



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/165582/>

Version: Accepted Version

---

**Article:**

Seuring, Till, Serneels, Pieter, Suhrcke, Marc et al. (2020) Diabetes, employment and behavioural risk factors in China:: Marginal structural models versus fixed effects models. *Economics and Human Biology*. 100925. ISSN: 1570-677X

<https://doi.org/10.1016/j.ehb.2020.100925>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

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

4 Till Seuring<sup>\*a</sup>, Pieter Serneels<sup>b</sup>, Marc Suhrcke<sup>a,d</sup>, and Max Bachmann<sup>c</sup>

5 <sup>a</sup>*Luxembourg Institute of Socio-Economic Research, 11 Porte des Sciences,*  
6 *4366 Esch-sur-Alzette/Belval, Luxembourg, till.seuring@liser.lu*

7 <sup>b</sup>*School of International Development, University of East Anglia, Norwich*  
8 *Research Park, Norwich, Norfolk, NR47TJ, UK*

9 <sup>c</sup>*Norwich Medical School, University of East Anglia, Norwich Research Park,*  
10 *Norwich, Norfolk, NR47TJ, UK*

11 <sup>d</sup>*Centre for Health Economics, University of York, Heslington, York,*  
12 *YO105DD, UK*

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

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

16 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 countries (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

52 alleviate the resulting economic burden, raising the question what the impact is of a diabetes  
53 diagnosis on these outcomes.

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

61 Nevertheless, existing research on the impact of diabetes on labour market outcomes  
62 has so far assumed that diabetes is unaffected by prior employment outcomes, or has used  
63 instrumental variable (IV) strategies (Brown, Pagán, Bastida 2005; Latif 2009; Seuring,  
64 Goryakin, Suhrcke 2015) with at least questionable instruments (for a discussion, see for  
65 instance Seuring, Serneels, Suhrcke (2019)). It has also not been possible to credibly control  
66 for the independent effect of body mass index (BMI) or hypertension on employment outcomes,  
67 as both are likely affected by diabetes themselves, leading to biased estimates of the effect  
68 of diabetes on employment in standard regression models (Angrist, Pischke 2009). Studies  
69 investigating behaviour change after a diabetes diagnosis remain scarce and focus primarily on  
70 high-income countries and the elderly population (Gaggero 2020), without accounting for the  
71 selection into a diabetes diagnosis based on prior behaviour change (Slade 2012). Evidence  
72 stratified by gender is also missing, even though differences in health behaviours between  
73 men and women may help to explain gender differences in the complication risk of diabetes  
74 (Kautzky-Willer, Harreiter, Pacini 2016; Huebschmann et al. 2019; Harreiter, Kautzky-Willer  
75 2018; The Lancet Diabetes & Endocrinology 2017; Kautzky-Willer, Harreiter 2017).

76 To assess the impact of a diabetes diagnosis on both employment probabilities and  
77 behavioural risk factors, this study uses longitudinal data from China, a country where about  
78 13% of adults between the age of 40 to 60 have diabetes<sup>1</sup>, and over 50% of those remain  
79 undiagnosed (Wang, Gao, et al. 2017). We take various sources of confounding into account,  
80 first by estimating marginal structural models (MSMs) to account for any time-dependent  
81 confounding (Robins, Hernán, Brumback 2000). Second, we complement this strategy with  
82 fixed effects (FE) models to account for any time-invariant unmeasured confounding. Apart  
83 from this methodological innovation, the study extends the scarce evidence base for the  
84 impact of diabetes on employment in LMICs and provides, as far as we are aware, the first

---

<sup>1</sup>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)

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

## 87 2 Data

88 The China Health and Nutrition Survey (CHNS) is a longitudinal survey providing information  
89 on socioeconomic outcomes, health, health behaviours and nutrition in nine provinces of  
90 China (Zhang et al. 2014). We use data from 1997 onwards (with survey rounds in 1997, 2000,  
91 2004, 2006, 2009 and 2011): 1997 was the first time diabetes information was provided. The  
92 sample is limited to the adult population aged 18–64, is not nationally representative and  
93 the CHNS does not provide sampling weights (Popkin et al. 2010). We exclude students and  
94 women who reported to be pregnant at the time of the survey. Further, due to relatively early  
95 retirement in China for those in formal employment and for women, once people reported to  
96 be retired they were excluded from the sample from this point onwards.

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

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

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

118 rate in relation to her resting or basal metabolic rate.<sup>2</sup> <sup>3</sup>

## 119 **3 Methods**

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

### 129 **3.1 Marginal structural models**

130 MSMs can, contrary to FE models, adjust for confounding and selection bias arising from  
131 time-varying confounders affected by prior exposure to treatment (Robins, Hernán, Brumback  
132 2000).

