I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11817

Diabetes, Employment and Behavioural Risk Factors in China: Marginal Structural Models versus Fixed Effects Models

Till Seuring
Pieter Serneels
Marc Suhrcke
Max Bachmann

SEPTEMBER 2018

DISCUSSION PAPER SERIES

IZA DP No. 11817

Diabetes, Employment and Behavioural Risk Factors in China: Marginal Structural Models versus Fixed Effects Models

Till Seuring
BIPS, University of Bremen and University of East Anglia

Pieter Serneels
University of East Anglia and IZA

Marc Suhrcke
University of York and Luxembourg Institute of Socio-Economic Research

Max Bachmann
University of East Anglia

SEPTEMBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Diabetes, Employment and Behavioural Risk Factors in China: Marginal Structural Models versus Fixed Effects Models

A diabetes diagnosis can motivate its recipients to reduce their health risks by changing lifestyles but can adversely affect their economic activity. We investigate the effect of a diabetes diagnosis on employment status and behavioural risk-factors taking into account their potentially intertwined relationships. Longitudinal data from the China Health and Nutrition Survey covering the years 1997 to 2011 are used to estimate the effect of a diabetes diagnosis on employment probabilities, alcohol consumption, smoking cessation, body mass index, physical activity and hypertension. To deal with potential confounding, two complementary statistical techniques - marginal structural and fixed effects models - are applied. The marginal structural and fixed effects models generate similar results despite their different underlying assumptions. Both strategies find patterns distinct for males and females, suggesting a decrease in employment probabilities after the diagnosis for women but not for men. Further, few improvements and even further deterioration of behavioural risk factors are found for women, while for men these risk factors either improve or remain the same. These results suggest differences in the impact of diabetes between sexes in China and highlight the potential of reducing behavioural risk factors for women to narrow these inequities.

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

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

Corresponding author:
Till Seuring
Leibniz Institute for Prevention Research and Epidemiology (BIPS)
Research Group for Evidence-Based Public Health
Achterstr. 30
28359 Bremen
Germany
E-mail: seuring@leibniz-bips.de

1 Introduction

The effect of diabetes on employment status has received little attention in low- and middle-income countries (LMICs) (Seuring, Archangelidi, & Suhrcke, 2015), despite its high prevalence in the working age population. Once diagnosed, the severity of diabetes, diabetes complications and their potential economic effects strongly depend on the patient's behaviour, which is itself likely affected by the development of diabetes in the first place. Research shows that behaviour changes after a diabetes diagnosis can positively affect health and the risk of subsequent cardiovascular events (Long, Cooper, Wareham, Griffin, & Simmons, 2014; Zhou et al., 2016). Thus, a diabetes diagnosis may present an important opportunity to reduce risk factors for diabetes complications (De Fine Olivarius, Siersma, Køster-Rasmussen, Heitmann, & Waldorff, 2015) and the related economic burden, raising the question of the current effect 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. So has unemployment been found to lead to weight gain but also to reduce smoking and fast-food consumption (Colman & Dave, 2014).

Despite that, 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 (Seuring, Serneels, & Suhrcke, 2016). Similarly, studies investigating behaviour change after a diabetes diagnosis are scarce, and have not accounted for the selection into a diabetes diagnosis based on prior behaviour change (Slade, 2012).

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, and over 50% of those remain undiagnosed (Wang 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, Hernan, & 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 longitudinal evidence for the effect of a diabetes diagnosis on behavioural risk factors in any LMICs country.

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, Zhai, Du, & Popkin, 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, Du, Zhai, & Zhang, 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, we exclude those who retired before age 65.

Because both the MSM and FE use changes in the treatment for identification, we only use incident cases of self-reported diabetes to construct our diabetes indicator, excluding individuals with self-reported diabetes at baseline. Given the chronic nature of diabetes, we assume that after diagnosis diabetes persists for the rest of one's life. To construct a measure of diabetes duration for incident cases we use self-reported information on the year of diagnosis.

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

The behavioural risk factor outcomes are binary variables for current smoking status, if alcohol was consumed equal to or more than three times per week and if 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 body mass index (BMI), daily calorie consumption and overall physical activity levels. We chose these outcomes as they present some of the most important risk-factors for diabetes and diabetes related complications. BMI is based on height and weight measurements, daily calorie consumption is based on the self-reported average daily consumption of carbohydrates, protein and fat of every individual, measured on three consecutive days, and was calculated by the CHNS investigators. Physical activity includes activity related to occupation, leisure, travel to work and homework and is expressed in metabolic equivalent of task (MET) hours per week. We followed the Compendium of Physical Activities (Ainsworth et al., 2011) to assign activity levels and the previous literature on calculating physical activity levels using the CHNS (Ng & Popkin, 2012; Ng, Norton, & Popkin, 2009).

3 Methods

3.1 Marginal structural models

MSMs can, contrary to FE models, adjust for confounding and selection bias as a result of time-varying confounders being affected by prior exposure to the treatment, using inverse probability of treatment weighting (Robins et al., 2000).[1]

This requires the estimation of inverse probability of treatment weights (IPTW) that are the inverse of the conditional probability of receiving a treatment given the past treatment and covariate history. 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 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. that have been diagnosed at time t, are given a weight of $\frac{1}{p}$, which assigns 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 $\frac{1}{1-p}$, which assigns higher weights to persons with higher probabilities and lower weights to those with lower probabilities. This allows for the creation of a pseudo population exchangeable with the study population within the levels of confounders (Cole & Hernan, 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, time-variant confounders lagged by one period and time-invariant confounders

[1] This relationship is typically presented in a causal directed acyclic graph (DAG) shown in Figure A1 of Supporting Information, which displays the association between confounders, outcomes and the treatment variable, in our case a diabetes diagnosis.

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 wave. Then IPTW from each wave after the baseline are multiplied with the IPTW from all previous waves to create IPTW reflecting cumulative probabilities over time.

