −.102 | −.067 | −.811 |  |  |  |
| 0 | 0.049 | 0.968 | 148.433 |  |
|  | (.087) | (.139) | (.138) | (.682) | (2.587) | (369.013) |  |
| 1-2 | 0.034 | −.028 | −.059 | −.535 | −1.517 | −121.424 |  |
|  | (.038) | (.046) | (.051) | (.340) | (.996) | (115.615) |  |
| 3-4 | 0.012 | −.050 | −.150∗∗ | −.795∗∗∗ | −2.312∗∗∗ | −96.393 |  |
|  | (.042) | (.063) | (.060) | (.288) | (.712) | (113.071) |  |
| 5-6 | −.139∗∗ | −.105 | −.112 | −.937∗∗ | −.843 | −237.957∗∗ |  |
|  | (.068) | (.090) | (.078) | (.438) | (1.050) | (107.745) |  |
| 7-8 | 0.010 | −.213∗ | −.067 | −.995∗ | −5.401∗∗∗ | −368.463∗∗∗ |  |
|  | (.087) | (.127) | (.132) | (.602) | (1.991) | (117.513) |  |
| 9-10 | −.014 | −.124 | −.139 | −2.648∗∗∗ | −7.791∗∗∗ | −300.917 |  |
|  | (.136) | (.155) | (.167) | (.699) | (2.294) | (184.269) |  |
| 11-12 | −.142 | −.065 | −.272 | −.751 | −4.542 | −218.570 |  |
|  | (.203) | (.205) | (.206) | (1.262) | (3.896) | (212.524) |  |
| 13-14 | −.304 |  |  |  |  |  |  |
|  | (.186) |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Female sample |  |  |  |  | −1.713 | −82.792 |  |
| 0 | 0.110 |  |  | 0.233 |  |
|  | (.132) |  |  | (1.158) | (4.483) | (158.560) |  |
| 1-2 | −.063 |  |  | −.153 | −.238 | −17.870 |  |
|  | (.049) |  |  | (.450) | (.850) | (59.249) |  |
| 3-4 | −.210∗∗∗ |  |  | −.409 | 0.506 | 11.054 |  |
|  | (.071) |  |  | (.537) | (1.229) | (67.419) |  |
| 5-6 | −.074 |  |  | −.743∗ | −1.038 | −104.696 |  |
|  | (.087) |  |  | (.396) | (1.063) | (64.981) |  |
| 7-8 | −.262∗∗ |  |  | −1.721∗∗∗ | −2.724∗∗ | −41.099 |  |
|  | (.122) |  |  | (.464) | (1.315) | (119.484) |  |
| 9-10 | −.355∗∗ |  |  | −1.095 | −.012 | 80.030 |  |
|  | (.172) |  |  | (.827) | (2.495) | (122.037) |  |
| 11-12 | −.072 |  |  | −1.739 | −10.515∗∗∗ | −275.508 |  |
|  | (.213) |  |  | (2.489) | (2.806) | (250.141) |  |

Notes: Other control variables: age, age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=13195 (male sample), N=14549 (female sample).

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01)

22

**Robustness checks**

**MSMs using truncated weights**

Table 0.8: Analysis of the eﬀect of a diabetes diagnosis on employment status and be-havioural outcomes using marginal structural models with truncated stabilized weights at 1st and 99th percentile

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Self-reported diabetes* | |  |  |  |
| Male sample | −.026 | −.053∗ | −.132∗∗∗ | −.707∗∗∗ | −1.426∗∗ | −200.750∗∗∗ |  |
| Self-reported diabetes |  |
| Female sample | (.025) | (.031) | (.034) | (.188) | (.560) | (54.156) |  |
| −.138∗∗∗ | −.017∗ | −.076∗∗∗ | −.243 |  | −61.998∗ |  |
| Self-reported diabetes | 0.058 |  |
|  | (.029) | (.009) | (.023) | (.252) | (.640) | (35.849) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Years since diagnosis* | |  |  |  |
| Male sample | −.009∗ | −.011∗ | −.021∗∗∗ | −.145∗∗∗ | −.377∗∗∗ | −34.287∗∗∗ |  |
| Time since diagnosis |  |
| Female sample | (.005) | (.006) | (.006) | (.037) | (.102) | (9.612) |  |
| −.027∗∗∗ | −.003∗ | −.015∗∗∗ | −.093∗∗ | −.086 | −10.750∗ |  |
| Time since diagnosis |  |
|  | (.007) | (.001) | (.006) | (.047) | (.128) | (6.455) |  |

Notes: Standard errors in parentheses. Other control variables: age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=13231 (male sample), N=14630 (female sample).