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

---

<sup>2</sup>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.

<sup>3</sup>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).

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

147 The IPTW are calculated as depicted in the following model:

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

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

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

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

171 To account for the potential of attrition bias, we estimate stabilized censoring weights  
 172 based on the probability to remain uncensored until the end of the panel. The model is similar  
 173 to the IPTW model above, now using as dependent variable a dummy variable indicating  
 174 censoring in the following wave. We then estimate the probability of remaining uncensored  
 175 until the last observation in the individual's panel  $t$  using the covariates  $X$  as described above,

176 additionally accounting for a person’s diabetes history  $\bar{D}_t$ .

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

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

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_0 + u_i \quad (3)$$

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

### 188 3.2 Fixed effects

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

202 We estimate the following FE model

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

203 where  $Y_{it}$  is the respective outcome of interest at time  $t$ ,  $D_{it}$  indicates a diabetes diagnosis  
204 at time  $t$  (or time since diagnosis in our duration analysis),  $X_{it}$  is a vector of control variables  
205 unaffected by prior treatment or outcomes,  $c_i$  represents the individual fixed effect, and  $u_{it}$  is  
206 the error term, which can vary over time and across individuals.  $X_{it}$  includes age squared, the  
207 level of urbanization, education, being married, health insurance, living in a rural area, region  
208 and time dummies as well as per capita household income. We do not use the fixed effects  
209 model to estimate the effect of time since diagnosis, since the increase in time since diagnosis  
210 is not distinguishable from the increase in age or overall time in the FE model (Wooldridge  
211 2012). For the same reason age is excluded from all FE specifications.

### 212 **3.3 Regression method**

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

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

### 221 **3.4 Robustness checks**

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

### 231 **3.5 Multiple imputation**

232 We use imputed data to avoid excluding participants with missing data on one or more variables.  
233 Chained multiple imputation is used to impute thirty data sets under the assumption that the

234 imputed data are missing at random, using the user written ICE command in Stata (Royston,  
 235 White 2009). All outcome and explanatory variables included in the MSM and FE models are  
 236 included in the multiple imputations. Table A1 details the number of missing observations for  
 237 each variable. We do not use multiple imputation for diabetes diagnosis and instead assume  
 238 that after the first reported diagnosis the individual had diabetes in every ensuing wave, even  
 239 when the observation was missing.

## 240 4 Results

241 To describe the distributions of our outcome and control variables at baseline, we report the  
 242 means separately for men and women and for those who did and did not report diabetes  
 243 over the observed period. Table 1 shows that both men and women who went on to report a  
 244 diabetes diagnosis are older, have higher BMI and lower physical activity levels and higher  
 245 rates of hypertension than those in the non-diabetes group. Further, men who report diabetes  
 246 drink more alcohol, live in more urbanized regions and have a higher socioeconomic status as  
 247 measured by education and income levels. Women who report diabetes, however, have lower  
 248 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.

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

251 (Table A3 of the Appendix). For women this is not the case, which suggests that MSM may  
 252 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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
<i>Marginal structural model</i>							
Men							
Diabetes	0.006 (.031)	-.046 (.038)	-.088* (.044)	0.024 (.043)	-.762*** (.200)	-117.299 (69.756)	-11.597 (10.787)
Women							
Diabetes	0.124** (.039)	-.033 (.022)	-.019*** (.006)	0.130*** (.039)	-.383 (.277)	-60.742 (41.220)	-33.855** (11.445)
<i>Fixed effects</i>							
Men							
Diabetes	0.014 (.029)	-.001 (.035)	-.100** (.038)	0.011 (.043)	-.797*** (.200)	-141.949* (69.219)	-1.392 (12.222)
Women							
Diabetes	-.155*** (.040)	-.016 (.012)	-.018 (.014)	0.065 (.040)	-.730** (.222)	-57.988 (58.055)	-33.787* (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.

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

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

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

269 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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
<b>Men</b>							
Time since diagnosis	-.001 (.006)	-.004 (.007)	-.015 (.008)	-.000 (.006)	-.142*** (.031)	-20.134 (11.549)	-1.741 (1.848)
<b>Women</b>							
Time since diagnosis	-.016** (.006)	-.003 (.003)	-.002** (.001)	.011 (.006)	-.058 (.053)	-12.718* (6.222)	-4.096 (2.145)