To reduce the variance of the IPTW, the numerator of Eq. 1 consists of an additional set of weights using only baseline values of the predictors as covariates. The result of calculating Eq. 1 are stabilized IPTW that only reflect confounding due to the time-varying covariates, which cannot be appropriately adjusted for by standard regression models (Cole & Hernan, 2008). Because our analysis is stratified by males and females, we create separate weights for each gender.

After the creation of the stabilized IPTW we 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 IPTW:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_0 + u_i \tag{2}$$

where Y_i represents the respective outcome variable, D_i is a binary variable indicating a diabetes diagnosis, 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.

3.2 Fixed effects

In contrast to the MSM, the FE model accounts for time-invariant unobserved confounders[2], relying on within-person variation for identification. This comes at a price: 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, 2016). Additionally, only confounders unaffected by a diabetes diagnosis can be included as control variables, as these would otherwise capture part of the causal effect

[2]See DAG in Figure A2 of Supporting Information.

of diabetes on the outcome of interest (Angrist & Pischke, 2009), so that we do not control for alcohol, smoking, BMI, physical activity, calorie consumption or hypertension in any FE model.

We estimate the following FE model

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 X_{it} + c_i + u_{it} \tag{3}$$

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.[3]

3.3 Regression method

For our analysis we use linear regression models to estimate effects throughout, including for binary outcomes, in order to increase comparability between the FE and the MSM and their ability to estimate 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 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

Because we use untruncated stabilized weights in our primary analysis of the MSMs as they did not exhibit extreme values (see Table A2 of Supporting Information), as a robustness check we truncate weights at the 1st and 99th percentiles to investigate the sensitivity of the MSM to the most extreme weights. While untruncated weights provide unbiased estimates under the assumptions of the MSM, they may not be the most efficient and tend to have larger standard errors (Cole & Hernan, 2008). Further, we test if increasing the comparability of the FE model to the MSM affects our results and reestimate the FE model using covariates

[3]For the estimation of the effect of time since diagnosis we have to rely on the presence of people without diabetes in the sample, for which diabetes duration does not increase at the same rate as time. Otherwise, the effect of an additional year since diagnosis could not be separately identified as it increases at the same rate between waves as the included time dummies (Wooldridge, 2012). For the same reason age is excluded from all FE specifications.

lagged by one period, and including all other outcome variables as covariates, using the same sample as the MSM.

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, with the user written ICE command in Stata 15 (Royston & White, 2009). All outcome and explanatory variables included in the MSM and FE models are included in the multiple imputations. Table A1 of Supporting Information 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 calculate means stratified into men and women and further into those that did not report diabetes in a later wave and those that did. Table 1 shows that men and women that went on to report a diabetes diagnosis are older, have higher BMI and lower physical activity (PA) levels and higher rates of hypertension than those in the non-diabetes group. Further, men in the diabetes group drink more alcohol, live in more urbanized regions and have a higher socioeconomic status as measured by education and income levels. Women in the diabetes group, however, have lower education levels and are less likely to be employed at baseline.

The calculation of the stabilized weights for the MSM indicates that in particular for men changes in employment, alcohol consumption and smoking predict a diagnosis of diabetes (Table A3 of Supporting Information). For women this is not the case suggesting that in particular for men the use of the MSM may help to reduce bias due to time-variant confounding.

The regression results in Table 2 show reductions of similar size for female employment probabilities due to a diabetes diagnosis in all models. For males no effects are found.

Looking at behavioural risk factors, lower alcohol consumption but not smoking are affected by a diabetes diagnosis in men. Further, BMI was reduced in the MSM and in the FE model in both sexes. For PA and the risk of hypertension, we find some evidence of women reducing their PA 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

Table 1: Sample baseline means for males and females, by diabetes status

	Males			Females		
	No diabetes	Diabetes	p-value (t-test)	No diabetes	Diabetes	p-value (t-test)
Employed	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	24.81	<0.001	22.42	25.85	<0.001
Physical activity (MET)	183.53	150.07	0.003	205.43	188.53	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	0.48	<0.001	0.23	0.21	0.301
Urbanization index	53.94	64.14	<0.001	53.93	51.18	0.021
Per capita household income (2011 Yuan)	5178.46	6086.41	0.014	5066.67	4790.30	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.

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			
Male sample							
Diabetes	−0.004	−0.046	−0.088**	0.031	−0.750***	−133.163*	−12.147
	(0.029)	(0.037)	(0.040)	(0.042)	(0.209)	(70.357)	(13.647)
Female sample							
Diabetes	−0.153***	−0.019	−0.017***	0.123***	−0.355	−78.717**	−41.793***
	(0.035)	(0.012)	(0.006)	(0.037)	(0.255)	(38.392)	(12.549)
				Fixed effects			
Male sample							
Diabetes	0.020	−0.001	−0.100***	0.007	−0.841***	−145.990**	11.512
	(0.029)	(0.035)	(0.038)	(0.040)	(0.211)	(72.276)	(15.797)
Female sample							
Diabetes	−0.157***	−0.014	−0.019	0.066	−0.709***	−69.394	−29.508*
	(0.040)	(0.011)	(0.015)	(0.040)	(0.229)	(55.122)	(16.785)

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 expenditures. 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 (male), N=16252 (female). Sample size for FE models: N=22319 (male), N=21913 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

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 female employment probabilities and PA levels, and potentially an increase on the risk of hypertension, but also small decreases in BMI and caloric consumption. The FE model only indicates a reduction in female employment probabilities and BMI levels. For males, only alcohol consumption and BMI are reduced using both estimation strategies.

