Diabetes, employment and behavioural risk factors in China: Marginal structural models versus fixed effects models

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

[a] *Luxembourg Institute of Socio-Economic Research, 11 Porte des Sciences, 4366 Esch-sur-Alzette/Belval, Luxembourg, till.seuring@liser.lu*

[b] *School of International Development, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, UK*

[c] *Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR47TJ, UK*

[d] *Centre for Health Economics, University of York, Heslington, York, YO105DD, UK*

Keywords: China, diabetes, employment, behavioural risk factors, marginal structural model

JEL codes: D83, E24, F61, I12, I14, J24

Declarations of interest: none.

*Corresponding author

Abstract

We use longitudinal data from the China Health and Nutrition Survey, covering the years 1997 to 2011, to estimate the effect of a diabetes diagnosis on an economic outcome (employment probabilities) and behavioural risk factors (alcohol consumption, smoking cessation, body mass index (BMI), physical activity and hypertension) for men and women. We apply two complementary statistical techniques—marginal structural models (MSMs) and fixed effects (FE) models—to deal with confounding. Both methods suggest, despite their different underlying assumptions, similar patterns that indicate important differences between men and women. Employment probabilities decline substantially after the diagnosis for women (-12.4 (MSM) and -15.5 (FE) percentage points), but do not change significantly for men. In particular, the MSM estimates indicate an increase in hypertension (13 percentage points) and a decrease in physical activity for women, while men have small and statistically insignificant changes in these outcomes. For BMI, the MSM results indicate statistically significant changes for men (-.76), but not for women, while the FE estimates show similar reductions for men and women (-.80 and -.73 respectively). Men also reduce their alcohol consumption, but do not cease to smoke. For women these risk factors have a prevalence close to zero to begin with, though women seem to still reduce alcohol consumption somewhat. These results suggest important gender differences in the impact of diabetes in China. To narrow these inequities policies supporting women to reduce diabetes related risk factors are likely important.

1 Introduction

The effect of diabetes on employment status has received little attention in low- and middle-income countrys (LMICs) (Seuring, Archangelidi, Suhrcke 2015). This is despite high prevalence rates in many LMICs like China, Mexico and South Africa, which have reached levels of over ten percent among the adult population, partly overtaking high-income countries like the United States (International Diabetes Federation 2019). The severity of diabetes, once it is diagnosed, as well as the potential for complications, and the economic impact of diabetes strongly depend on individual patient behaviour. This is the case in particular for the most common type of diabetes, type 2 diabetes, whose development may also be related to health behaviours. Previous research shows that behaviour changes such as increased physical activity, dietary changes and reduced alcohol consumption after a type 2 diabetes diagnosis are related to health and reductions in the risk of subsequent cardiovascular events (Long, Cooper, et al. 2014; Zhou et al. 2016). Thus, a diabetes diagnosis may present an important opportunity to reduce risk factors for complications (De Fine Olivarius et al. 2015) and to alleviate the resulting economic burden, raising the question what the impact is of a diabetes diagnosis on these outcomes.

Diabetes, economic outcomes and behavioural risk factors are likely interrelated, making it difficult to establish causal pathways. For example, transitioning from unemployment to employment may reduce physical activity by decreasing available leisure time; or may promote risk factors such as smoking and higher energy intake by changing the available income, thereby affecting the probability of developing diabetes and its complications (Colombo, Rotondi, Stanca 2018). Similarly, unemployment can lead to weight gain but also reduce smoking and fast-food consumption (Colman, Dave 2014).

Nevertheless, existing research on the impact of diabetes on labour market outcomes has so far assumed that diabetes is unaffected by prior employment outcomes, or has used instrumental variable (IV) strategies (Brown, Pagán, Bastida 2005; Latif 2009; Seuring, Goryakin, Suhrcke 2015) with at least questionable instruments (for a discussion, see for instance Seuring, Serneels, Suhrcke (2019)). It has also not been possible to credibly control for the independent effect of body mass index (BMI) or hypertension on employment outcomes, as both are likely affected by diabetes themselves, leading to biased estimates of the effect of diabetes on employment in standard regression models (Angrist, Pischke 2009). Studies investigating behaviour change after a diabetes diagnosis remain scarce and focus primarily on high-income countries and the elderly population (Gaggero 2020), without accounting for the selection into a diabetes diagnosis based on prior behaviour change (Slade 2012). Evidence stratified by gender is also missing, even though differences in health behaviours between men and women may help to explain gender differences in the complication risk of diabetes (Kautzky-Willer, Harreiter, Pacini 2016; Huebschmann et al. 2019; Harreiter, Kautzky-Willer 2018; The Lancet Diabetes & Endocrinology 2017; Kautzky-Willer, Harreiter 2017).

To assess the impact of a diabetes diagnosis on both employment probabilities and behavioural risk factors, this study uses longitudinal data from China, a country where about 13% of adults between the age of 40 to 60 have diabetes[1] , and over 50% of those remain undiagnosed (Wang, Gao, et al. 2017). We take various sources of confounding into account, first by estimating marginal structural models (MSMs) to account for any time-dependent confounding (Robins, Hernán, Brumback 2000). Second, we complement this strategy with fixed effects (FE) models to account for any time-invariant unmeasured confounding. Apart from this methodological innovation, the study extends the scarce evidence base for the impact of diabetes on employment in LMICs and provides, as far as we are aware, the first

[1]Here we refer to any type of diabetes. However, it is generally assumed that about 90% of all diabetes cases are type 2 diabetes. This is largely confirmed for China by recent evidence which found five to six percent of newly diagnosed diabetes cases among people 30 years or older having type 1 diabetes (Tang et al. 2019)

longitudinal evidence for the effect of a diabetes diagnosis on behavioural risk factors in a LMIC.

2 Data

The China Health and Nutrition Survey (CHNS) is a longitudinal survey providing information on socioeconomic outcomes, health, health behaviours and nutrition in nine provinces of China (Zhang et al. 2014). We use data from 1997 onwards (with survey rounds in 1997, 2000, 2004, 2006, 2009 and 2011): 1997 was the first time diabetes information was provided. The sample is limited to the adult population aged 18–64, is not nationally representative and the CHNS does not provide sampling weights (Popkin et al. 2010). We exclude students and women who reported to be pregnant at the time of the survey. Further, due to relatively early retirement in China for those in formal employment and for women, once people reported to be retired they were excluded from the sample from this point onwards.

Our main interest lies with the effect of developing diabetes, and we therefore exclude individuals with self-reported diabetes at baseline. Given the chronic nature of diabetes, we assume that it persists after diagnosis for the rest of one's life. We also investigate the effect of time since diabetes diagnosis on our outcomes and therefore construct a measure of diabetes duration using self-reported information on the year of diagnosis.

The economic outcome of interest is employment status, based on a self-reported response stating the respondent's current work status. This includes working in informal jobs, family businesses and farms.

The behavioural risk factor outcomes are binary variables for currently smoking, whether alcohol was consumed equal to or more than three times per week and whether the person had hypertension based on the average blood pressure from three consecutive readings of ≥ 140 mm Hg for systolic blood pressure or ≥ 90 mm Hg for diastolic blood pressure. We further assess the effect on BMI, daily calorie consumption and overall level of physical activity. We chose these outcomes because they present among the most important risk-factors for diabetes and diabetes related complications (American Diabetes Association 2020; Long, Johansson, et al. 2015; Long, Cooper, et al. 2014). BMI is based on height and weight measurements, daily calorie consumption is based on an individual's self-reported average daily consumption of carbohydrates, protein and fat, measured on three consecutive days, and was calculated by the CHNS investigators. Physical activity includes activities related to different types of occupation, leisure, travel to work and homework and is expressed in metabolic equivalent of task (MET) hours per week. One MET is defined as the ratio of a person's working metabolic rate in relation to her resting or basal metabolic rate.[2] [3]

3 Methods

We apply two distinct estimation methods: marginal structural model (MSM) and fixed effects (FE) estimation. Figure A1 and figure A2 presents the directed acyclic graph (DAG) for the respective models, providing a visual overview of the key differences between MSM and FE models. We estimate models separately for men and women as it is likely that diabetes has differential effects on employment and behavioural risk factors given results from previous studies and evidence for gender differences in the severity of diabetes (Kautzky-Willer, Harreiter, Pacini 2016; Huebschmann et al. 2019; Harreiter, Kautzky-Willer 2018; The Lancet Diabetes & Endocrinology 2017; Kautzky-Willer, Harreiter 2017; Minor 2011; Latif 2009; Harris 2009; Seuring, Serneels, Suhrcke 2016; Rodríguez-Sánchez, Cantarero-Prieto 2019).