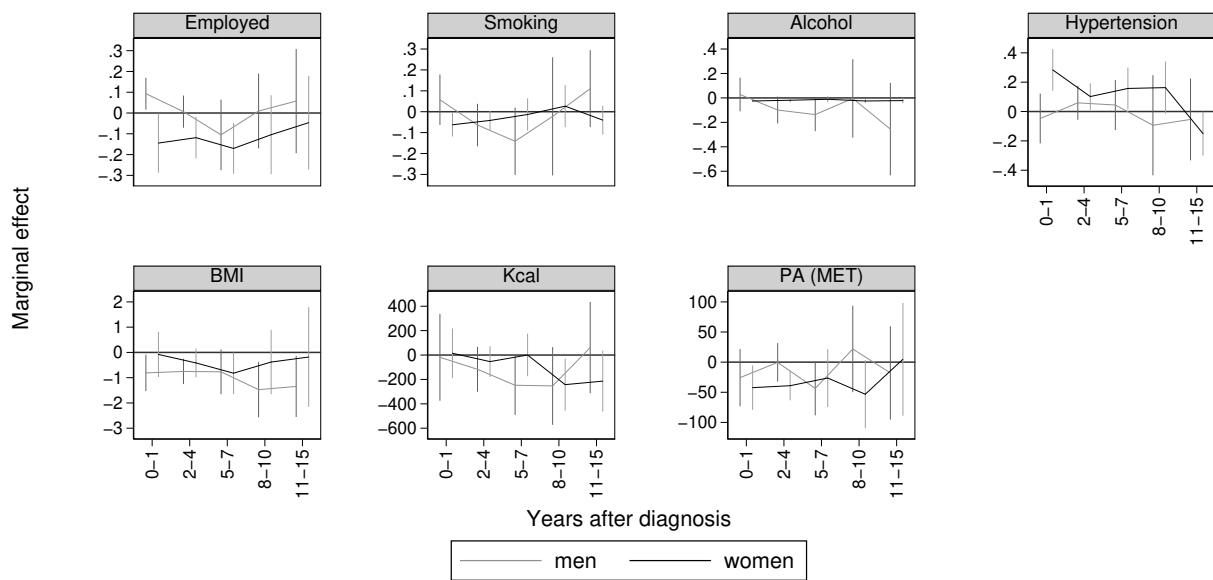
*Note* Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. MSM controls for baseline values of age, age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, 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.

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

280 Using weights that do not account for censoring in the MSM yields very similar results,  
 281 suggesting little bias due to censoring (Table A8, A9 and A10 of the Appendix). Likewise,  
 282 using truncated weights leads to qualitative similar estimates (Table A5, A6 and A7 of the  
 283 Appendix). Finally, estimating the FE model with lagged covariates and a smaller sample,  
 284 gives results very similar to those of the MSM, with a slight reduction in the adverse effect of  
 285 diabetes on employment, and the hypertension risk in women now being adversely affected by  
 286 a diabetes diagnosis (Table A11 of the Appendix). Finally, the results of logistic regression  
 287 models used to re-estimate the effect of a diabetes diagnosis on employment, smoking, alcohol  
 288 consumption and the risk of hypertension support the findings from the linear probability  
 289 models (LPMs), although they are not directly comparable due to the need to present results  
 290 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.

## 291 5 Discussion

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

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

### 302 5.1 Methodological considerations

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

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

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

## 329 **5.2 Potential mechanisms**

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

352 The evidence we find for a worsening of risk factors of women may be explained in  
353 several ways. Generally lower educational attainment and income of women may reduce their  
354 exposure to health information and limit the access to treatment (Luo et al. 2015; Ma, Nolan,  
355 Smith 2018). Women may also receive less spousal support or support of their close network  
356 in the management of their disease, making it more difficult to change health behaviours  
357 (Albanese et al. 2019). Women have also been found to be in a worse metabolic health state  
358 compared to men when crossing the diabetes threshold, with a higher risk of cardiovascular  
359 disease and stroke after diagnosis (Kautzky-Willer, Harreiter, Pacini 2016; Huebschmann

360 et al. 2019; Harreiter, Kautzky-Willer 2018; The Lancet Diabetes & Endocrinology 2017;  
361 Kautzky-Willer, Harreiter 2017). Potentially as a result of these factors, Chinese women with  
362 diabetes experience more comorbidities than men (Liu et al. 2010).

363 This has been the first study to use MSM to explore the impact of a diabetes diagnosis  
364 on employment longitudinally. Previous longitudinal studies used fixed effects models only,  
365 finding reductions in employment probabilities for men and women of about 5 percentage  
366 points in Mexico (Seuring, Serneels, Suhrcke 2019). Taking into account the lower overall  
367 employment rate of Mexican women compared to men, this translated into a 16% reduction  
368 in female employment probabilities, a figure comparable to the effect observed in this study.  
369 Overall, the adverse effect of diabetes on employment is in line with other studies that have  
370 found diabetes to reduce employment probabilities for women (Minor 2011; Latif 2009; Harris  
371 2009; Seuring, Serneels, Suhrcke 2016)—often more than for men. The large gender differences  
372 in the employment impact may, at least partly, be driven by the observed differences in  
373 behaviour change and in risk factors for complications, leading to worse health outcomes in  
374 women that result in a decrease in their employment probabilities. Further, evidence from  
375 Mexico points towards a larger employment penalty of diabetes for those in the informal  
376 labour market (Seuring, Goryakin, Suhrcke 2015). Given the considerable informal sector  
377 in China and the over-representation of women in this sector (Wang, Klugman 2020), it is  
378 possible that women are more exposed to low job security, increasing their chances to be  
379 laid-off due to their diabetes, be it due to actual health problems, or the stigma surrounding  
380 the disease.