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

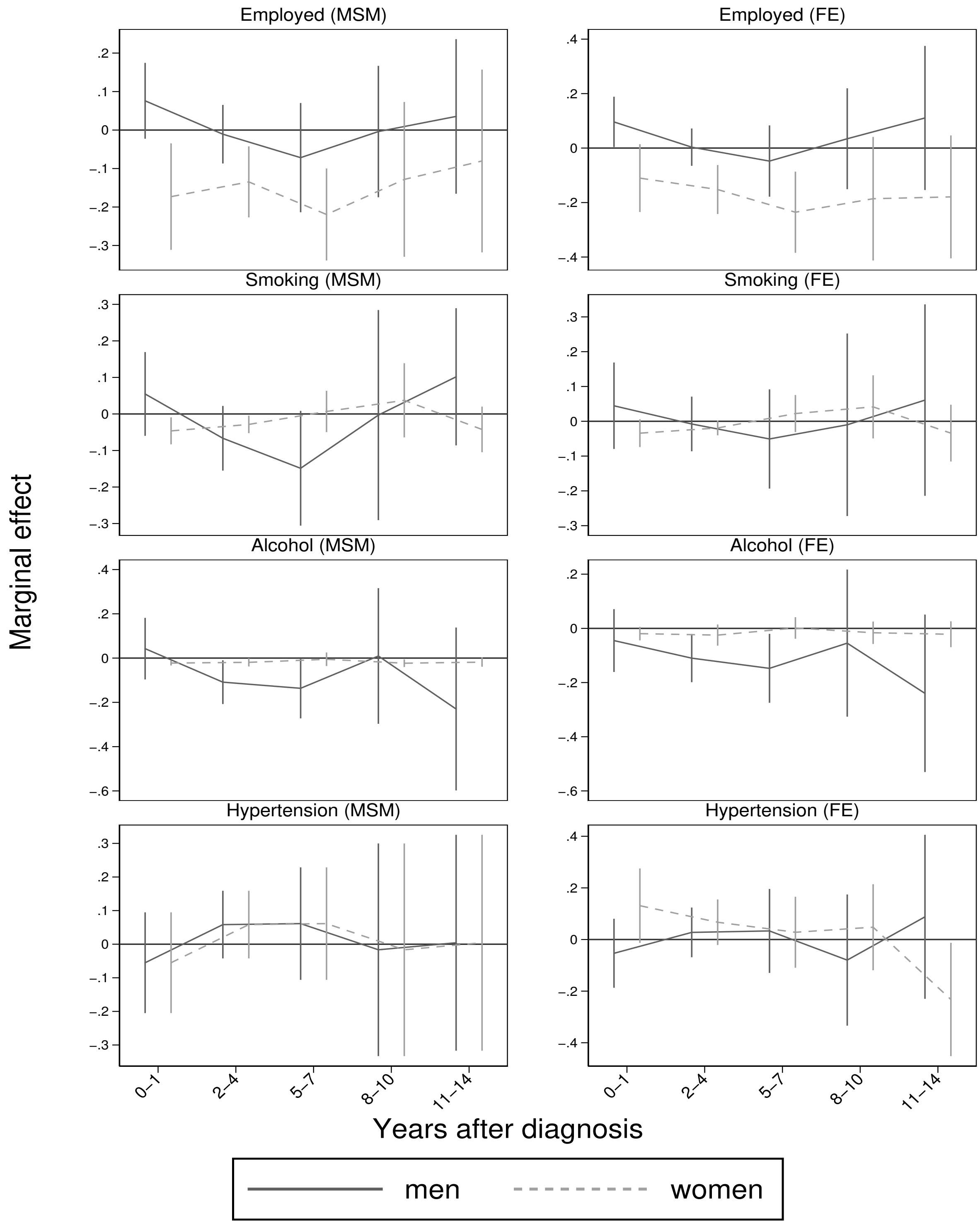
	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
				Marginal structural model			
Male sample							
Time since diagnosis	-0.002	-0.005	-0.015^{**}	0.004	-0.134^{***}	-19.997	-2.100
	(0.005)	(0.006)	(0.007)	(0.007)	(0.034)	(12.234)	(2.340)
Female sample							
Time since diagnosis	-0.021^{***}	-0.002	-0.002^{**}	0.013^{**}	-0.050	-16.496^{***}	-5.518^{**}
	(0.007)	(0.002)	(0.001)	(0.006)	(0.048)	(5.735)	(2.533)
				Fixed effects			
Male sample							
Time since diagnosis	0.002	0.000	-0.017^{**}	0.002	-0.186^{***}	-19.781	2.792
	(0.007)	(0.006)	(0.007)	(0.008)	(0.041)	(13.256)	(3.124)
Female sample							
Time since diagnosis	-0.022^{***}	0.000	-0.001	-0.003	-0.091^{**}	-11.006	-2.660
	(0.008)	(0.002)	(0.002)	(0.006)	(0.046)	(8.222)	(3.223)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other control variables for FE: Age squared, region, urban, education, Han ethnicity, marital status, urbanization index, time dummies, health insurance status, household income. 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 (male), N=16252 (female). Sample size for FE models: N=22319 (male), N=21913 (female). * p<0.10, ** p<0.05, *** p<0.01.

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 Figures 1 and 2 and presented in Tables A4 and A5 of Supporting Information. Both estimation methods indicate a reduction in female employment probabilities in at least the first eight years after diagnosis. Further, both show consistent reductions in male, and to a lesser extent, in female BMI. For physical activity, the MSM indicates a consistent reduction for females over the first ten years after diagnosis, the FE model shows a similar pattern, however, the effects are not statistically significant. No consistent associations over time were found for the other risk factors.