23

Table 0.9: Eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using MSM with truncated stabilized weights (1st and 99th pct; im-puted)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  |  |  |  |  |
| Male sample |  | −.102 | −.067 | −.345 |  | −86.368 |  |
| 0 | 0.075 | 2.302 |  |
|  | (.077) | (.139) | (.138) | (.839) | (2.872) | (247.229) |  |
| 1-2 | 0.012 | −.028 | −.059 | −.375 | −.812 | −224.716∗∗ |  |
|  | (.036) | (.046) | (.051) | (.297) | (.957) | (95.028) |  |
| 3-4 | −.018 | −.050 | −.150∗∗ | −.723∗∗ | −1.965∗∗∗ | −167.769∗ |  |
|  | (.042) | (.063) | (.060) | (.305) | (.735) | (86.996) |  |
| 5-6 | −.139∗∗ | −.105 | −.112 | −1.034∗∗∗ | −.891 | −203.814∗∗ |  |
|  | (.066) | (.090) | (.078) | (.392) | (1.096) | (102.051) |  |
| 7-8 | 0.018 | −.213∗ | −.067 | −.944∗ | −4.935∗∗∗ | −347.441∗∗∗ |  |
|  | (.084) | (.127) | (.132) | (.563) | (1.846) | (125.660) |  |
| 9-10 | −.033 | −.124 | −.139 | −2.627∗∗∗ | −6.998∗∗∗ | −232.656 |  |
|  | (.122) | (.155) | (.167) | (.689) | (2.453) | (177.019) |  |
| 11-12 | −.151 | −.065 | −.272 | −.889 | −4.363 | −212.318 |  |
|  | (.202) | (.205) | (.206) | (1.123) | (3.500) | (189.967) |  |
| 13-14 | −.305∗ |  |  |  |  |  |  |
|  | (.185) |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Female sample |  |  |  |  | −.060 | −104.652 |  |
| 0 | 0.112 |  |  | 0.685 |  |
|  | (.128) |  |  | (1.289) | (4.993) | (158.546) |  |
| 1-2 | −.104∗∗ |  |  | 0.205 | 0.698 | −18.334 |  |
|  | (.047) |  |  | (.452) | (.834) | (64.456) |  |
| 3-4 | −.208∗∗∗ |  |  | −.299 | 0.889 | 6.982 |  |
|  | (.068) |  |  | (.520) | (1.191) | (65.213) |  |
| 5-6 | −.138∗ |  |  | −.569 | −.793 | −122.064∗ |  |
|  | (.082) |  |  | (.364) | (1.087) | (64.349) |  |
| 7-8 | −.249∗∗ |  |  | −1.662∗∗∗ | −2.125 | −5.039 |  |
|  | (.121) |  |  | (.502) | (1.315) | (120.139) |  |
| 9-10 | −.373∗∗ |  |  | −1.019 | 0.409 | 72.475 |  |
|  | (.169) |  |  | (.829) | (2.523) | (132.027) |  |
| 11-12 | −.087 |  |  | −1.858 | −10.814∗∗∗ | −251.975 |  |
|  | (.214) |  |  | (2.510) | (2.749) | (234.610) |  |

Notes: Standard errors in parentheses. Other control variables: age squared, region, urban, education, Han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=13195 (male sample), N=14549 (female sample).

24

**Only covariate adjustment**

Table 0.10: Analysis of the eﬀect of a diabetes diagnosis on employment status and be-havioural outcomes only using covariate adjustment