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

## 388 6 Acknowledgements

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

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

## 402 References

- 403 Ainsworth, B. E., W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett, C. Tudor-  
404 Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, A. S. Leon (2011), 2011 compendium  
405 of physical activities: A second update of codes and MET values, DOI: 10.1249/MSS.  
406 0b013e31821ece12.
- 407 Albanese, A. M., J. C. Huffman, C. M. Celano, L. M. Malloy, D. J. Wexler, M. E. Freedman,  
408 R. A. Millstein (2019), The role of spousal support for dietary adherence among type 2  
409 diabetes patients: a narrative review, *Social Work in Health Care* 58 (3) 304–323, DOI:  
410 10.1080/00981389.2018.1563846.
- 411 American Diabetes Association (2020), 10. Cardiovascular Disease and Risk Management:  
412 Standards of Medical Care in Diabetes—2020, *Diabetes Care* 43 (Supplement 1) S111–S134,  
413 DOI: 10.2337/dc20-S010.
- 414 Angrist, J., J. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist’s Companion*,  
415 Princeton University Press.
- 416 Bays, H. E., R. H. Chapman, S. Grandy (2007), The relationship of body mass index to  
417 diabetes mellitus, hypertension and dyslipidaemia: Comparison of data from two national  
418 surveys, *International Journal of Clinical Practice* 61 (5) 737–747, DOI: 10.1111/j.1742-  
419 1241.2007.01336.x.
- 420 Brown, H. S., J. A. Pagán, E. Bastida (2005), The Impact of Diabetes on Employment: Genetic  
421 IVs in a Bivariate Probit, *Health Economics* 14 (5) 537–544, DOI: 10.1002/hec.942.
- 422 Cole, S. R., M. A. Hernán (2008), Constructing Inverse Probability Weights for Marginal  
423 Structural Models, *American Journal of Epidemiology* 168 (6) 656–664, DOI: 10.1093/  
424 aje/kwn164.
- 425 Colman, G., D. Dave (2014), *Unemployment and Health Behaviors Over the Business Cycle:  
426 a Longitudinal View*, NBER Working Paper (20748).

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

430 De Fine Olivarius, N., V. D. Siersma, R. Køster-Rasmussen, B. L. Heitmann, F. B. Waldorff  
431 (2015), Weight changes following the diagnosis of type 2 diabetes: The impact of recent and  
432 past weight history before diagnosis. Results from the Danish Diabetes Care in General  
433 Practice (DCGP) Study, *PLoS ONE* 10 (4) 1–14, DOI: 10.1371/journal.pone.0122219.

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

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

438 Harris, A. (2009), Diabetes, Cardiovascular Disease and Labour Force Participation in Aus-  
439 tralia: An Endogenous Multivariate Probit Analysis of Clinical Prevalence Data, *Economic*  
440 *Record* 85 (271) 472–484, DOI: 10.1111/j.1475-4932.2009.00572.x.

441 Huebschmann, A. G., R. R. Huxley, W. M. Kohrt, P. Zeitler, J. G. Regensteiner, J. E. Reusch  
442 (2019), Sex differences in the burden of type 2 diabetes and cardiovascular risk across the  
443 life course, *Diabetologia* 62 (10) 1761–1772, DOI: 10.1007/s00125-019-4939-5.

444 Imai, K., I. S. Kim (2019), When Should We Use Unit Fixed Effects Regression Models for  
445 Causal Inference with Longitudinal Data?, *American Journal of Political Science* 63 (2)  
446 467–490, DOI: 10.1111/ajps.12417.

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

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

452 Kautzky-Willer, A., J. Harreiter, G. Pacini (2016), Sex and gender differences in risk, patho-  
453 physiology and complications of type 2 diabetes mellitus, *Endocrine Reviews* 37 (3) 278–  
454 316, DOI: 10.1210/er.2015-1137.

455 Latif, E. (2009), The impact of diabetes on employment in Canada, *Health Economics* 18 (5)  
456 577–589, DOI: 10.1002/hec.1390.