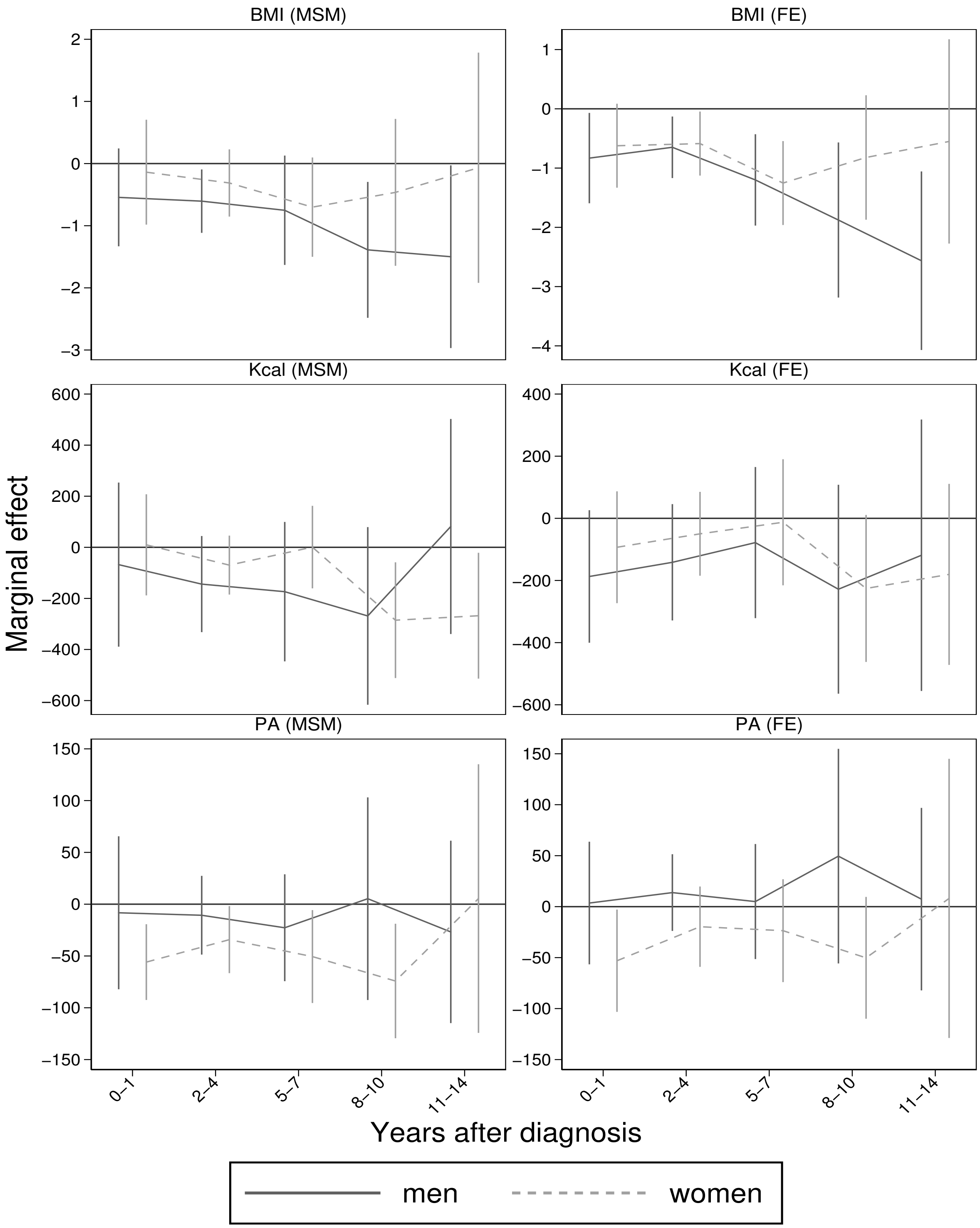
Using truncated instead of untruncated weights for the MSM indicates very similar effects, suggesting no important influence of extreme weights in our MSMs (Table A6 and A7 of Supporting Information). Similarly, testing the robustness of the FE model to the use of lagged covariates and a smaller sample does not lead to qualitative changes in the results, apart from alcohol consumption of men which is no longer affected by a diabetes diagnosis, and the hypertension risk in women which now is adversely affected by a diabetes diagnosis (Tables A8, A9 and A10 of Supporting Information).

Figure 1: The effect of time since diabetes diagnosis on employment, smoking, alcohol consumption and hypertension (duration groups)



Note The visualized coefficients are based on the results of the regression models shown in Tables A4 and A5. The bars indicate 95% confidence intervals. The coefficients present marginal effects compared to baseline.

Figure 2: The effect of time since diabetes diagnosis on BMI, calorie consumption and physical activity (duration groups)



Note The visualized coefficients are based on the results of the regression models shown in Tables A4 and A5. 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 in China a diabetes diagnosis lead to a strong and lasting reduction in female, but not male, employment probabilities. At the same time, men and 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 PA levels and hypertension risk also over time, contrary to women who reduce PA levels after diagnosis and may also experience an increased risk of hypertension.

5.1 Methodological considerations

The MSMs and FE models overall indicate similar results. 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 nonetheless is that the potential source of bias remains unknown and therefore the estimates may not be interpreted as causal. However, given the very similar results of both estimation strategies, we believe that the results very strongly suggest that women are more adversely affected by diabetes than are men. Additionally, with the used methodologies we are not able to assess in how far positive changes in behavioural outcomes contribute to improved diabetes and consequently economic outcomes.

5.2 Potential mechanisms

The results regarding weight loss after a diabetes diagnosis are consistent with other studies. Slade found reductions in overweight and obesity after a diabetes diagnosis, however only over the short term (Slade, 2012). Our results using BMI indicated that weight loss might be reached 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). The decline was attributed to motivation changes as a result of the diabetes diagnosis, suggesting that the diagnosis may represent a window of opportunity to obtain long lasting weight reductions. This may also be the case here, though weight reductions may also be—at least partly—the result of treatment initiation with diabetes drugs causing weight loss (Yang & Weng, 2014).

The worsening of the other risk factors in women after diagnosis could have several reasons, including their lower educational attainment and lower income levels, limiting the access to treatment and reducing exposure to health information (Luo et al., 2015). Moreover, women have 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; Peters, Huxley, Sattar, & Woodward, 2015; Peters, Huxley, & Woodward, 2014b, 2014a; Bertram & Vos, 2010). These factors may help explain the greater burden of comorbidities in Chinese women compared to men (Liu, Fu, Wang, & Xu, 2010), and may also be a contributing factor to the reduced levels of physical activity in women, their increased risk for hypertension as well as the reduction in their employment probabilities.