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Self-reported diabetes* | |  |  |  |
| Male sample |  | −.074∗∗ | −.119∗∗∗ |  |  | −165.647∗∗∗ |  |
| Self-reported diabetes | 0.006 | 0.925∗∗∗ | 2.964∗∗∗ |  |
| Female sample | (.023) | (.036) | (.033) | (.289) | (.817) | (50.101) |  |
| −.100∗∗∗ | −.007 | −.072∗∗∗ |  |  | −33.807 |  |
| Self-reported diabetes | 1.805∗∗∗ | 4.619∗∗∗ |  |
|  | (.029) | (.011) | (.023) | (.354) | (.877) | (33.783) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Years since diagnosis* | |  |  |  |
| Male sample | −.002 | −.052∗ | −.017∗∗ |  |  | −27.700∗∗∗ |  |
| Time since diagnosis | 0.107∗ | 0.318∗∗ |  |
| Female sample | (.005) | (.031) | (.007) | (.058) | (.161) | (9.297) |  |
| −.019∗∗∗ | −.030 | −.014∗∗ |  |  | −4.186 |  |
| Time since diagnosis | 0.245∗∗∗ | 0.645∗∗∗ |  |
|  | (.007) | (.079) | (.006) | (.067) | (.178) | (6.113) |  |

Notes: Standard errors in parentheses. Other control variables: Age, age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=23697 (male sample), N=23913 (female sample).

25

Table 0.11: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using covariate adjustment (duration groups) (im-puted)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  |  |  |  |  |
| Male sample |  | −.095 | −.045 |  |  | −7.725 |  |
| 0 | 0.067 | 1.763∗∗ | 7.505∗∗∗ |  |
|  | (.060) | (.136) | (.136) | (.737) | (2.581) | (222.633) |  |
| 1-2 | 0.027 | −.087∗ | −.104∗∗ | 1.247∗∗∗ | 3.518∗∗∗ | −214.457∗∗∗ |  |
|  | (.027) | (.049) | (.047) | (.294) | (.944) | (78.407) |  |
| 3-4 | 0.002 | −.019 | −.146∗∗ | 1.406∗∗∗ | 3.208∗∗∗ | −127.663 |  |
|  | (.039) | (.060) | (.058) | (.435) | (1.131) | (86.377) |  |
| 5-6 | −.078 | −.114 | −.132 | 0.857 | 4.054∗∗∗ | −146.275 |  |
|  | (.058) | (.084) | (.084) | (.619) | (1.425) | (101.505) |  |
| 7-8 | 0.075 | −.291∗∗ | 0.001 | 0.308 | −.579 | −326.239∗∗ |  |
|  | (.053) | (.126) | (.129) | (.532) | (2.014) | (129.013) |  |
| 9-10 | 0.082 | −.025 | −.037 | −.603 | −1.740 | −144.106 |  |
|  | (.083) | (.138) | (.149) | (.959) | (2.495) | (171.510) |  |
| 11-12 | −.151 |  | −.175 | −1.601 | −6.070 | −140.067 |  |
|  | (.230) |  | (.220) | (1.094) | (4.376) | (195.607) |  |
| 13-14 | −.246 |  |  | 0.265 | −1.131 | −34.363 |  |
|  | (.152) |  |  | (2.421) | (4.340) | (176.819) |  |
|  |  |  |  |  |  |  |  |
| Female sample |  |  |  |  |  | −83.493 |  |
| 0 | 0.083 | 0.025 | 0.030 | 3.742∗∗∗ | 8.177∗ |  |
|  | (.115) | (.067) | (.107) | (.995) | (4.255) | (139.371) |  |
| 1-2 | −.101∗∗ | −.014 | −.056∗∗∗ | 2.096∗∗∗ | 5.256∗∗∗ | −23.578 |  |
|  | (.044) | (.009) | (.018) | (.455) | (.943) | (59.415) |  |
| 3-4 | −.132∗∗ | −.018 | −.059∗∗ | 1.833∗∗∗ | 5.405∗∗∗ | 57.985 |  |
|  | (.066) | (.012) | (.023) | (.544) | (1.432) | (63.186) |  |
| 5-6 | −.071 | −.014 | −.059∗ | 1.416∗∗∗ | 3.092∗∗ | −77.020 |  |
|  | (.079) | (.015) | (.032) | (.485) | (1.519) | (64.878) |  |
| 7-8 | −.132 | 0.013 |  | 0.536 | 2.251 | 69.285 |  |
|  | (.125) | (.030) |  | (.851) | (1.878) | (116.795) |  |
| 9-10 | −.223 | 0.012 |  | 1.453 | 5.297∗ | 151.466 |  |
|  | (.194) | (.030) |  | (1.357) | (2.988) | (127.461) |  |
| 11-12 | 0.007 |  |  | −.352 | −7.653∗∗∗ | −162.492 |  |
|  | (.230) |  |  | (1.479) | (2.436) | (213.782) |  |

Notes: Standard errors in parentheses. Other control variables: age, age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=23661 (male sample), N=23830 (female sample).