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

460 Liu, Z., C. Fu, W. Wang, B. Xu (2010), Prevalence of chronic complications of type 2 diabetes  
461 mellitus in outpatients - a cross-sectional hospital based survey in urban China. *Health*  
462 *and quality of life outcomes* 8 (1) 62, DOI: 10.1186/1477-7525-8-62.

463 Long, G. H., I. Johansson, O. Rolandsson, P. Wennberg, E. Fhärm, L. Weinehall, S. J. Griffin,  
464 R. K. Simmons, M. Norberg (2015), Healthy behaviours and 10-year incidence of diabetes:  
465 A population cohort study, *Preventive Medicine* 71 121–127, DOI: 10.1016/j.ypmed.  
466 2014.12.013.

467 Long, G. H., A. J. Cooper, N. J. Wareham, S. J. Griffin, R. K. Simmons (2014), Healthy  
468 Behavior Change and Cardiovascular Outcomes in Newly Diagnosed Type 2 Diabetic  
469 Patients: A Cohort Analysis of the ADDITION-Cambridge Study, *Diabetes Care* 37 (6)  
470 1712–1720, DOI: 10.2337/dc13-1731.

471 Luo, X., T. Liu, X. Yuan, S. Ge, J. Yang, C. Li, W. Sun (2015), Factors Influencing Self-  
472 Management in Chinese Adults with Type 2 Diabetes: A Systematic Review and Meta-  
473 Analysis, *International Journal of Environmental Research and Public Health* 12 (9)  
474 11304–11327, DOI: 10.3390/ijerph120911304.

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

477 Minor, T. (2011), The effect of diabetes on female labor force decisions: new evidence  
478 from the National Health Interview Survey, *Health Economics* 20 (12) 1468–1486, DOI:  
479 10.1002/hec.1685.

480 Ng, S. W., B. M. Popkin (2012), Time use and physical activity: a shift away from movement  
481 across the globe, *Obesity Reviews* 13 (8) 659–680, DOI: 10.1111/j.1467-789X.2011.  
482 00982.x.

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

486 Popkin, B. M., S. Du, F. Zhai, B. Zhang (2010), Cohort profile: The China Health and Nutrition  
487 Survey-monitoring and understanding socio-economic and health change in China, 1989-  
488 2011, *International Journal of Epidemiology* 39 (6) 1435–1440, DOI: 10.1093/ije/dyp322.

489 Robins, J. M., M. Á. Hernán, B. Brumback (2000), Marginal Structural Models and Causal  
490 Inference in Epidemiology, *Epidemiology* 11 (5) 550–560, DOI: 10.1097/00001648-  
491 200009000-00011.

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

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

498 Seuring, T., O. Archangelidi, M. Suhrcke (2015), The Economic Costs of Type 2 Diabetes: A  
499 Global Systematic Review, *PharmacoEconomics* 33 (8) 811–831, DOI: 10.1007/s40273-  
500 015-0268-9.

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

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

506 — (2019), The impact of diabetes on labour market outcomes in Mexico: A panel data and  
507 biomarker analysis, *Social Science & Medicine* 233 252–261, DOI: 10.1016/j.socscimed.  
508 2019.05.051.

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

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

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

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

520 Wang, L., J. Klugman (2020), How women have fared in the labour market with China’s  
521 rise as a global economic power, *Asia and the Pacific Policy Studies* 7 (1) 43–64, DOI:  
522 10.1002/app5.293.

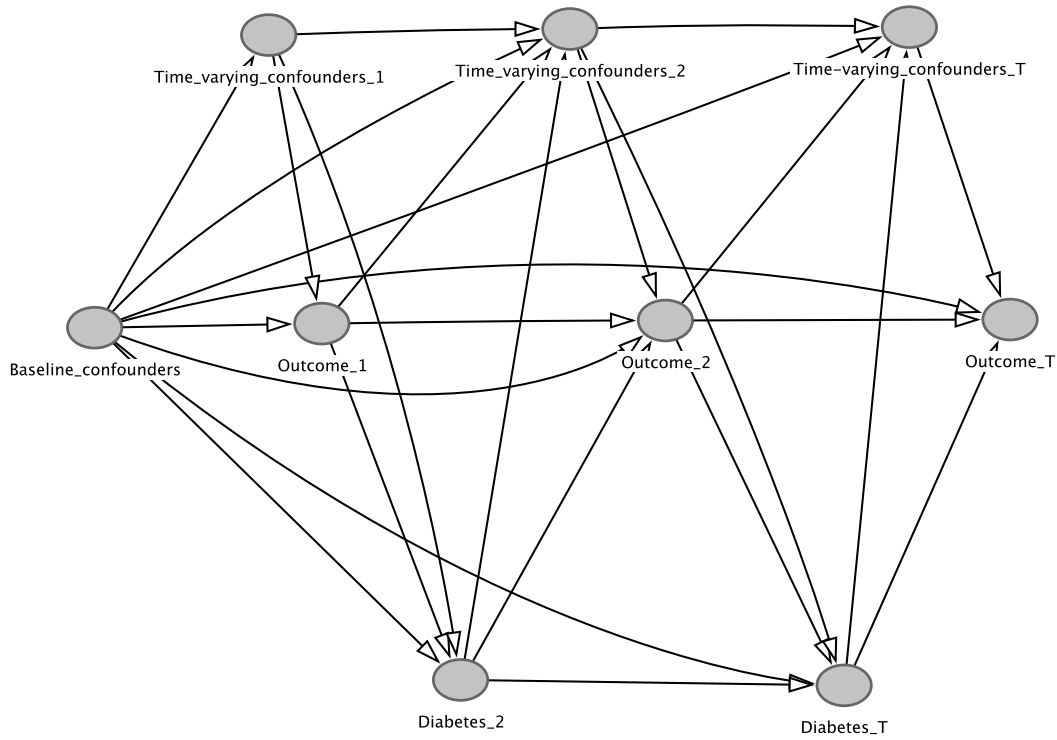
523 Wooldridge, J. (2012), *Introductory Econometrics. A Modern Approach*, 5th ed., Cengage  
524 Learning.