3.1 Marginal structural models

MSMs can, contrary to FE models, adjust for confounding and selection bias arising from time-varying confounders affected by prior exposure to treatment (Robins, Hernán, Brumback 2000).

This requires the estimation of inverse probability of treatment weights (IPTW), which are the inverse of the probability of receiving treatment, conditional on past treatment and covariate history. Because our analysis is stratified by gender, we calculate separate weights for men and women. For the calculation of IPTW, we first calculate the probability, p, that a person will have received a diabetes diagnosis by a given time, conditional on the prior history of diabetes and observed time-constant and time-varying covariates. Then each person is weighted by the inverse of her conditional probability. Those in the treated group, i.e. who have been diagnosed at time t, are given a weight of $\frac{1}{p}$ assigning lower weights to persons with higher probabilities and higher weights to persons with lower probabilities. Those in the comparison group, i.e. those who were not diagnosed at time t, are given a weight of $\frac{1}{1-p}$ assigning higher weights to persons with higher probabilities and lower weights to those with

[2] We followed the Compendium of Physical Activities (Ainsworth et al. 2011) and the previous literature on calculating physical activity levels in the CHNS (Ng, Popkin 2012; Ng, Norton, Popkin 2009) to assign an MET to each reported activity in the survey and then multiplied them with the number of hours per week spend on carrying out this activity.

[3] BMI and MET were analysed as continuous variables instead of categorising them into overweight and obesity groups or physical activity categories, because used continuously they provide more information and are thus more sensitive to potential changes than when categorised. Furthermore, BMI, and to an extend also physical activity, have continuous associations with the risk of type 2 diabetes and its complications, that are not necessarily well captured using categorised variables (Bays, Chapman, Grandy 2007).

lower probabilities. This allows for the creation of a pseudo population exchangeable with the study population within the levels of confounders (Cole, Hernán 2008), ensuring that confounders and treatment are independent of each other in a weighted regression model.

The IPTW are calculated as depicted in the following model:

$$IPTW_{it} = \prod_{t=0}^{T} \frac{Pr(D_t = z|\bar{D}_{t-1}, X_0)}{Pr(D_t = z|\bar{D}_{t-1}, X_0, \bar{X}_{t-1})} \tag{1}$$

where t indexes time, i indexes the person, $D_t = z$ is the treatment actually received (diabetes diagnosis), X is a vector of time-invariant and time-dependent confounders including our outcome variables, variables subscripted with a 0 represent baseline values, and variables subscripted with $t-1$ are one period lags. We use overbars to denote covariate history up to time t for time-variant confounders.

The denominator is calculated using a logistic regression model to predict the probability of a diabetes diagnosis as indicated in Eq. 1, conditional on time-variant confounders measured at baseline when the individual was first observed in the sample, time-variant confounders lagged by one period (e.g. using BMI from the 2004 to predict diabetes in 2006) and time-invariant confounders as independent variables. We use lagged time-variant confounders to ensure that predictors of diabetes were determined previous to the manifestation of diabetes. X consists of age and age squared; an urbanization index pre-constructed within the CHNS data (Zhang et al. 2014); having secondary or university education, being married, having health insurance, Han ethnicity, region and time dummies, inflation adjusted per-capita household income, survey year dummies, employment status, alcohol consumption, smoking status, BMI, calorie consumption, physical activity levels and measured hypertension. The resulting IPTW for being diagnosed with diabetes are calculated for each individual at each survey wave. Then the IPTW from each wave after the baseline is multiplied with the IPTW from all previous waves to create the overall IPTW that reflects cumulative probabilities over time.

To reduce the variance of the overall IPTW, the numerator of Eq. 1 consists of an additional set of weights using only baseline values of the predictors as covariates. Eq. 1 gives stabilized IPTW that only reflect confounding due to the time-varying covariates, which cannot be appropriately adjusted for by standard regression models (Cole, Hernán 2008).

To account for the potential of attrition bias, we estimate stabilized censoring weights based on the probability to remain uncensored until the end of the panel. The model is similar to the IPTW model above, now using as dependent variable a dummy variable indicating censoring in the following wave. We then estimate the probability of remaining uncensored until the last observation in the individual's panel t using the covariates X as described above, additionally accounting for a person's diabetes history $\bar{D}_t$.

$$IPCW_{it} = \prod_{t=0}^{T} \frac{Pr(C_t = z|\bar{C}_{t-1}, X_0)}{Pr(C_t = z|\bar{C}_{t-1}, X_0\bar{X}_t\bar{D}_t)} \tag{2}$$

After the creation of the inverse probability of censoring weights (IPCW), the weights to be used in our MSMs are calculated as the product of IPCW and IPTW. We then estimate the following linear regression models of the effect of a diabetes diagnosis on our outcomes of interest, while accounting for any time-variant confounding by applying the resulting weights:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_0 + u_i \tag{3}$$

where Y_i represents the respective outcome variable, D_i is a binary variable indicating a diabetes diagnosis after baseline, X_0 is a vector containing any baseline and time-invariant confounders used in the calculation of the IPTW and u_i is the error term. Robust standard errors clustered at the individual level are used throughout. The same model is used to estimate the effect of the time since diabetes diagnosis on our outcomes. The calculated stabilized weights used in our primary analysis of the MSMs are shown in Table A2 of the Appendix.

3.2 Fixed effects

In contrast to the MSM, the FE model accounts for time-invariant unobserved confounders, relying on within-person variation for identification. This comes at a cost: effects of variables that are invariant over time cannot be estimated. Further, as with any non-dynamic regression model and contrary to the MSM, past treatments are assumed to have no direct effect on current outcomes, and past outcomes are assumed to have no direct effect on current treatment (Imai, Kim 2019). Additionally, only confounders unaffected by a diabetes diagnosis should be included as control variables, as these would otherwise capture part of the causal effect of diabetes on the outcome of interest (Angrist, Pischke 2009; Imai, Kim 2019). Hence, while we can control for the intermediate effects of alcohol, smoking, BMI, physical activity, calorie consumption or hypertension on the outcome of interest and on diabetes in MSMs, we should not include these in the FE model. For the employment model we additionally do not control for household income or health insurance status as they are closely related to employment status.

We estimate the following FE model

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 X_{it} + c_i + u_{it} \tag{4}$$

where Y_{it} is the respective outcome of interest at time t, D_{it} indicates a diabetes diagnosis at time t (or time since diagnosis in our duration analysis), X_{it} is a vector of control variables unaffected by prior treatment or outcomes, c_i represents the individual fixed effect, and u_{it} is the error term, which can vary over time and across individuals. X_{it} includes age squared, the level of urbanization, education, being married, health insurance, living in a rural area, region and time dummies as well as per capita household income. We do not use the fixed effects model to estimate the effect of time since diagnosis, since the increase in time since diagnosis is not distinguishable from the increase in age or overall time in the FE model (Wooldridge 2012). For the same reason age is excluded from all FE specifications.

3.3 Regression method

We use linear regression models for our analysis throughout, including for binary outcomes, to facilitate comparability between FE and the MSM and to allow for cluster-robust standard errors. Further, linear probability models have been shown to produce similar results to non-linear models (Angrist, Pischke 2009).