26

**Results using non-imputed data**

Table 0.12: Analysis of the eﬀect of a diabetes diagnosis on employment status and be-havioural outcomes using fixed eﬀects and marginal structural models (no imputation)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Fixed eﬀects* | |  |  |  |
| Male sample |  | −.000 | −.063∗ | −.868∗∗∗ | −2.448∗∗∗ | −152.027∗∗ |  |
| Diabetes | 0.028 |  |
| Female sample | (.030) | (.033) | (.034) | (.170) | (.517) | (68.421) |  |
| −.107∗∗∗ | −.024∗∗ | −.022 | −.638∗∗ | −1.049∗ | −81.554∗ |  |
| Diabetes |  |
|  | (.034) | (.012) | (.019) | (.288) | (.636) | (48.970) |  |
|  |  |  |  | |  |  |  |
|  |  |  | *Marginal structural model* | |  |  |  |
| Male sample |  | −.042 | −.112∗∗ | −.641∗∗∗ | −1.224 | −152.149 |  |
| Diabetes | 0.050 |  |
| Female sample | (.046) | (.041) | (.049) | (.220) | (.803) | (123.921) |  |
| −.081∗ | −.028∗ | −.075 | −.583 | −.935 | −50.350 |  |
| Diabetes |  |
|  | (.049) | (.015) | (.045) | (.393) | (.866) | (56.914) |  |

Notes: Standard errors in parentheses. Other control variables: age (only MSM), age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. FE: N=21444 (male sample), N=23128 (female sample), MSM: N=10039 (male sample), N=11489 (female sample).

27

Table 0.13: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using fixed eﬀects and marginal structural models (non-imputed)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | | (3) | | (4) | | (5) | |  | (6) |  |
|  | Employment | Smoking | | Any alcohol | | BMI | | Waist (cm) | | Calories (kcal) | |  |
|  |  |  |  |  |  | | |  |  |  |  |  |
|  |  |  |  |  | *Fixed eﬀects* | | |  |  |  |  |  |
| Male sample | .003 | 0.005 | |  | .003 |  | .164∗∗∗ |  | .581∗∗∗ |  | 22.700 |  |
| Time since diagnosis | − | − | − | − |  |
|  | − |  |  |  |  |  |  |  |
| Female sample | (.008) | (.007) | | (.007) | | (.044) | | (.121) | | (12.487) | |  |
| .026∗∗ |  | .006∗ |  | .005 |  | .163∗∗ |  | .325∗ |  | 16.888 |  |
| Time since diagnosis | − | − | − | − | − |  |
|  | − |  |  |  |  |  |  |
|  | (.009) | (.002) | | (.004) | | (.060) | | (.148) | | (10.337) | |  |
|  |  |  |  |  | | | |  |  |  |  |  |
|  |  |  |  | *Marginal structural model* | | | |  |  |  |  |  |
| Male sample |  | −.020 | | −.024 | | −.221∗∗ | | −.705∗∗ | | −80.594∗∗ | |  |
| Time since diagnosis | 0.027 |  |
| Female sample | (.018) | (.016) | | (.017) | | (.087) | | (.295) | | (39.681) | |  |
| .030 |  | .009 |  | .039 |  | .413∗ |  | .598 |  | 14.552 |  |
| Time since diagnosis | − | − | − | − | − |  |
|  | − |  |  |  |  |  |  |
|  | (.022) | (.006) | | (.025) | | (.216) | | (.376) | | (24.269) | |  |

Notes: Standard errors in parentheses. Other control variables: age (only MSM) age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expen-ditures. FE: N=22066 (male sample), N=23051 (female sample), MSM: N=10028 (male sample), N=11465 (female sample).