Given the high prevalence of undiagnosed diabetes, an earlier diagnosis may be a good way to foster early behaviour change leading to more positive health and economic outcomes for people with diabetes over time. However, greater emphasis needs to be placed on women to reduce the observed inequities in the impact of diabetes. Future research should try to unravel the mechanisms behind these differential outcomes for men and women, investigating more formally whether differences in behavioural risk factors could be a potential explanation.

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., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., ... Leon, A. S. (2011). 2011 compendium of physical activities: A second update of codes and MET values. doi:10.1249/MSS.0b013e31821ece12

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

Bertram, M. Y. & Vos, T. (2010). Quantifying the duration of pre-diabetes. *Australian and New Zealand Journal of Public Health*, *34*, 311–314. doi:10.1111/j.1753-6405.2010.00532.x

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

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

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

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

Imai, K. & Kim, I. S. (2016). When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data? Retrieved from http://imai.princeton.edu/research/files/FEmatch.pdf

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

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

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

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

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

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

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

Peters, S. A. E., Huxley, R. R., Sattar, N., & Woodward, M. (2015). Sex Differences in the Excess Risk of Cardiovascular Diseases Associated with Type 2 Diabetes: Potential Explanations and Clinical Implications. *Current Cardiovascular Risk Reports*, *9*, 1–7. doi:10.1007/s12170-015-0462-5

Peters, S. A. E., Huxley, R. R., & Woodward, M. (2014a). Diabetes as a risk factor for stroke in women compared with men: A systematic review and meta-analysis of 64 cohorts, including 775 385 individuals and 12 539 strokes. *The Lancet*, *383*, 1973–1980. doi:10.1016/S0140-6736(14)60040-4

Peters, S. A. E., Huxley, R. R., & Woodward, M. (2014b). Diabetes as risk factor for incident coronary heart disease in women compared with men: A systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. *Diabetologia*, *57*, 1542–1551. doi:10.1007/s00125-014-3260-6

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

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

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

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

Seuring, T., Goryakin, Y., & Suhrcke, M. (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., Serneels, P., & Suhrcke, M. (2016). The Impact of Diabetes on Labor Market Outcomes in Mexico: A Panel Data and Biomarker Analysis. *IZA Discussion Papers*. Discussion Paper Series. Retrieved from http://ftp.iza.org/dp10123.pdf

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

Wang, L., Gao, P., Zhang, M., Huang, Z., Zhang, D., Deng, Q., . . . Wang, L. (2017). Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. *Jama*, *317*, 2515. doi:10.1001/jama.2017.7596

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

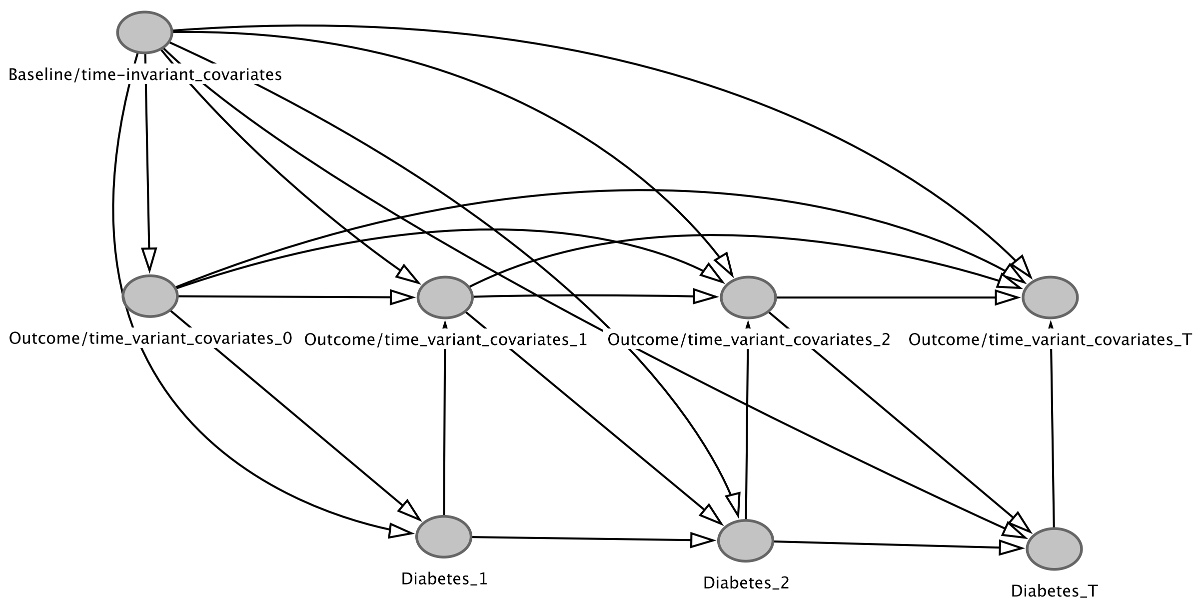
Yang, W. & Weng, J. (2014). Early therapy for type 2 diabetes in China. *The Lancet Diabetes & Endocrinology*, *2*, 992–1002. doi:10.1016/S2213-8587(14)70136-6