Because we use lagged independent variables to construct stabilized weights for the MSMs, the reported number of observations in the MSMs is lower compared to the FE models, where we do not use lagged variables. The summary statistics shown in Table 1 are based on the observations used in the FE models. The number of observations is stated below each table.

3.4 Robustness checks

We carry out several robustness checks. Because the FE model does not control for a potential bias introduced by censoring, we also estimate the MSM without censoring weights to increase comparability between the two models. Second, we re-estimate the MSMs truncating weights at the 1st and 99th percentile to reduce the influence of very extreme weights. Third, we estimate the FE model using time-variant confounders lagged by one period to test the robustness of the results to using lagged confounders and the same sample as the MSM. Finally, we re-estimate the effect of diabetes on the binary outcomes using logistic regression instead of linear regression models. Because the calculation of marginal effects after fixed effects logistic regression can be problematic, we present the results as odds ratios.

3.5 Multiple imputation

We use imputed data to avoid excluding participants with missing data on one or more variables. Chained multiple imputation is used to impute thirty data sets under the assumption that the imputed data are missing at random, using the user written ICE command in Stata (Royston, White 2009). All outcome and explanatory variables included in the MSM and FE models are included in the multiple imputations. Table A1 details the number of missing observations for each variable. We do not use multiple imputation for diabetes diagnosis and instead assume that after the first reported diagnosis the individual had diabetes in every ensuing wave, even when the observation was missing.

4 Results

To describe the distributions of our outcome and control variables at baseline, we report the means separately for men and women and for those who did and did not report diabetes over the observed period. Table 1 shows that both men and women who went on to report a diabetes diagnosis are older, have higher BMI and lower physical activity levels and higher rates of hypertension than those in the non-diabetes group. Further, men who report diabetes drink more alcohol, live in more urbanized regions and have a higher socioeconomic status as measured by education and income levels. Women who report diabetes, however, have lower education levels and are less likely to be employed at baseline.

Table 1

Sample baseline means for men and women, by diabetes status.

	Men			Women		
	No diabetes	Diabetes	p-value (t-test)	No diabetes	Diabetes	p-value (t-test)
Employed	0.90	0.92	0.475	0.81	0.77	0.148
Smoking	0.61	0.63	0.450	0.03	0.06	0.023
Alcohol consumption	0.27	0.43	<0.001	0.02	0.04	0.038
3-Day Ave: Energy (kcal)	2547.74	2505.69	0.412	2167.37	2172.70	0.897
BMI	22.22	24.80	<0.001	22.42	25.86	<0.001
Physical activity (MET)	178.67	158.58	0.003	214.53	193.62	0.138
Hypertension (biomarker)	0.14	0.27	<0.001	0.09	0.39	<0.001
Age	36.16	42.07	<0.001	36.98	45.28	<0.001
Han ethnicity	0.13	0.10	0.246	0.13	0.08	0.018
Married	0.75	0.93	<0.001	0.89	0.93	0.028
Secondary or higher education	0.68	0.73	0.124	0.51	0.31	<0.001
Any health insurance	0.26	0.47	<0.001	0.23	0.21	0.301
Urbanization index	53.94	64.14	<0.001	53.93	51.18	0.021
Rural area	0.70	0.56		0.71	0.60	
Per capita household income (2011 Yuan)	5182.25	6090.24	0.014	5065.56	4804.45	0.419
Number of individuals	5761	121		5659	115	

Note The table shows the average baseline values, i.e as individuals joined the sample, stratified into groups depending on whether they went on to develop (report) diabetes in any of the following waves or not. People with diabetes reported at baseline are excluded.

The calculation of the stabilized weights for the MSM indicates that, in particular for men, changes in employment, alcohol consumption and smoking predict self-reporting of diabetes

(Table A3 of the Appendix). For women this is not the case, which suggests that MSM may help to reduce bias due to time-variant confounding in particular for men.

Table 2
The effect of a diabetes diagnosis on employment status and behavioural outcomes using MSM and FE.

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
				Marginal structural model			
Men							
Diabetes	0.006	−.046	−.088*	0.024	−.762***	−117.299	−11.597
	(.031)	(.038)	(.044)	(.043)	(.200)	(69.756)	(10.787)
Women							
Diabetes	0.124**	−.033	−.019***	0.130***	−.383	−60.742	−33.855**
	(.039)	(.022)	(.006)	(.039)	(.277)	(41.220)	(11.445)
				Fixed effects			
Men							
Diabetes	0.014	−.001	−.100**	0.011	−.797***	−141.949*	−1.392
	(.029)	(.035)	(.038)	(.043)	(.200)	(69.219)	(12.222)
Women							
Diabetes	−.155***	−.016	−.018	0.065	−.730**	−57.988	−33.787*
	(.040)	(.012)	(.014)	(.040)	(.222)	(58.055)	(13.993)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Control variables for FE: age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income. For the FE model on employment, we do not control for income or insurance status as they are likely affected by changes in employment. MSM controls for baseline values of the same variables as the FE models additionally to baseline values of age, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. Sample size for MSM: N=16557 (men), N=16252 (women). Sample size for FE models: N=22319 (men), N=21913 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

The regression results in Table 2 show reductions in women employment probabilities due to a diabetes diagnosis in all models. These reductions are somewhat larger in the FE model compared to the MSM. For men, the effects are qualitatively and statistically insignificant in both models.

Looking at behavioural risk factors, alcohol consumption but not smoking is reduced after a diabetes diagnosis in men. Further, BMI decreases for men to a similar extent in the MSM and the FE model. For women, only the FE model indicates a reduction in BMI, similar in size to that of men. The MSM shows a smaller and statistically insignificant reduction in BMI for women. We find some evidence of women reducing their physical activity levels and having a higher risk of hypertension after a diabetes diagnosis using the MSM, while men do not experience such changes. Overall, the evidence points to less favourable changes in behavioural risk factors and similarly a larger employment penalty for women compared to men.

Using time since diagnosis as a continuous variable, the MSMs (Table 3) indicates a steady reduction of women employment probabilities and physical activity levels, and potentially an increase on the risk of hypertension, but also small decreases in BMI and caloric consumption.

For men, BMI is reduced.

Table 3
The effect of each year since diabetes diagnosis on employment status and behavioural outcomes using MSM.

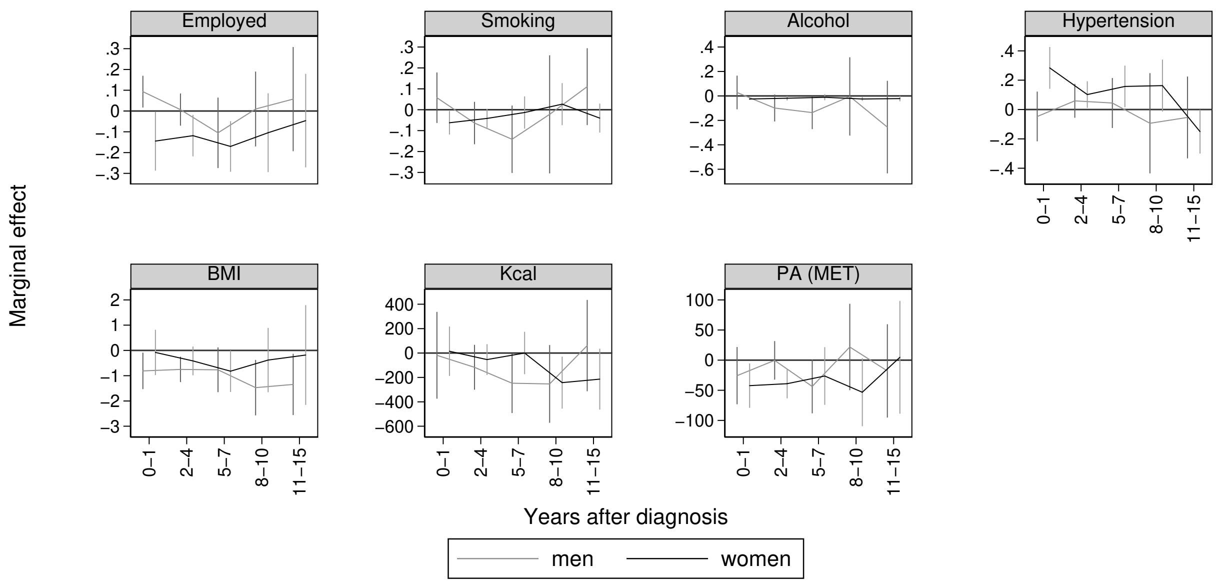
	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Time since diagnosis	−.001	-.004	−.015	-.000	−.142***	-20.134	−1.741
	(.006)	(.007)	(.008)	(.006)	(.031)	(11.549)	(1.848)
Women							
Time since diagnosis	−.016**	-.003	−.002**	.011	−.058	-12.718*	−4.096
	(.006)	(.003)	(.001)	(.006)	(.053)	(6.222)	(2.145)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income, alcohol consumption, smoking status, BMI, calorie consumption, physical activity and hypertension. Sample size: N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Dummy variables capturing time-periods after the diagnosis are used to investigate potential non-linearities in the effects over time. The results are visualized in Figure 1 and presented in Table A4 of the Appendix and indicate a reduction in employment probabilities of women in at least the first eight years after diagnosis. Further, they show consistent reductions of BMI for men, and to a lesser extent, for women. For physical activity, the MSM indicates a consistent reduction for women over the first ten years after diagnosis. No consistent associations over time were found for the other risk factors. Overall it appears that after 10–15 years observed differences between people with and without diabetes become smaller and are no longer distinguishable from zero, possibly also because the reduced sample of people with long term diabetes increases standard errors.