28

Table 0.14: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using fixed eﬀects (duration groups) (non-imputed)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  |  |  |  |  |
| Male sample |  | −.014 |  | −.009 |  | −268.692 |  |
| 0 | 0.132 | 0.125 | 1.439 |  |
|  | (.072) | (.085) | (.136) | (.702) | (1.879) | (215.527) |  |
| 1-2 | 0.065 | −.013 | −.053 | −.849∗∗∗ | −2.391∗∗∗ | −244.075∗∗ |  |
|  | (.039) | (.041) | (.045) | (.212) | (.663) | (91.888) |  |
| 3-4 | 0.009 | 0.058 | −.035 | −.720∗ | −2.642∗∗ | −105.039 |  |
|  | (.048) | (.054) | (.053) | (.334) | (.824) | (100.459) |  |
| 5-6 | −.127 | 0.029 | −.006 | −1.168∗∗∗ | −1.733 | −14.828 |  |
|  | (.080) | (.079) | (.078) | (.351) | (1.204) | (104.926) |  |
| 7-8 | 0.118 | −.089 | 0.030 | −.617 | −4.227∗ | −319.563 |  |
|  | (.081) | (.090) | (.087) | (.537) | (1.871) | (166.278) |  |
| 9-10 | 0.018 | 0.030 | −.115 | −2.199∗∗∗ | −9.351∗∗∗ | −189.090 |  |
|  | (.154) | (.132) | (.145) | (.659) | (2.503) | (224.518) |  |
| 11-12 | −.032 | 0.097 | 0.058 | −2.060 | −9.185∗∗ | −162.422 |  |
|  | (.250) | (.200) | (.097) | (1.074) | (2.868) | (94.541) |  |
| 13-14 | −.116 | 0.341∗ | 0.599 | −.527 | −3.481∗∗∗ | −146.776∗ |  |
|  | (.312) | (.153) | (.311) | (1.959) | (.864) | (68.253) |  |
|  |  |  |  |  |  |  |  |
| Female sample |  | −.019∗ |  | −.811 | −4.079 | −372.137∗ |  |
| 0 | 0.117 | 0.066 |  |
|  | (.154) | (.008) | (.072) | (.912) | (3.403) | (171.871) |  |
| 1-2 | −.091∗ | −.032∗ | −.037 | −.324 | −.383 | −133.764∗ |  |
|  | (.037) | (.014) | (.021) | (.417) | (.681) | (61.475) |  |
| 3-4 | −.144∗ | −.024 | −.025 | −.857 | −1.051 | 26.559 |  |
|  | (.061) | (.014) | (.039) | (.458) | (1.184) | (85.752) |  |
| 5-6 | −.118 | −.080 | −.062∗ | −1.243∗∗∗ | −2.531∗∗ | −190.484∗ |  |
|  | (.080) | (.046) | (.031) | (.348) | (.958) | (94.117) |  |
| 7-8 | −.255∗ | 0.004 | −.050 | −1.823∗∗∗ | −3.976∗∗ | −47.473 |  |
|  | (.102) | (.020) | (.026) | (.454) | (1.217) | (133.625) |  |
| 9-10 | −.249∗ | 0.016 | 0.064 | −1.121 | −.803 | −89.938 |  |
|  | (.125) | (.033) | (.102) | (.653) | (2.277) | (136.611) |  |
| 11-12 | −.197 | −.224 | −.069 | −.878 | −10.898∗∗∗ | −563.138∗ |  |
|  | (.170) | (.172) | (.041) | (2.888) | (1.724) | (240.085) |  |

Notes: Standard errors in parentheses. Other control variables: age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=22066 (male sample), N=23051 (female sample).

29

Table 0.15: Analysis of the eﬀect of time since diabetes diagnosis on employment status and behavioural outcomes using marginal structural models (duration groups) (non-imputed)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |  |
|  | Employment | Smoking | Any alcohol | BMI | Waist (cm) | Calories (kcal) |  |
|  |  |  |  |  |  |  |  |
| Male sample |  |  | −.149 | −1.004∗ |  |  |  |
| 0 | 0.107 | 0.051 | 1.113 | 314.866 |  |
|  | (.080) | (.150) | (.175) | (.562) | (1.103) | (409.370) |  |
| 1-2 | 0.038 | −.050 | −.050 | −.609∗∗ | −1.727∗ | −146.884 |  |
|  | (.047) | (.052) | (.057) | (.284) | (.973) | (148.251) |  |
| 3-4 | 0.000 | −.038 | −.095 | −1.028∗∗ | −3.164 | −695.698∗∗∗ |  |
|  | (.) | (.159) | (.117) | (.460) | (2.066) | (190.065) |  |
|  |  |  |  |  |  |  |  |
| Female sample |  |  |  | −.102 | −1.362 | −111.379 |  |
| 0 | 0.134 | 0.000 | 0.000 |  |
|  | (.179) | (.) | (.) | (1.544) | (6.099) | (207.839) |  |
| 1-2 | −.079 | −.019∗∗ | −.052∗ | −.484 | −.581 | −32.105 |  |
|  | (.069) | (.008) | (.029) | (.487) | (1.006) | (67.670) |  |
| 3-4 | 0.000 | 0.000 | 0.000 | −5.590∗ | −8.485∗∗∗ | 1.258 |  |
|  | (.) | (.) | (.) | (3.286) | (1.792) | (258.264) |  |