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

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

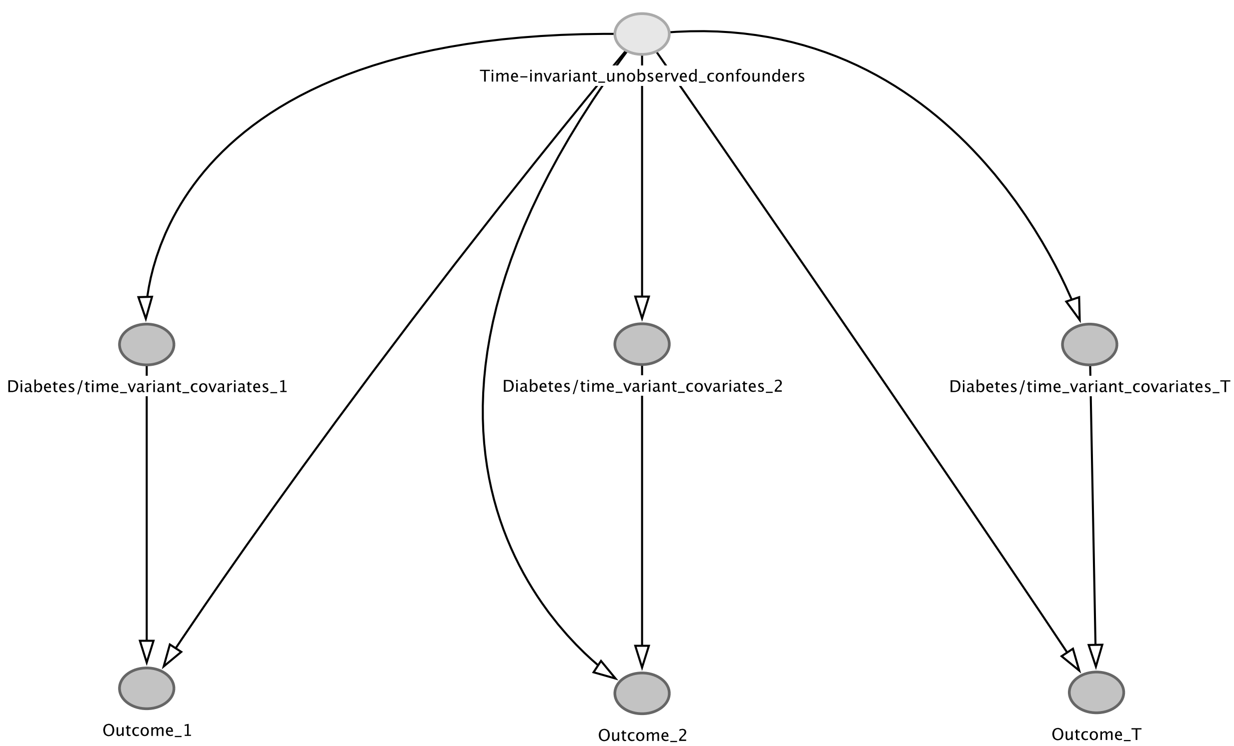
Appendix

Figure A1: Direct acyclic graph for the marginal structural model



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

Figure A2:   Direct acyclic graph for the fixed effects model



Note FE models account for time-invariant unobserved confounding (light grey circle), 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.

Missing data

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

Stabilized weights

Table A2: Summary of stabilized weights

	Mean	Minimum	Maximum
Untruncated (men)	1.001	0.071	3.222
Untruncated (women)	1.000	0.248	2.935
Truncated 1 and 99 percentile (men)	1.000	0.124	3.014
Truncated 1 and 99 percentile (women)	1.000	0.345	1.864

Predicting diabetes

Table A3: Time variant and invariant predictors of a diabetes diagnosis (denominator of stabilized weights): logistic regression models

	Males		Females	
Baseline and time-invariant variables				
Age	0.767**	(0.087)	1.295	(0.214)
Age squared	1.004***	(0.001)	0.998	(0.002)
Urbanization index	1.002	(0.013)	1.008	(0.015)
BMI	1.231***	(0.060)	1.217***	(0.067)
3-Day Ave: Energy (kcal)	1.000	(0.000)	1.000	(0.000)
Smoking	1.380	(0.351)	1.039	(0.850)
Alcohol consumption	1.519*	(0.349)	1.541	(1.136)
Secondary or higher education	0.626	(0.249)	0.567	(0.259)
Married	1.076	(0.546)	0.987	(0.559)
Any health insurance	1.279	(0.322)	0.985	(0.301)
Employed	1.997	(0.846)	1.727*	(0.553)
Per capita household income (2011 Yuan)	1.000	(0.000)	1.000	(0.000)
Hypertension (biomarker)	0.992	(0.260)	1.674*	(0.461)
Physical activity (MET)	0.999*	(0.001)	0.999	(0.001)
Survey year				
2004	1.311	(0.518)	0.670	(0.216)
2006	1.262	(0.527)	0.481*	(0.184)
2009	2.357**	(1.007)	0.800	(0.317)
2011	0.923	(0.453)	0.845	(0.378)
Lagged time-varying variables				
Age	1.658***	(0.257)	0.927	(0.157)
Age squared	0.995***	(0.002)	1.001	(0.002)
BMI	0.983	(0.047)	1.024	(0.055)
Urbanization index	1.018	(0.013)	0.997	(0.014)
3-Day Ave: Energy (kcal)	1.000	(0.000)	1.000	(0.000)
Smoking	0.586**	(0.142)	0.929	(0.755)
Alcohol consumption	0.641*	(0.158)	0.980	(0.795)
Secondary or higher education	1.642	(0.677)	2.363**	(1.034)
Married	1.011	(0.506)	0.853	(0.415)
Any health insurance	1.165	(0.286)	1.042	(0.316)
Employed	0.487***	(0.135)	0.696	(0.192)
Physical activity (MET)	1.000	(0.001)	1.000	(0.001)
Hypertension (biomarker)	1.260	(0.301)	1.193	(0.315)
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. N=16439 (male), N=16113 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Duration groups results