Using weights that do not account for censoring in the MSM yields very similar results, suggesting little bias due to censoring (Table A8, A9 and A10 of the Appendix). Likewise, using truncated weights leads to qualitative similar estimates (Table A5, A6 and A7 of the Appendix). Finally, estimating the FE model with lagged covariates and a smaller sample, gives results very similar to those of the MSM, with a slight reduction in the adverse effect of diabetes on employment, and the hypertension risk in women now being adversely affected by a diabetes diagnosis (Table A11 of the Appendix). Finally, the results of logistic regression models used to re-estimate the effect of a diabetes diagnosis on employment, smoking, alcohol consumption and the risk of hypertension support the findings from the linear probability models (LPMs), although they are not directly comparable due to the need to present results as odds ratios (Table A12 of the appendix).

Figure 1
The effect of each year since diabetes diagnosis on employment status and behavioural outcomes using MSM (duration groups).



Note The visualized coefficients are based on the results of the regression models shown in Tables A4. The bars indicate 95% confidence intervals. The coefficients present marginal effects compared to baseline.

5 Discussion

This study adds to the scarce evidence of the effect of a diabetes diagnosis on diabetes risk factors and employment status using longitudinal data from China, improving upon previously used methodologies by taking into account potential confounding over time.

Our results suggest that a diabetes diagnosis leads to a strong and lasting reduction in employment probabilities for women but not for men in this context. At the same time, men and potentially women reduce their BMI as a result of the diagnosis. Overall, men appear to achieve greater positive changes in their risk behaviours post diagnosis, maintaining their physical activity levels and keeping hypertension risk the same, contrary to women who reduce physical activity levels after diagnosis and may also experience an increased risk of hypertension.

5.1 Methodological considerations

The MSMs and FE models overall show similar trends and effect sizes. Because none of the models can simultaneously account for both unobserved and time-variant confounding, this could mean that either both models correct for distinct but more or less similar sized biases, or that both models are able to account for the same source of bias. The latter would be the case if a combination of both time-invariant unobserved factors—such as a genetic predisposition to diabetes that increases the risk to develop diabetes—and time-variant factors—such as job loss or increases in weight—would cause the onset of diabetes in those genetically predisposed to its development.

A limitation of the study is that the estimates cannot be interpreted as fully causal, as we cannot completely exclude potential omitted variable bias. However, given the closely similar results of both estimation strategies, we believe that the results strongly suggest that women are more adversely affected by diabetes than men. Unfortunately, with the methodologies used we are not able to assess in how far changes in behavioural outcomes have played a role in improving diabetes and consequently economic outcomes. Further limitations arise from the nature of the data. A first one is related to the way alcohol consumption is measured, which does not capture the actual quantities of alcohol consumed at each occasion, potentially missing changes among people that are infrequent or non-heavy drinkers. Second, the diabetes diagnosis was self-reported so that there may have been some false reports of diabetes; this also prohibits us distinguishing between different types of diabetes. The number of cases reporting the use of insulin immediately after diagnosis, which can be used as an indicator for type 1 diabetes, in our sample is around 10 percent. Re-estimating our models dropping these cases only leads to marginal changes in our estimates (results available on request).

Third, while the data covers a large part of China, the data and therefore out results are not nationally representative. Finally, given the overall small number of new diabetes diagnoses observed over time, especially the results using duration groups should be interpreted with caution due to the small number of cases especially in the longer duration groups.

5.2 Potential mechanisms

The results regarding weight loss after a diabetes diagnosis are consistent with those from other studies. Slade (2012) found reductions in overweight and obesity immediately after a diabetes diagnosis, though not over the long term. Our results indicate that weight loss may be more permanently, in particular for men. Permanent reductions in weight after diagnosis were also observed in a cohort of Danish patients (De Fine Olivarius et al. 2015). In that setting the decline was attributed to motivational changes stemming from the diabetes diagnosis, which may represent a window of opportunity to initiate long lasting weight reductions. Similarly, Gaggero (2020) finds a reduction in BMI shortly after a diabetes diagnosis, without reporting longer term effects. Nonetheless, weight reductions may also be—at least partly—the result of treatment initiation with diabetes drugs that cause weight loss (Yang, Weng 2014). Our study did not investigate changes in dietary quality and if these changes may explain reductions in weight loss. While it is not clear if changes in dietary quality can directly cause weight loss without also causing changes in a person's energy balance, better dietary quality may still be of importance for the prevention of diabetes complications. It may help with achieving reductions in calories and independently can allow for a better control of blood glucose levels and the reduction in risk factors such as hypertension or high cholesterol levels (Ley et al. 2014). Potential changes in dietary quality after a diabetes diagnosis will present an interesting subject for future research. With regards to alcohol consumption, we find a significant reduction for women using the MSM. One possibility is that reducing alcohol consumption for women with diabetes is a relatively easily achieved task, given the already low prevalence rates and potentially also, because those women may not have been heavy users to begin with.

The evidence we find for a worsening of risk factors of women may be explained in several ways. Generally lower educational attainment and income of women may reduce their exposure to health information and limit the access to treatment (Luo et al. 2015; Ma, Nolan, Smith 2018). Women may also receive less spousal support or support of their close network in the management of their disease, making it more difficult to change health behaviours (Albanese et al. 2019). Women have also been found to be in a worse metabolic health state compared to men when crossing the diabetes threshold, with a higher risk of cardiovascular disease and stroke after diagnosis (Kautzky-Willer, Harreiter, Pacini 2016; Huebschmann et al. 2019; Harreiter, Kautzky-Willer 2018; The Lancet Diabetes & Endocrinology 2017; Kautzky-Willer, Harreiter 2017). Potentially as a result of these factors, Chinese women with diabetes experience more comorbidities than men (Liu et al. 2010).

This has been the first study to use MSM to explore the impact of a diabetes diagnosis on employment longitudinally. Previous longitudinal studies used fixed effects models only, finding reductions in employment probabilities for men and women of about 5 percentage points in Mexico (Seuring, Serneels, Suhrcke 2019). Taking into account the lower overall employment rate of Mexican women compared to men, this translated into a 16% reduction in female employment probabilities, a figure comparable to the effect observed in this study. Overall, the adverse effect of diabetes on employment is in line with other studies that have found diabetes to reduce employment probabilities for women (Minor 2011; Latif 2009; Harris 2009; Seuring, Serneels, Suhrcke 2016)—often more than for men. The large gender differences in the employment impact may, at least partly, be driven by the observed differences in behaviour change and in risk factors for complications, leading to worse health outcomes in women that result in a decrease in their employment probabilities. Further, evidence from Mexico points towards a larger employment penalty of diabetes for those in the informal labour market (Seuring, Goryakin, Suhrcke 2015). Given the considerable informal sector in China and the over-representation of women in this sector (Wang, Klugman 2020), it is possible that women are more exposed to low job security, increasing their chances to be laid-off due to their diabetes, be it due to actual health problems, or the stigma surrounding the disease.