Notes: Due to Standard errors in parentheses. Other control variables: Age, age squared, region, ur-ban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=10028 (male sample), N=11465 (female sample).

30

**Overweight and obesity results (only FE so far)**

Table 0.16: Analysis of the eﬀect of a diabetes diagnosis on overweight and obesity using FE models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Males |  |  |  | Females | |
|  |  |  |  |  |  |  |
|  | (1) | (2) |  | (3) | (4) |  |
|  | Overweight | Obese |  | Overweight | Obese | |
|  |  |  |  |  |  | |
| Self-reported diabetes | −.029 | −.052∗∗ |  | −.085∗∗ | −.044 | |
|  | (.035) | (.025) | (.037) | | (.028) |  |

Standard errors in parentheses

Other control variables: age squared, region, urban, education, han, marital status, urbanicity index, time dummies, health insurance status

∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01

Notes: Standard errors in parentheses. Other control variables: Age, age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=19897 (male sample), N=21592 (female sample).

Table 0.17: Analysis of the eﬀect of time since diagnosis on overweight and obesity using FE models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Males |  |  |  | Females | |
|  |  |  |  |  |  |  |
|  | (1) | (2) |  | (3) | (4) |  |
|  | Overweight | Obese |  | Overweight | Obese | |
|  |  |  |  |  |  | |
| Self-reported diabetes | −.009 | −.008∗∗ |  | −.009 | −.012∗∗ | |
|  | (.008) | (.004) | (.007) | | (.006) |  |

Standard errors in parentheses

Other control variables: age squared, region, urban, education, han, marital status, urbanicity index, time dummies, health insurance status

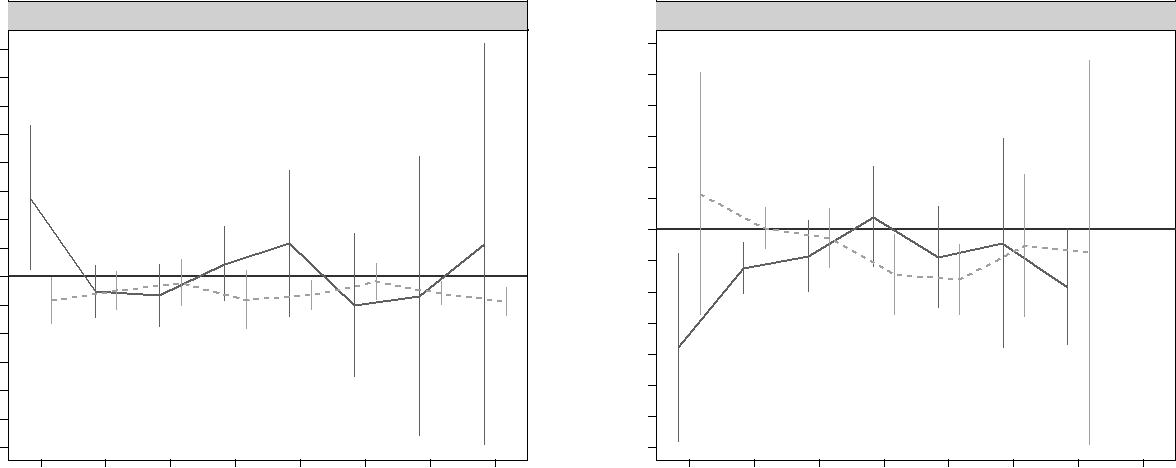
∗ *p <* 0*.*10,∗∗ *p <* 0*.*05,∗∗∗ *p <* 0*.*01

Notes: Standard errors in parentheses. Other control variables: Age, age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=23697 (male sample), N=23913 (female sample).