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)
Male sample							
0–1	0.076	0.055	0.043	−0.055	−0.544	−67.765	−8.269
	(0.050)	(0.058)	(0.071)	(0.077)	(0.402)	(163.832)	(37.673)
2–4	−0.011	−0.067	−0.109**	0.058	−0.605**	−144.008	−10.663
	(0.039)	(0.045)	(0.050)	(0.051)	(0.260)	(95.912)	(19.375)
5–7	−0.072	−0.149*	−0.136**	0.061	−0.751*	−173.575	−22.736
	(0.072)	(0.080)	(0.069)	(0.085)	(0.448)	(138.711)	(26.226)
8–10	−0.004	−0.003	0.010	−0.017	−1.389**	−268.656	5.304
	(0.087)	(0.147)	(0.156)	(0.161)	(0.557)	(177.255)	(49.887)
11–14	0.035	0.102	−0.230	0.004	−1.499**	81.493	−26.759
	(0.102)	(0.095)	(0.188)	(0.164)	(0.746)	(213.478)	(44.666)
Female sample							
0–1	−0.173**	−0.046**	−0.022***	0.239***	−0.138	9.740	−55.904***
	(0.071)	(0.019)	(0.006)	(0.073)	(0.430)	(100.961)	(18.654)
2–4	−0.135***	−0.029**	−0.019**	0.094**	−0.312	−69.653	−34.203**
	(0.047)	(0.012)	(0.010)	(0.047)	(0.276)	(58.846)	(16.509)
5–7	−0.219***	0.007	−0.005	0.157**	−0.701*	0.846	−50.572**
	(0.061)	(0.029)	(0.016)	(0.071)	(0.407)	(82.297)	(22.856)
8–10	−0.128	0.037	−0.023***	0.189**	−0.464	−285.322**	−74.143***
	(0.102)	(0.052)	(0.009)	(0.093)	(0.603)	(115.193)	(28.115)
11–14	−0.080	−0.042	−0.018*	−0.126	−0.067	−267.810**	5.430
	(0.121)	(0.032)	(0.011)	(0.079)	(0.945)	(125.235)	(66.142)

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 expenditures, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (male), N=16252 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A5: The effect of time since diabetes diagnosis on employment status and behavioural outcomes using FE (duration groups)

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Male sample							
0–1	0.096**	0.045	−0.045	−0.053	−0.832**	−187.105*	3.558
	(0.047)	(0.063)	(0.059)	(0.068)	(0.388)	(108.694)	(30.675)
2–4	0.003	−0.008	−0.110**	0.028	−0.649**	−141.444	13.798
	(0.035)	(0.040)	(0.045)	(0.049)	(0.264)	(95.451)	(19.196)
5–7	−0.048	−0.051	−0.148**	0.033	−1.200***	−78.052	5.005
	(0.067)	(0.073)	(0.065)	(0.083)	(0.392)	(123.931)	(28.780)
8–10	0.034	−0.010	−0.054	−0.080	−1.876***	−228.164	49.552
	(0.094)	(0.134)	(0.138)	(0.129)	(0.666)	(171.342)	(53.691)
11–14	0.111	0.061	−0.240	0.088	−2.563***	−118.726	7.368
	(0.135)	(0.140)	(0.148)	(0.162)	(0.764)	(222.198)	(45.312)
Female sample							
0–1	−0.110*	−0.034*	−0.019	0.131*	−0.623*	−93.052	−53.050**
	(0.064)	(0.020)	(0.013)	(0.074)	(0.361)	(91.793)	(25.597)
2–4	−0.152***	−0.019*	−0.025	0.067	−0.587**	−49.764	−19.664
	(0.046)	(0.011)	(0.020)	(0.045)	(0.275)	(68.814)	(20.100)
5–7	−0.236***	0.022	0.001	0.028	−1.252***	−12.610	−23.556
	(0.076)	(0.027)	(0.020)	(0.070)	(0.360)	(103.427)	(25.729)
8–10	−0.186	0.041	−0.016	0.048	−0.821	−225.650*	−50.190*
	(0.116)	(0.046)	(0.021)	(0.085)	(0.535)	(120.364)	(30.440)
11–14	−0.179	−0.034	−0.022	−0.232**	−0.551	−180.333	8.105
	(0.115)	(0.042)	(0.024)	(0.112)	(0.879)	(148.014)	(69.827)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other control variables: age squared, region, urban, education, han, marital status, urbanization index, time dummies, health insurance status, household expenditures. N=22319 (male), N=21913 (female). * p<0.10, ** p<0.05, *** p<0.01.

Robustness checks

MSMs using truncated weights

Table A6: The effect of a diabetes diagnosis on employment status and behavioural outcomes using MSM with truncated stabilized weights at 1st and 99th percentile

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
				Diabetes diagnosis			
Male sample							
Diabetes	-0.019	-0.053	-0.104^{***}	0.038	-0.804^{***}	-184.027^{***}	-17.245
	(0.029)	(0.033)	(0.032)	(0.038)	(0.198)	(59.250)	(12.885)
Female sample							
Diabetes	-0.172^{***}	-0.020	-0.018^{***}	0.134^{***}	-0.331	-92.718^{**}	-43.329^{***}
	(0.035)	(0.013)	(0.006)	(0.036)	(0.245)	(37.459)	(12.438)
				Years since diagnosis			
Male sample							
Time since diagnosis	-0.005	-0.008	-0.016^{**}	0.005	-0.157^{***}	-26.809^{**}	-2.937
	(0.006)	(0.006)	(0.006)	(0.007)	(0.034)	(11.155)	(2.202)
Female sample							
Time since diagnosis	-0.023^{***}	-0.002	-0.002^{**}	0.013^{**}	-0.047	-17.504^{***}	-5.597^{**}
	(0.007)	(0.002)	(0.001)	(0.006)	(0.047)	(5.732)	(2.561)

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 expenditures, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (male), N=16252 (female). * p<0.10, ** p<0.05, *** p<0.01.