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

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

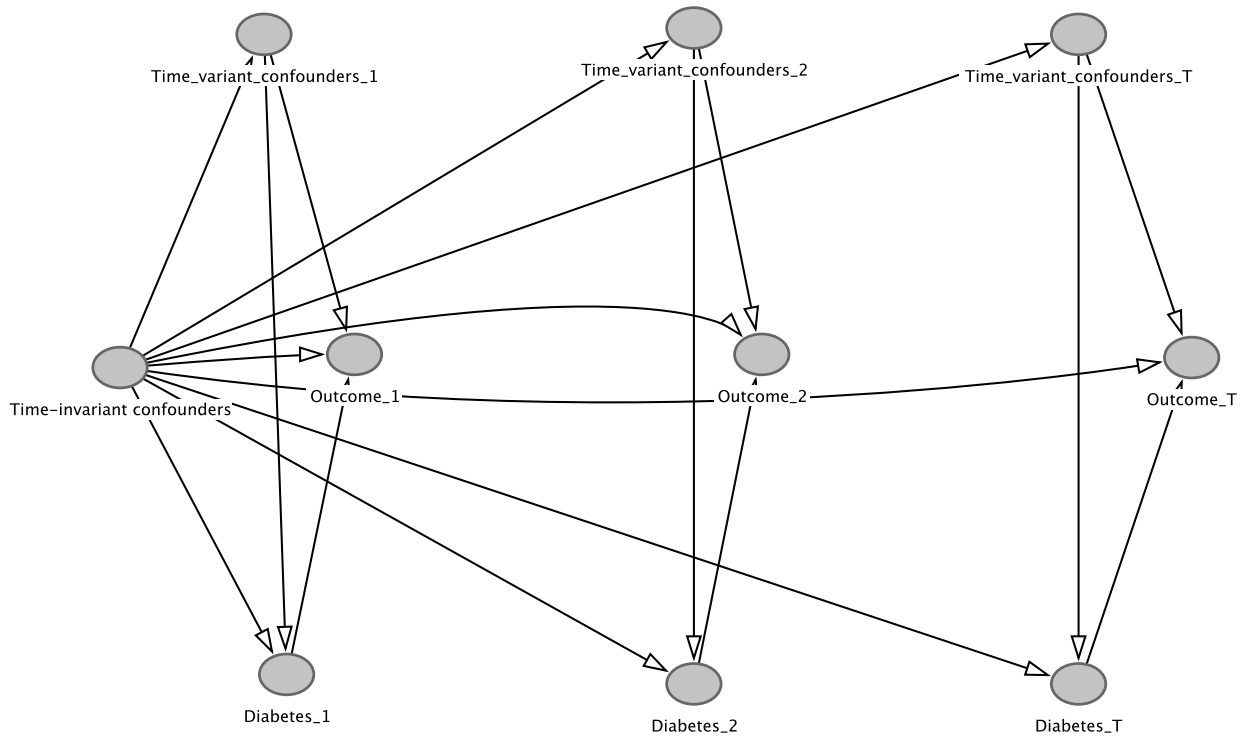
529 Zhou, X., L. Ji, X. Ran, B. Su, Q. Ji, C. Pan, J. Weng, C. Ma, C. Hao, D. Zhang, D. Hu (2016),  
530 Prevalence of Obesity and Its Influence on Achievement of Cardiometabolic Therapeutic  
531 Goals in Chinese Type 2 Diabetes Patients: An Analysis of the Nationwide, Cross-Sectional  
532 3B Study, *PLOS ONE* 11 (1) e0144179, DOI: 10.1371/journal.pone.0144179.