31

Figure 0.3: Analysis of the eﬀect of time since diabetes diagnosis on overweight and obesity using fixed eﬀects (duration groups)

Fixed eﬀects



|  |
| --- |
| Marginal effect |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Males | |  |  |  |  |  |  | Females | |  |  |  |  |
| .8 |  |  |  |  |  |  |  | .6 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| .7 |  |  |  |  |  |  |  | .5 |  |  |  |  |  |  |  |  |
| .6 |  |  |  |  |  |  |  | .4 |  |  |  |  |  |  |  |  |
| .5 |  |  |  |  |  |  |  | .3 |  |  |  |  |  |  |  |  |
| .4 |  |  |  |  |  |  |  | .2 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| .3 |  |  |  |  |  |  |  | .1 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| .2 |  |  |  |  |  |  |  | 0 |  |  |  |  |  |  |  |  |
| .1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | −.1 |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | −.2 |  |  |  |  |  |  |  |  |
| −.1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | −.3 |  |  |  |  |  |  |  |  |
| −.2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| −.3 |  |  |  |  |  |  |  | −.4 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| −.4 |  |  |  |  |  |  |  | −.5 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| −.5 |  |  |  |  |  |  |  | −.6 |  |  |  |  |  |  |  |  |
| −.6 |  |  |  |  |  |  |  | −.7 |  |  |  |  |  |  |  |  |
| 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 | 0 | 1−2 | 3−4 | 5−6 | 7−8 | 9−10 | 11−12 | 13−14 |  |
|  |  |  |  |  |  | Years after diagnosis | | |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | overweight | | obese |  |  |  |  |  |  |  |  |

32

**Bibliography**

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist’s Com-panion*. Princeton University Press.

Attard, S. M., Herring, a. H., Mayer-Davis, E. J., Popkin, B. M., Meigs, J. B., and Gordon-Larsen, P. (2012). “Multilevel examination of diabetes in modernising China: what ele-ments of urbanisation are most associated with diabetes?” *Diabetologia* 55 (12), 3182– 92.

Batis, C., Mendez, M. a., Sotres-Alvarez, D., Gordon-Larsen, P., and Popkin, B. (2014). “Dietary pattern trajectories during 15 years of follow-up and HbA1c, insulin resistance and diabetes prevalence among Chinese adults.” *Journal of epidemiology and community* *health* 29 (6), 773–779. arXiv:[NIHMS15000](http://arxiv.org/abs/NIHMS150003)3.

Bertram, M. Y. and Vos, T. (2010). “Quantifying the duration of pre-diabetes.” *Australian* *and New Zealand journal of public health* 34 (3), 311–314.

Chan, J. C. N., Zhang, Y., and Ning, G. (2014). “Diabetes in China: a societal solution for a personal challenge.” *The Lancet Diabetes & Endocrinology* 2 (12), 969–979.

Cole, S. R. and Hernan, M. A. (2008). “Constructing Inverse Probability Weights for Marginal Structural Models.” *American Journal of Epidemiology* 168 (6), 656–664.

De Fine Olivarius, N., Siersma, V. D., Køster-Rasmussen, R., Heitmann, B. L., and Wal-dorﬀ, F. B. (2015). “Weight changes following the diagnosis of type 2 diabetes: The impact of recent and past weight history before diagnosis. Results from the Danish Diabetes Care in General Practice (DCGP) Study.” *PLoS ONE* 10 (4), 1–14.

Hu, F. B. (2011). “Globalization of diabetes: the role of diet, lifestyle, and genes.” *Diabetes* *care* 34 (6), 1249–57.

Klein, D. (2014). *MIMRGNS: Stata module to run margins after mi estimate*. Statistical Software Components, Boston College Department of Economics.

Klein, S., Allison, D. B., Heymsfield, S. B., Kelley, D. E., Leibel, R. L., Nonas, C., and Kahn, R. (2007). “Waist circumference and cardiometabolic risk: A consensus state-ment from Shaping America’s Health: Association for Weight Management and Obesity Prevention; NAASO, the Obesity Society; the American Society for Nutrition; and the American Diabetes Associat.” *Diabetes Care* 30 (6), 1647–1652.

33

Liu, Z., Fu, C., Wang, W., and Xu, B. (2010). “Prevalence of chronic complications of type 2 diabetes mellitus in outpatients - a cross-sectional hospital based survey in urban China.” *Health and quality of life outcomes* 8, 62.