Table A7: Effect of time since diagnosis on employment status and behavioural outcomes using MSM with truncated stabilized weights (1st and 99th percentile, duration groups)

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Male sample							
0–1	0.045	0.037	−0.028	−0.036	−0.625*	−195.065*	−21.883
	(0.052)	(0.064)	(0.063)	(0.070)	(0.370)	(114.473)	(34.068)
2–4	−0.023	−0.075*	−0.131***	0.072	−0.628**	−192.976**	−13.626
	(0.036)	(0.039)	(0.039)	(0.046)	(0.248)	(80.026)	(16.405)
5–7	−0.064	−0.116*	−0.138**	0.058	−0.864**	−180.027	−24.982
	(0.064)	(0.069)	(0.062)	(0.077)	(0.378)	(111.538)	(22.471)
8–10	−0.047	−0.031	0.023	−0.081	−1.927***	−240.486	4.223
	(0.099)	(0.130)	(0.132)	(0.132)	(0.565)	(152.693)	(49.417)
11–14	−0.011	0.063	−0.166	0.042	−1.697**	43.220	−45.690
	(0.119)	(0.107)	(0.175)	(0.152)	(0.668)	(203.611)	(40.097)
Female sample							
0–1	−0.194***	−0.046**	−0.022***	0.244***	−0.088	−35.929	−58.333***
	(0.068)	(0.019)	(0.006)	(0.068)	(0.413)	(90.643)	(16.941)
2–4	−0.165***	−0.030**	−0.020**	0.120**	−0.308	−77.304	−35.728**
	(0.045)	(0.013)	(0.010)	(0.047)	(0.264)	(58.291)	(16.843)
5–7	−0.219***	0.007	−0.004	0.149**	−0.691*	−14.513	−51.723**
	(0.062)	(0.030)	(0.016)	(0.069)	(0.393)	(83.307)	(23.213)
8–10	−0.143	0.040	−0.024**	0.192**	−0.371	−291.341**	−74.909***
	(0.102)	(0.053)	(0.009)	(0.090)	(0.598)	(112.655)	(27.188)
11–14	−0.085	−0.041	−0.019*	−0.130*	−0.078	−256.217**	4.161
	(0.119)	(0.031)	(0.011)	(0.077)	(0.925)	(123.032)	(66.481)

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 expenditures, alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension. N=16557 (male), N=16252 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Lagged covariates

Table A8: 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)
Male sample							
Diabetes	0.058*	0.003	−0.073	0.009	−0.894***	−187.768**	10.622
	(0.035)	(0.045)	(0.049)	(0.049)	(0.243)	(89.057)	(20.357)
Female sample							
Diabetes	−0.140**	−0.011	−0.011	0.162***	−0.672**	−41.826	−51.978*
	(0.058)	(0.009)	(0.017)	(0.058)	(0.302)	(74.000)	(27.663)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other lagged control variables: Age squared, region, education, marital status, urbanization index, time dummies, health insurance status, household expenditures and the outcome variables alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension, though we do not include the respective lagged dependent variable. N=16557 (male), N=16252 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A9: The effect of each year since 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)
Male sample							
Time since diagnosis	0.006	0.001	-0.009	0.005	-0.209^{***}	-28.564*	2.573
	(0.009)	(0.008)	(0.009)	(0.009)	(0.047)	(16.315)	(4.407)
Female sample							
Time since diagnosis	-0.024^{**}	0.002	0.000	0.007	-0.071	-10.565	-5.739
	(0.010)	(0.002)	(0.002)	(0.009)	(0.053)	(11.502)	(4.046)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other lagged control variables: Age squared, region, education, marital status, urbanization index, time dummies, health insurance status, household expenditures and the outcome variables alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension, though we do not include the respective lagged dependent variable. N=16557 (male), N=16252 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A10: The effect of time since diabetes diagnosis on employment status and behavioural outcomes using FE (duration groups, lagged covariates)

	Employed	Smoking	Alcohol	Hypertension	BMI (kg/m^2)	Calories (kcal)	Physical activity (hours/week)
Male sample							
0–1	0.151^{***}	0.031	−0.045	−0.059	-1.019^{**}	-278.220^{**}	6.269
	(0.052)	(0.074)	(0.064)	(0.076)	(0.405)	(120.839)	(32.336)
2–4	0.017	−0.014	-0.094^{*}	0.031	-0.773^{**}	−182.325	9.825
	(0.042)	(0.051)	(0.057)	(0.057)	(0.309)	(111.411)	(24.093)
5–7	−0.017	−0.049	−0.094	0.053	-1.363^{***}	−123.780	−1.017
	(0.074)	(0.080)	(0.078)	(0.091)	(0.428)	(145.661)	(39.132)
8–10	0.089	−0.018	0.017	−0.020	-2.014^{***}	−288.200	43.493
	(0.107)	(0.140)	(0.143)	(0.124)	(0.672)	(209.609)	(63.514)
11–14	0.165	0.045	−0.174	0.109	-2.912^{***}	−348.511	2.184
	(0.155)	(0.186)	(0.152)	(0.174)	(0.760)	(228.023)	(56.682)
Female sample							
0–1	−0.095	−0.031	−0.016	0.185^{**}	-0.713^{*}	−77.341	-78.226^{**}
	(0.074)	(0.020)	(0.011)	(0.083)	(0.377)	(101.184)	(33.738)
2–4	-0.176^{***}	−0.011	−0.022	0.161^{**}	−0.531	−46.306	−48.470
	(0.067)	(0.010)	(0.021)	(0.063)	(0.373)	(86.789)	(32.173)
5–7	-0.258^{***}	0.040	0.014	0.157^{*}	-1.318^{***}	−9.542	−53.157
	(0.090)	(0.030)	(0.030)	(0.082)	(0.437)	(124.234)	(33.340)
8–10	-0.256^{**}	0.055	−0.002	0.214^{**}	−0.846	−216.734	-93.327^{**}
	(0.129)	(0.046)	(0.021)	(0.106)	(0.569)	(145.445)	(39.082)
11–14	-0.269^{**}	−0.017	−0.001	−0.080	−0.558	−200.894	−43.512
	(0.132)	(0.034)	(0.023)	(0.137)	(0.938)	(183.520)	(72.320)

Note Standard errors clustered at the individual level in parentheses. Employed, smoking, alcohol and hypertension are binary outcomes. Other lagged control variables: Age squared, region, education, marital status, urbanization index, time dummies, health insurance status, household expenditures and the outcome variables alcohol consumption, smoking status, BMI, calorie consumption, physical activity, hypertension, though we do not include the respective lagged dependent variable. N=16557 (male), N=16252 (female). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.