**Figure A1**  
 Direct acyclic graph for the marginal structural model.



*Note* MSMs assume the absence of unobserved time-invariant and 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	
<i>Baseline and time-invariant variables</i>				
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)
<i>Lagged time-varying variables</i>				
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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
Men							
0-1	0.095* (.039)	0.062 (.061)	0.027 (.071)	-.054 (.083)	-.784* (.370)	-40.964 (171.094)	-4.428 (34.371)
2-4	0.004 (.041)	-.066 (.052)	-.106* (.054)	0.044 (.058)	-.715** (.256)	-70.155 (311.593)	-6.245 (20.168)
5-7	-.087 (.076)	-.130 (.080)	-.133 (.071)	0.054 (.087)	-.749 (.448)	-229.235* (107.881)	-32.295 (24.850)
8-10	0.023 (.086)	-.026 (.145)	0.003 (.162)	-.099 (.174)	-1.531** (.544)	-252.410 (167.490)	15.587 (49.932)
11-14	0.064 (.114)	0.103 (.102)	-.272 (.190)	-.063 (.147)	-1.264 (.660)	104.464 (172.888)	-26.757 (47.953)
0-1	0.093* (.039)	0.058 (.062)	0.028 (.070)	-.047 (.086)	-.810* (.368)	-18.598 (181.515)	-25.703 (24.220)
2-4	0.007 (.039)	-.064 (.052)	-.099 (.056)	0.059 (.059)	-.751** (.256)	-116.940 (93.817)	-.326 (16.303)
5-7	-.105 (.086)	-.141 (.082)	-.136* (.069)	0.044 (.087)	-.767 (.451)	-247.440* (124.535)	-43.579 (22.659)
8-10	0.010 (.092)	-.022 (.144)	-.004 (.163)	-.094 (.174)	-1.472** (.559)	-253.240 (162.492)	21.873 (36.522)
11-14	0.057 (.127)	0.110 (.093)	-.255 (.193)	-.054 (.142)	-1.345* (.613)	61.299 (190.266)	-17.822 (39.230)
Women							
0-1	-.145* (.073)	-.062* (.029)	-.025*** (.004)	0.284*** (.073)	-.079 (.457)	14.668 (103.154)	-42.297* (18.782)
2-4	-.118* (.051)	-.041 (.023)	-.019* (.009)	0.102* (.046)	-.415 (.290)	-53.783 (64.346)	-39.018** (12.414)
5-7	-.170** (.062)	-.013 (.039)	-.011 (.013)	0.157* (.072)	-.822* (.419)	0.328 (88.597)	-26.308 (24.410)
8-10	-.104 (.097)	0.027 (.051)	-.025** (.008)	0.163 (.090)	-.381 (.649)	-242.489* (108.205)	-53.214 (28.754)
11-14	-.046 (.115)	-.040 (.035)	-.022 (.011)	-.150* (.076)	-.182 (1.007)	-213.862 (127.050)	4.717 (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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
<b>Men</b>							
Diabetes	-.004 (.030)	-.058 (.036)	-.095** (.036)	0.030 (.040)	-.741*** (.193)	-142.008* (63.428)	-14.485 (10.240)
<b>Women</b>							
Diabetes	-.128*** (.037)	-.030 (.020)	-.019*** (.006)	0.130*** (.038)	-.376 (.272)	-58.374 (40.861)	-34.827** (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)
<b>Men</b>							
Time since diagnosis	-.003 (.006)	-.007 (.006)	-.015* (.007)	.001 (.007)	-.147*** (.031)	-22.087* (11.252)	-2.282 (1.772)
<b>Women</b>							
Time since diagnosis	-.017** (.006)	-.003 (.003)	-.002** (.001)	.011 (.006)	-.056 (.052)	-12.308* (6.213)	-4.200* (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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
<b>Men</b>							
0-1	0.087* (.042)	0.054 (.061)	-.002 (.065)	-.057 (.071)	-.744* (.365)	-128.911 (132.013)	-29.781 (24.819)
2-4	-.004 (.039)	-.081 (.046)	-.118* (.046)	0.073 (.053)	-.691** (.235)	-136.257 (86.271)	-3.093 (15.131)
5-7	-.108 (.080)	-.145 (.077)	-.149* (.062)	0.046 (.082)	-.761 (.416)	-224.930 (119.446)	-42.806* (20.719)
8-10	-.013 (.100)	-.030 (.142)	0.024 (.145)	-.131 (.135)	-1.615** (.565)	-232.071 (156.783)	18.467 (40.230)