Long, G. H., Cooper, A. J., Wareham, N. J., Griﬃn, S. J., and Simmons, R. K. (2014). “Healthy Behavior Change and Cardiovascular Outcomes in Newly Diagnosed Type 2 Diabetic Patients: A Cohort Analysis of the ADDITION-Cambridge Study.” *Diabetes* *Care* 37 (6), 1712–1720.

Luo, X., Liu, T., Yuan, X., Ge, S., Yang, J., Li, C., and Sun, W. (2015). “Factors Influencing Self-Management in Chinese Adults with Type 2 Diabetes: A Systematic Review and Meta-Analysis.” *International Journal of Environmental Research and Public Health*

12 (9), 11304–11327.

Ma, R. C. W., Lin, X., and Jia, W. (2014). “Causes of type 2 diabetes in China.” *The* *Lancet Diabetes & Endocrinology* 2 (12), 980–991.

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

Pan, X. (2015). “Gender Dissimilarity in Type 2 Diabetes Risk Factors: a Chinese Study.”

*International Journal of Behavioral Medicine* 22 (5), 614–624.

Peters, S. A. E., Huxley, R. R., Sattar, N., and Woodward, M. (2015). “Sex Diﬀerences in the Excess Risk of Cardiovascular Diseases Associated with Type 2 Diabetes: Potential Explanations and Clinical Implications.” *Current Cardiovascular Risk Reports* 9 (7), 1– 7.

Peters, S. A. E., Huxley, R. R., and Woodward, M. (2014a). “Diabetes as a risk factor for stroke in women compared with men: A systematic review and meta-analysis of 64 cohorts, including 775 385 individuals and 12 539 strokes.” *The Lancet* 383 (9933), 1973–1980.

– (2014b). “Diabetes as risk factor for incident coronary heart disease in women com-pared with men: A systematic review and meta-analysis of 64 cohorts including 858,507

individuals and 28,203 coronary events.” *Diabetologia* 57 (8), 1542–1551.

Popkin, B. M., Du, S., Zhai, F., and Zhang, B. (2010). “Cohort profile: The China Health and Nutrition Survey-monitoring and understanding socio-economic and health change in China, 1989-2011.” *International Journal of Epidemiology* 39 (6), 1435–1440.

Robins, J. M., Hernan, M. Á., and Brumback, B. (2000). “Marginal Structural Models and Causal Inference in Epidemiology.” *Epidemiology* 11, 550–560.

34

Royston, P. and White, I. (2009). “Multiple imputation by chained equations (MICE): Implementation in Stata.” *Journal of Statistical Software* 45 (4), pages. arXiv:  [arXiv](http://arxiv.org/abs/arXiv:1501.0228):

[1501.022](http://arxiv.org/abs/arXiv:1501.0228)8.

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., Serneels, P., and Suhrcke, M. (2016). “The Impact of Diabetes on Labor Market Outcomes in Mexico: A Panel Data and Biomarker Analysis.”

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

Textor, J., Hardt, J., and Knüppel, S. (2011). “DAGitty: a graphical tool for analyzing causal diagrams.” *Epidemiology* 22 (5), 745.

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

Yang, W. and Weng, J. (2014). “Early therapy for type 2 diabetes in China.” *The Lancet* *Diabetes & Endocrinology* 2 (12), 992–1002.

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.

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

Zhang, B., Zhai, F. Y., Du, S. F., and Popkin, B. M. (2014). “The China Health and Nutrition Survey, 1989-2011.” *Obesity Reviews* 15 (S1), 2–7. arXiv:  [NIHMS15000](http://arxiv.org/abs/NIHMS150003)3.

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.

Zhao, X., Zhu, X., Zhang, H.-s., Zhao, W., Li, J., Shu, Y., Li, S., Yang, M., Cai, L., Zhou, J., and Li, Y. (2012). “Prevalence of diabetes and predictions of its risks using anthropometric measures in southwest rural areas of China.” *BMC Public Health* 12 (1), 821.

Zhou, X., Ji, L., Ran, X., Su, B., Ji, Q., Pan, C., Weng, J., Ma, C., Hao, C., Zhang, D., and Hu, D. (2016). “Prevalence of Obesity and Its Influence on Achievement of Cardiometabolic Therapeutic Goals in Chinese Type 2 Diabetes Patients: An Analysis of the Nationwide, Cross-Sectional 3B Study.” *PLOS ONE* 11 (1). Ed. by J. Devaney, e0144179.

35