Given the high prevalence of undiagnosed diabetes, early diagnosis is to be encouraged to foster positive behaviour change, and potentially reduce the individual economic burden of diabetes. Our results also suggest greater emphasis needs to be placed on women to reduce the observed inequities in the impact of diabetes. Future research may want to study in more detail the mechanisms behind these impacts, including the potential mediating role of behavioural risk factors for the economic impact of diabetes. This may also improve our understanding of the difference in impact of diabetes between men and women.

6 Acknowledgements

We would like to thank participants of the congress of the German Health Economics Association (2017), the Student Prize Special Organized Session at the congress of the International Health Economics Association (2017), and the brown bag seminar at the Centre for Modern Indian Studies at the University of Göttingen (2017) for their comments. This research uses data from China Health and Nutrition Survey (CHNS). We thank the National

Institute for Nutrition and Health, China Center for Disease Control and Prevention, Carolina Population Center (P2C HD050924, T32 HD007168), the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, R24 HD050924, and R01-HD38700) and the NIH Fogarty International Center (D43 TW009077, D43 TW007709) for financial support for the CHNS data collection and analysis files from 1989 to 2011 and future surveys, and the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009, Chinese National Human Genome Center at Shanghai since 2009, and Beijing Municipal Center for Disease Prevention and Control since 2011.

References

Ainsworth, B. E., W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, A. S. Leon (2011), 2011 compendium of physical activities: A second update of codes and MET values, DOI: 10.1249/MSS.0b013e31821ece12.

Albanese, A. M., J. C. Huffman, C. M. Celano, L. M. Malloy, D. J. Wexler, M. E. Freedman, R. A. Millstein (2019), The role of spousal support for dietary adherence among type 2 diabetes patients: a narrative review, Social Work in Health Care 58 (3) 304–323, DOI: 10.1080/00981389.2018.1563846.

American Diabetes Association (2020), 10. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes—2020, Diabetes Care 43 (Supplement 1) S111–S134, DOI: 10.2337/dc20-S010.

Angrist, J., J. Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.

Bays, H. E., R. H. Chapman, S. Grandy (2007), The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: Comparison of data from two national surveys, International Journal of Clinical Practice 61 (5) 737–747, DOI: 10.1111/j.1742-1241.2007.01336.x.

Brown, H. S., J. A. Pagán, E. Bastida (2005), The Impact of Diabetes on Employment: Genetic IVs in a Bivariate Probit, Health Economics 14 (5) 537–544, DOI: 10.1002/hec.942.

Cole, S. R., M. A. Hernán (2008), Constructing Inverse Probability Weights for Marginal Structural Models, American Journal of Epidemiology 168 (6) 656–664, DOI: 10.1093/aje/kwn164.

Colman, G., D. Dave (2014), Unemployment and Health Behaviors Over the Business Cycle: a Longitudinal View, NBER Working Paper (20748).

Colombo, E., V. Rotondi, L. Stanca (2018), Macroeconomic conditions and health: Inspecting the transmission mechanism, Economics and Human Biology 28 29–37, DOI: 10.1016/j.ehb.2017.11.005.

De Fine Olivarius, N., V. D. Siersma, R. Køster-Rasmussen, B. L. Heitmann, F. B. Waldorff (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, DOI: 10.1371/journal.pone.0122219.

Gaggero, A. (2020), The effect of type 2 diabetes diagnosis in the elderly, Economics and Human Biology 37 1–23, DOI: 10.1016/j.ehb.2019.100830.

Harreiter, J., A. Kautzky-Willer (2018), Sex and gender differences in prevention of type 2 diabetes, Frontiers in Endocrinology 9 (MAY) 1–15, DOI: 10.3389/fendo.2018.00220.

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, DOI: 10.1111/j.1475-4932.2009.00572.x.

Huebschmann, A. G., R. R. Huxley, W. M. Kohrt, P. Zeitler, J. G. Regensteiner, J. E. Reusch (2019), Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course, Diabetologia 62 (10) 1761–1772, DOI: 10.1007/s00125-019-4939-5.

Imai, K., I. S. Kim (2019), When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?, American Journal of Political Science 63 (2) 467–490, DOI: 10.1111/ajps.12417.

International Diabetes Federation (2019), Diabetes Atlas, tech. rep., Brussels, Belgium: International Diabetes Federation.

Kautzky-Willer, A., J. Harreiter (2017), Sex and gender differences in therapy of type 2 diabetes, Diabetes Research and Clinical Practice 131 (July) 230–241, DOI: 10.1016/j.diabres.2017.07.012.

Kautzky-Willer, A., J. Harreiter, G. Pacini (2016), Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus, Endocrine Reviews 37 (3) 278–316, DOI: 10.1210/er.2015-1137.

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

Ley, S. H., O. Hamdy, V. Mohan, F. B. Hu (2014), Prevention and management of type 2 diabetes: Dietary components and nutritional strategies, The Lancet 383 (9933) 1999–2007, DOI: 10.1016/S0140-6736(14)60613-9.

Liu, Z., C. Fu, W. Wang, B. Xu (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 (1) 62, DOI: 10.1186/1477-7525-8-62.

Long, G. H., I. Johansson, O. Rolandsson, P. Wennberg, E. Fhärm, L. Weinehall, S. J. Griffin, R. K. Simmons, M. Norberg (2015), Healthy behaviours and 10-year incidence of diabetes: A population cohort study, Preventive Medicine 71 121–127, DOI: 10.1016/j.ypmed.2014.12.013.

Long, G. H., A. J. Cooper, N. J. Wareham, S. J. Griffin, R. K. Simmons (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, DOI: 10.2337/dc13-1731.

Luo, X., T. Liu, X. Yuan, S. Ge, J. Yang, C. Li, W. Sun (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, DOI: 10.3390/ijerph120911304.

Ma, Y., A. Nolan, J. P. Smith (2018), The value of education to health: Evidence from Ireland, Economics and Human Biology 31 14–25, DOI: 10.1016/j.ehb.2018.07.006.

Minor, T. (2011), The effect of diabetes on female labor force decisions: new evidence from the National Health Interview Survey, Health Economics 20 (12) 1468–1486, DOI: 10.1002/hec.1685.

Ng, S. W., B. M. Popkin (2012), Time use and physical activity: a shift away from movement across the globe, Obesity Reviews 13 (8) 659–680, DOI: 10.1111/j.1467-789X.2011.00982.x.

Ng, S. W., E. C. Norton, B. M. Popkin (2009), Why have physical activity levels declined among Chinese adults? Findings from the 1991–2006 China health and nutrition surveys, Social Science & Medicine 68 (7) 1305–1314, DOI: 10.1016/j.socscimed.2009.01.035.

Popkin, B. M., S. Du, F. Zhai, B. Zhang (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, DOI: 10.1093/ije/dyp322.

Robins, J. M., M. Á. Hernán, B. Brumback (2000), Marginal Structural Models and Causal Inference in Epidemiology, Epidemiology 11 (5) 550–560, DOI: 10.1097/00001648-200009000-00011.

Rodríguez-Sánchez, B., D. Cantarero-Prieto (2019), Socioeconomic differences in the associations between diabetes and hospital admission and mortality among older adults in Europe, Economics and Human Biology 33 89–100, DOI: 10.1016/j.ehb.2018.12.007.

Royston, P., I. White (2009), Multiple imputation by chained equations (MICE): Implementation in Stata, Journal of Statistical Software 45 (4) 1–2–, DOI: 10.1093/ije/dyh299, arXiv: arXiv:1501.0228.