11-14	0.039 (.131)	0.105 (.100)	-.172 (.184)	-.018 (.143)	-1.495* (.635)	54.355 (200.571)	-26.721 (40.121)
<b>Women</b>							
0-1	-.146* (.073)	-.060* (.027)	-.025*** (.004)	0.275*** (.071)	-.112 (.450)	18.543 (103.734)	-44.753* (18.404)
2-4	-.124* (.049)	-.039 (.020)	-.019* (.009)	0.108* (.046)	-.400 (.285)	-55.985 (62.172)	-39.865*** (12.099)
5-7	-.174** (.060)	-.009 (.035)	-.011 (.013)	0.153* (.072)	-.771 (.411)	14.742 (88.550)	-26.326 (23.843)
8-10	-.105 (.096)	0.026 (.051)	-.025** (.008)	0.161 (.090)	-.385 (.646)	-244.078* (107.443)	-54.441 (28.467)
11-14	-.047 (.115)	-.040 (.035)	-.022 (.011)	-.151* (.076)	-.180 (1.007)	-213.686 (126.873)	4.424 (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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
Men							
Diabetes	-.007 (.032)	-.041 (.038)	-.088 (.046)	0.038 (.041)	-.687*** (.200)	-124.116 (69.861)	-15.568 (10.748)
Women							
Diabetes	-.135*** (.036)	-.030 (.019)	-.020*** (.006)	0.122** (.037)	-.365 (.275)	-57.607 (41.479)	-39.302*** (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)
<b>Men</b>							
Time since diagnosis	-.003 (.006)	-.003 (.006)	-.016 (.009)	.002 (.006)	-.133*** (.030)	-19.610 (11.780)	-2.166 (1.857)
<b>Women</b>							
Time since diagnosis	-.019** (.006)	-.003 (.003)	-.002** (.001)	.012* (.006)	-.055 (.053)	-12.494* (6.213)	-5.087* (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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
<b>Men</b>							
0-1	0.080 (.048)	0.057 (.061)	0.035 (.071)	-.042 (.079)	-.606 (.412)	-57.745 (158.819)	-31.227 (24.953)
2-4	-.011 (.040)	-.065 (.049)	-.104 (.058)	0.077 (.054)	-.684** (.247)	-131.817 (98.646)	-5.237 (16.973)
5-7	-.101 (.086)	-.133 (.083)	-.120 (.072)	0.041 (.084)	-.807 (.421)	-232.150 (129.047)	-41.588 (23.065)
8-10	-.006 (.094)	0.015 (.136)	-.012 (.172)	-.037 (.179)	-1.324* (.554)	-227.770 (162.758)	18.969 (37.833)
11-14	0.053 (.118)	0.120 (.086)	-.274 (.195)	-.051 (.142)	-1.265* (.586)	96.735 (168.556)	-19.427 (36.253)
<b>Women</b>							
0-1	-.152* (.069)	-.059* (.025)	-.025*** (.005)	0.244*** (.074)	-.060 (.435)	30.656 (105.419)	-49.597** (18.811)
2-4	-.121* (.050)	-.039* (.019)	-.021* (.009)	0.094* (.045)	-.367 (.293)	-60.189 (62.349)	-41.364** (12.584)
5-7	-.194** (.061)	-.005 (.035)	-.012 (.012)	0.155* (.072)	-.874* (.429)	19.858 (88.260)	-31.087 (24.872)
8-10	-.123 (.097)	0.026 (.053)	-.026** (.009)	0.180* (.088)	-.442 (.632)	-262.560* (108.815)	-62.185* (28.959)
11-14	-.066 (.116)	-.041 (.036)	-.021 (.011)	-.145 (.075)	-.009 (1.021)	-208.206 (118.884)	-7.337 (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 <sup>2</sup> )	Calories (kcal)	Physical activity (hours/week)
Men							
Diabetes	0.054 (.035)	0.004 (.045)	-.069 (.049)	0.014 (.050)	-.830*** (.227)	-181.109* (86.015)	-2.425 (16.033)
Women							
Diabetes	-.132* (.058)	-.011 (.009)	-.010 (.015)	0.160** (.057)	-.672* (.294)	-38.070 (76.524)	-53.022* (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)	(2)	(3)	(4)
	Employment	Smoking	Any alcohol	Hypertension
<i>Marginal structural models</i>				
Men				
Diabetes	1.062 (.273)	0.775 (.162)	0.613 (.158)	1.063 (.241)
Women				
Diabetes	0.561** (.106)	0.306 (.190)	0.212* (.156)	1.674** (.326)
<i>Fixed effects</i>				
Men				
Diabetes	1.327 (.458)	1.046 (.322)	0.482** (.130)	0.922 (.245)
Women				
Diabetes	0.293** (.121)	0.212 (.276)	0.320 (.293)	1.313 (.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.