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

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

Seuring, T., P. Serneels, M. Suhrcke (2016), The Impact of Diabetes on Labor Market Outcomes in Mexico: A Panel Data and Biomarker Analysis, IZA Discussion Papers, Discussion Paper Series (10123).

— (2019), The impact of diabetes on labour market outcomes in Mexico: A panel data and biomarker analysis, Social Science & Medicine 233 252–261, DOI: 10.1016/j.socscimed.2019.05.051.

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

Tang, X., X. Yan, H. Zhou, X. Yang, X. Niu, J. Liu, Q. Ji, L. Ji, X. Li, Z. Zhou (2019), Prevalence and identification of type 1 diabetes in Chinese adults with newly diagnosed diabetes, Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy 12 1527–1541, DOI: 10.2147/DMSO.S202193.

The Lancet Diabetes & Endocrinology (2017), Sex disparities in diabetes: bridging the gap, The Lancet Diabetes and Endocrinology 5 (11) 839, DOI: 10.1016/S2213-8587(17)30336-4.

Wang, L., P. Gao, M. Zhang, Z. Huang, D. Zhang, Q. Deng, Y. Li, Z. Zhao, X. Qin, D. Jin, M. Zhou, X. Tang, Y. Hu, L. Wang (2017), Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013, Jama 317 (24) 2515, DOI: 10.1001/jama.2017.7596.

Wang, L., J. Klugman (2020), How women have fared in the labour market with China's rise as a global economic power, Asia and the Pacific Policy Studies 7 (1) 43–64, DOI: 10.1002/app5.293.

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

Yang, W., J. Weng (2014), Early therapy for type 2 diabetes in China, The Lancet Diabetes & Endocrinology 2 (12) 992–1002, DOI: 10.1016/S2213-8587(14)70136-6.

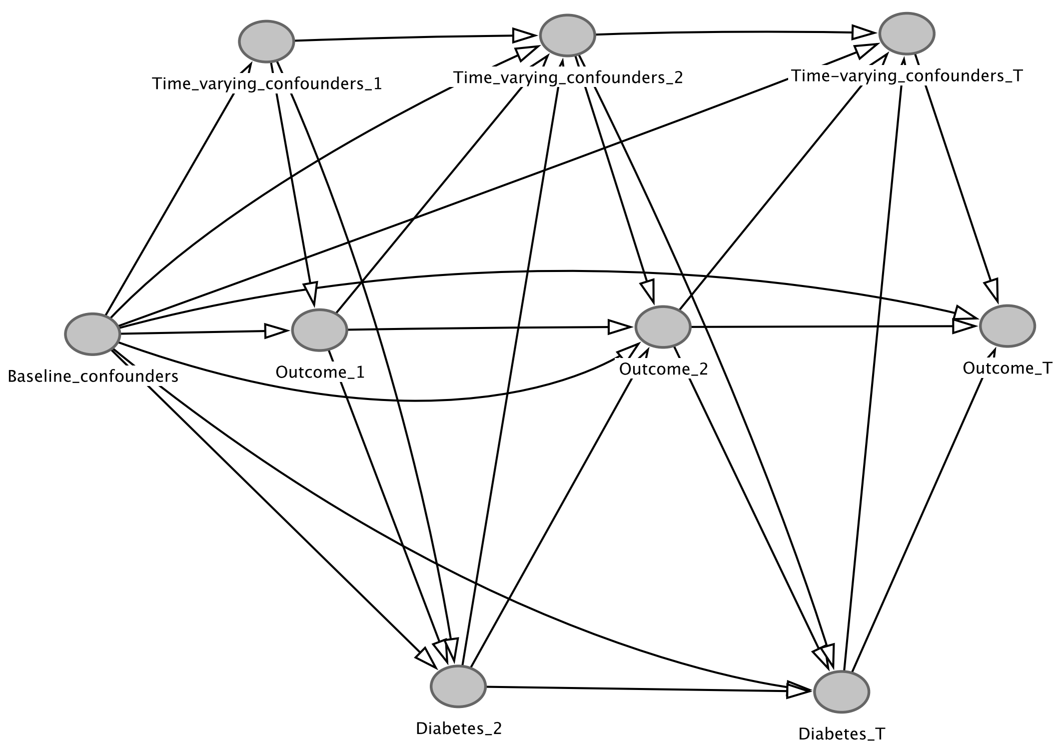
Zhang, B., F. Y. Zhai, S. F. Du, B. M. Popkin (2014), The China Health and Nutrition Survey, 1989-2011, Obesity Reviews 15 (S1) 2–7, DOI: 10.1111/obr.12119.

Zhou, X., L. Ji, X. Ran, B. Su, Q. Ji, C. Pan, J. Weng, C. Ma, C. Hao, D. Zhang, D. Hu (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) e0144179, DOI: 10.1371/journal.pone.0144179.

Appendix

Figure A1

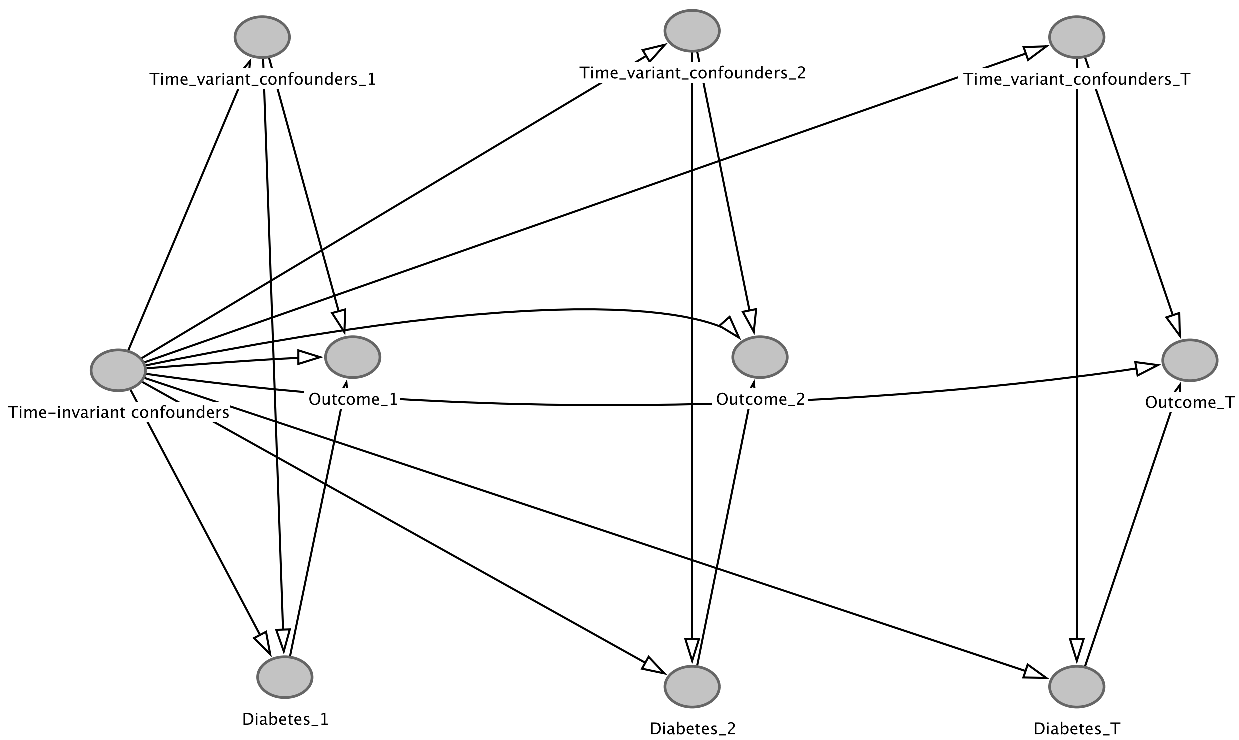
Direct acyclic graph for the marginal structural model.



Note MSMs assume the absence of unobserved time-invariant an.unobserved time-variant confounders but allow the past treatments to affect the current outcomes (arrows going from Diabetes to Outcome in the same wave) and the past outcomes to affect the current treatment (arrows going from Outcome in previous wave to current diabetes). Lagged time-variant confounders and baseline confounders predict current diabetes status and lagged outcomes.

Figure A2

Direct acyclic graph for the fixed effects model.



Note FE models account for any time-invariant confounding both observed and unobserved, but still assume the absence of unobserved time-variant confounding. They further do not allow for past outcomes to affect the current treatment, i.e. diabetes status.

Table A1
Number of imputed observations.

Variable	Missing	Non-missing	Missing (%)
Employed	2498	41734	5.6
Smokes	3174	41058	7.2
Alcohol consumption	3290	40942	7.4
Daily Kcal eaten (3-day average)	3485	40747	7.9
BMI	5849	38383	13.2
PA (MET)	2103	42129	13.35
Hypertension (biomarker)	5620	44579	4.8
Age	0	44579	0.00
Han ethnicity	0	44579	0.00
Married	2462	41770	5.6
Secondary and higher education	2413	41819	5.5
Any health insurance	2414	41818	5.5
Urbanization Index	0	44579	0.00
Diabetes	0	44579	0.00
Per capita household income (Yuan (2011))	512	43720	1.2
Years since diabetes diagnosis	20	44212	0.0

Table A2
Summary of stabilized weights.

	Mean	Minimum	Maximum
Untruncated (men)	1.02	0.17	3.67
Untruncated (women)	1.02	0.02	7.40
Truncated (men)	1.01	0.60	1.65
Truncated (women)	1.02	0.58	1.87

Note N=16557 (men), N=16252 (women).

Table A3

Time variant and invariant predictors of a diabetes diagnosis (denominator of stabilized weights): logistic regression models.

	Men		Women	
Baseline and time-invariant variables				
Age	0.758*	(0.087)	1.266	(0.208)
Age squared	1.004**	(0.001)	0.998	(0.002)
Urbanization index	1.001	(0.013)	1.007	(0.015)
Rural area	0.787	(0.179)	0.487**	(0.115)
BMI	1.222***	(0.063)	1.221***	(0.071)
3-Day Ave: Energy (kcal)	1.000	(0.000)	1.000	(0.000)
Smoking	1.378	(0.353)	1.000	(0.802)
Alcohol consumption	1.548	(0.356)	1.514	(1.077)
Secondary	0.706	(0.281)	0.645	(0.281)
University	0.642	(0.473)	–	–
Married	1.146	(0.584)	0.926	(0.531)
Any health insurance	1.245	(0.312)	0.967	(0.300)
Employed	2.115	(0.910)	1.644	(0.531)
Han ethnicity	0.988	(0.373)	0.632	(0.263)
Per capita household income (2011 Yuan)	1.000	(0.000)	1.000	(0.000)
Hypertension (biomarker)	0.992	(0.259)	1.704	(0.473)
Physical activity (MET)	0.998	(0.001)	0.999	(0.001)
Survey year				
2004	1.323	(0.523)	0.723	(0.234)
2006	1.308	(0.549)	0.532	(0.204)
2009	2.454*	(1.056)	0.897	(0.358)
2011	0.970	(0.480)	0.983	(0.445)
Lagged time-varying variables				
Age	1.664**	(0.258)	0.930	(0.157)
Age squared	0.995**	(0.002)	1.001	(0.002)
BMI	0.986	(0.049)	1.022	(0.058)
Urbanization index	1.016	(0.013)	0.989	(0.014)
3-Day Ave: Energy (kcal)	1.000	(0.000)	1.000	(0.000)
Smoking	0.583*	(0.142)	0.896	(0.715)
Alcohol consumption	0.633	(0.156)	0.821	(0.662)
Secondary	1.499	(0.622)	2.203	(0.946)
University	1.296	(0.890)	0.804	(0.858)
Married	0.981	(0.492)	0.907	(0.446)
Any health insurance	1.178	(0.289)	1.050	(0.320)
Employed	0.526*	(0.152)	0.727	(0.204)
Physical activity (MET)	1.000	(0.001)	1.000	(0.001)
Hypertension (biomarker)	1.268	(0.304)	1.164	(0.311)
Per capita household income (2011 Yuan)	1.000	(0.000)	1.000	(0.000)

Note Odds ratios. Standard errors in parenthesis. Results for province dummies omitted to preserve space. The variable University could not be estimated for women at baseline, as it perfectly predicted diabetes status. Base N=16439 (men), N=16113 (women). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A4
The effect of time since diabetes diagnosis on employment status and behavioural outcomes using MSM (duration groups).

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
0–1	0.095*	0.062	0.027	−.054	−.784*	−40.964	−4.428
	(.039)	(.061)	(.071)	(.083)	(.370)	(171.094)	(34.371)
2–4	0.004	−.066	−.106*	0.044	−.715**	−70.155	−6.245
	(.041)	(.052)	(.054)	(.058)	(.256)	(311.593)	(20.168)
5–7	−.087	−.130	−.133	0.054	−.749	−229.235*	−32.295
	(.076)	(.080)	(.071)	(.087)	(.448)	(107.881)	(24.850)
8–10	0.023	−.026	0.003	−.099	−1.531**	−252.410	15.587
	(.086)	(.145)	(.162)	(.174)	(.544)	(167.490)	(49.932)
11–14	0.064	0.103	−.272	−.063	−1.264	104.464	−26.757
	(.114)	(.102)	(.190)	(.147)	(.660)	(172.888)	(47.953)
0–1	0.093*	0.058	0.028	−.047	−.810*	−18.598	−25.703
	(.039)	(.062)	(.070)	(.086)	(.368)	(181.515)	(24.220)
2–4	0.007	−.064	−.099	0.059	−.751**	−116.940	−.326
	(.039)	(.052)	(.056)	(.059)	(.256)	(93.817)	(16.303)
5–7	−.105	−.141	−.136*	0.044	−.767	−247.440*	−43.579
	(.086)	(.082)	(.069)	(.087)	(.451)	(124.535)	(22.659)
8–10	0.010	−.022	−.004	−.094	−1.472**	−253.240	21.873
	(.092)	(.144)	(.163)	(.174)	(.559)	(162.492)	(36.522)
11–14	0.057	0.110	−.255	−.054	−1.345*	61.299	−17.822
	(.127)	(.093)	(.193)	(.142)	(.613)	(190.266)	(39.230)
Women							
0–1	−.145*	−.062*	−.025***	0.284***	−.079	14.668	−42.297*
	(.073)	(.029)	(.004)	(.073)	(.457)	(103.154)	(18.782)
2–4	−.118*	−.041	−.019*	0.102*	−.415	−53.783	−39.018**
	(.051)	(.023)	(.009)	(.046)	(.290)	(64.346)	(12.414)
5–7	−.170**	−.013	−.011	0.157*	−.822*	0.328	−26.308
	(.062)	(.039)	(.013)	(.072)	(.419)	(88.597)	(24.410)
8–10	−.104	0.027	−.025**	0.163	−.381	−242.489*	−53.214
	(.097)	(.051)	(.008)	(.090)	(.649)	(108.205)	(28.754)
11–14	−.046	−.040	−.022	−.150*	−.182	−213.862	4.717
	(.115)	(.035)	(.011)	(.076)	(1.007)	(127.050)	(47.724)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other control variables: baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A5

The effect of a diabetes diagnosis on employment status and behavioural outcomes using MSM with truncated weights at 1st and 99th percentile.

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Diabetes	−.004	−.058	−.095**	0.030	−.741***	−142.008*	−14.485
	(.030)	(.036)	(.036)	(.040)	(.193)	(63.428)	(10.240)
Women							
Diabetes	−.128***	−.030	−.019***	0.130***	−.376	−58.374	−34.827**
	(.037)	(.020)	(.006)	(.038)	(.272)	(40.861)	(11.119)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income alcohol consumption, smoking status, BMI, calorie consumption, physical activity and hypertension. Sample size: N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A6
The effect of each year since diabetes diagnosis on employment status and behavioural outcomes using MSM with truncated weights at 1st and 99th percentile.

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Time since diagnosis	−.003	-.007	−.015*	.001	−.147***	-22.087*	−2.282
	(.006)	(.006)	(.007)	(.007)	(.031)	(11.252)	(1.772)
Women							
Time since diagnosis	−.017**	-.003	−.002**	.011	−.056	-12.308*	−4.200*
	(.006)	(.003)	(.001)	(.006)	(.052)	(6.213)	(2.122)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income alcohol consumption, smoking status, BMI, calorie consumption, physical activity and hypertension. Sample size: N=16557 (men), N=16252 (women). * p<0.05, ** p<0.01, *** p<0.001.

Table A7
The effect of time since diabetes diagnosis on employment status and behavioural outcomes using MSM with truncated weights at 1st and 99th percentile (duration groups).

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
0–1	0.087*	0.054	–.002	–.057	–.744*	–128.911	–29.781
	(.042)	(.061)	(.065)	(.071)	(.365)	(132.013)	(24.819)
2–4	–.004	–.081	–.118*	0.073	–.691**	–136.257	–3.093
	(.039)	(.046)	(.046)	(.053)	(.235)	(86.271)	(15.131)
5–7	–.108	–.145	–.149*	0.046	–.761	–224.930	–42.806*
	(.080)	(.077)	(.062)	(.082)	(.416)	(119.446)	(20.719)
8–10	–.013	–.030	0.024	–.131	–1.615**	–232.071	18.467
	(.100)	(.142)	(.145)	(.135)	(.565)	(156.783)	(40.230)
11–14	0.039	0.105	–.172	–.018	–1.495*	54.355	–26.721
	(.131)	(.100)	(.184)	(.143)	(.635)	(200.571)	(40.121)
Women							
0–1	–.146*	–.060*	–.025***	0.275***	–.112	18.543	–44.753*
	(.073)	(.027)	(.004)	(.071)	(.450)	(103.734)	(18.404)
2–4	–.124*	–.039	–.019*	0.108*	–.400	–55.985	–39.865***
	(.049)	(.020)	(.009)	(.046)	(.285)	(62.172)	(12.099)
5–7	–.174**	–.009	–.011	0.153*	–.771	14.742	–26.326
	(.060)	(.035)	(.013)	(.072)	(.411)	(88.550)	(23.843)
8–10	–.105	0.026	–.025**	0.161	–.385	–244.078*	–54.441
	(.096)	(.051)	(.008)	(.090)	(.646)	(107.443)	(28.467)
11–14	–.047	–.040	–.022	–.151*	–.180	–213.686	4.424
	(.115)	(.035)	(.011)	(.076)	(1.007)	(126.873)	(47.687)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other control variables: baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A8
The effect of a diabetes diagnosis on employment status and behavioural outcomes using MSM with uncensored weights.

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Diabetes	−.007	−.041	−.088	0.038	−.687***	−124.116	−15.568
	(.032)	(.038)	(.046)	(.041)	(.200)	(69.861)	(10.748)
Women							
Diabetes	−.135***	−.030	−.020***	0.122**	−.365	−57.607	−39.302***
	(.036)	(.019)	(.006)	(.037)	(.275)	(41.479)	(11.170)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income alcohol consumption, smoking status, BMI, calorie consumption, physical activity and hypertension. Sample size: N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A9
The effect of each year since diabetes diagnosis on employment status and behavioural outcomes using MSM with uncensored weights.

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Time since diagnosis	−.003	-.003	−.016	.002	−.133***	-19.610	−2.166
	(.006)	(.006)	(.009)	(.006)	(.030)	(11.780)	(1.857)
Women							
Time since diagnosis	−.019**	-.003	−.002**	.012*	−.055	-12.494*	−5.087*
	(.006)	(.003)	(.001)	(.006)	(.053)	(6.213)	(2.118)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income alcohol consumption, smoking status, BMI, calorie consumption, physical activity and hypertension. Sample size: N=16557 (men), N=16252 (women). * p<0.05, ** p<0.01, *** p<0.001.

Table A10
The effect of time since diabetes diagnosis on employment status and behavioural outcomes using MSM with uncensored weights (duration groups).

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
0–1	0.080	0.057	0.035	−.042	−.606	−57.745	−31.227
	(.048)	(.061)	(.071)	(.079)	(.412)	(158.819)	(24.953)
2–4	−.011	−.065	−.104	0.077	−.684**	−131.817	−5.237
	(.040)	(.049)	(.058)	(.054)	(.247)	(98.646)	(16.973)
5–7	−.101	−.133	−.120	0.041	−.807	−232.150	−41.588
	(.086)	(.083)	(.072)	(.084)	(.421)	(129.047)	(23.065)
8–10	−.006	0.015	−.012	−.037	−1.324*	−227.770	18.969
	(.094)	(.136)	(.172)	(.179)	(.554)	(162.758)	(37.833)
11–14	0.053	0.120	−.274	−.051	−1.265*	96.735	−19.427
	(.118)	(.086)	(.195)	(.142)	(.586)	(168.556)	(36.253)
Women							
0–1	−.152*	−.059*	−.025***	0.244***	−.060	30.656	−49.597**
	(.069)	(.025)	(.005)	(.074)	(.435)	(105.419)	(18.811)
2–4	−.121*	−.039*	−.021*	0.094*	−.367	−60.189	−41.364**
	(.050)	(.019)	(.009)	(.045)	(.293)	(62.349)	(12.584)
5–7	−.194**	−.005	−.012	0.155*	−.874*	19.858	−31.087
	(.061)	(.035)	(.012)	(.072)	(.429)	(88.260)	(24.872)
8–10	−.123	0.026	−.026**	0.180*	−.442	−262.560*	−62.185*
	(.097)	(.053)	(.009)	(.088)	(.632)	(108.815)	(28.959)
11–14	−.066	−.041	−.021	−.145	−.009	−208.206	−7.337
	(.116)	(.036)	(.011)	(.075)	(1.021)	(118.884)	(47.266)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other control variables: baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A11
The effect of a diabetes diagnosis on employment status and behavioural outcomes using FE (lagged covariates).

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Men							
Diabetes	0.054	0.004	−.069	0.014	−.830***	−181.109*	−2.425
	(.035)	(.045)	(.049)	(.050)	(.227)	(86.015)	(16.033)
Women							
Diabetes	−.132*	−.011	−.010	0.160**	−.672*	−38.070	−53.022*
	(.058)	(.009)	(.015)	(.057)	(.294)	(76.524)	(20.710)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Control variables: Age squared, region, education, lagged marital status, lagged urbanization index, time dummies, lagged health insurance status, lagged household income. N=16557 (men), N=16252 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A12
The effect of a diabetes diagnosis on employment status and behavioural outcomes using logistic regression.

	(1) Employment	(2) Smoking	(3) Any alcohol	(4) Hypertension
		Marginal structural models		
Men				
Diabetes	1.062	0.775	0.613	1.063
	(.273)	(.162)	(.158)	(.241)
Women				
Diabetes	0.561**	0.306	0.212*	1.674**
	(.106)	(.190)	(.156)	(.326)
		Fixed effects		
Men				
Diabetes	1.327	1.046	0.482**	0.922
	(.458)	(.322)	(.130)	(.245)
Women				
Diabetes	0.293**	0.212	0.320	1.313
	(.121)	(.276)	(.293)	(.379)

Note Odds ratios; Standard errors in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Control variables for FE: age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income. For the FE model on employment, we do not control for income or insurance status as they are likely affected by changes in employment. MSM controls for baseline values of the same variables as the FE models additionally to baseline values of age, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. Sample size for MSM: N=16557 (men), N=16252 (women). Sample size for FE models: N=22319 (men), N=21913 (women). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.