

Can Chronic Disease Diagnosis Urge the Patients to Quit Smoking? — Evidence from the China Health and Nutrition Survey

Yue Hu ¹
Qihui Chen ²
Bo Zhang ³

¹College of Economics and Management, China Agricultural University, Beijing, People's Republic of China; ²Beijing Food Safety Policy & Strategy Research Base, China Agricultural University, Beijing, People's Republic of China; ³Department of Neurology and ICCTR Biostatistics and Research Design Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

Background: Tobacco consumption is one of the world's largest public health threats. Yet little is known about how chronic disease diagnoses affect individuals' smoking behavior in China, where the world's largest smoking population resides.

Methods: This study analyzes an unbalanced panel dataset on 2986 Chinese males aged 50 or above from the China Health and Nutrition Survey, a household survey covering nine Chinese provinces. We adopt a zero-inflated negative binomial (ZINB) regression framework to account for the count-data nature of the outcome variable of interest, the number of cigarettes one smokes per day. Logit regressions are also adopted to predict one's likelihood of smoking cessation.

Results: First, the estimated ZINB model suggests that the number of chronic disease diagnoses only affects whether one smokes, but conditional on one being a smoker, it does not affect the number of cigarettes one smokes per day. Logit estimates suggest that an additional diagnosed chronic disease is associated with a 4.8 percentage-point increase in the likelihood of smoking cessation. Second, while the diagnoses of all four chronic conditions examined are found to increase the likelihood of smoking cessation, the diagnosis of myocardial infarction has the largest impact, followed by diabetes diagnosis.

Conclusion: While chronic disease diagnoses reduce smoking in China, their effects are small. Healthcare policies and relevant measures (such as helping smoking patients switch to a diet with more foods enriched with antioxidants) are thus needed to reduce the adverse effects that continued smoking might impose on their health.

Keywords: chronic disease, diagnosis, smoking, count-data models, China

Introduction

Cigarette smoking is one of the world's largest public health threats.¹ It has been widely documented that smoking can lead to a large number of chronic diseases, including hypertension, diabetes, strokes, cardiovascular disease, and cancer, imposing enormous health and economic burdens on the human population.^{2–13} Yet, less known is: Will one quit smoking, or at least reduce the number of cigarettes smoked if he/she is diagnosed with one or more chronic conditions?

The answer to this question is not only of interest in its own right but also of significant policy relevance. If chronic disease diagnoses failed to serve as a wake-up call to urge patients to quit smoking, presumably due to smoking's addictive nature, special medical measures might be needed to reduce the adverse effects of continued smoking on their worsened health. For example, medical professionals

Correspondence: Qihui Chen
Email chen1006@umn.edu

Received: 12 April 2021
Accepted: 8 July 2021
Published: 22 July 2021



may need to help smoking patients adopt a diet with more antioxidants to reduce the harmful effects of continued smoking. Somewhat puzzlingly, while there have been numerous studies conducted in many other countries,^{14–18} little has been done in China, where the number of smokers has recently reached 300 million,¹³ and chronic diseases are becoming increasingly prevalent.^{19–24}

To help provide an answer to this question in the context of China, we analyze a dataset spanning 11 years (2004–2015) drawn from the China Health and Nutrition Survey (CHNS), a household survey covering nine provinces in China. To account for the count-data nature of the outcome variable of primary interest, the number of cigarettes one smokes per day, as well as the fact that many respondents never smoked, we adopt a Zero-inflated Negative Binomial (ZINB) framework to model a respondent's daily smoking behavior. Given the low proportion of female smokers (6.18%) and the low incidence of chronic diseases among individuals under 50 in China,²⁵ our analysis focuses on male respondents aged 50 or above at the time of the survey in 2004. Based on an analytical sample of 2,986 Chinese elder males, our estimation suggests that chronic disease diagnoses do affect whether one smokes—they significantly raise the likelihood of smoking cessation among Chinese elder males. However, conditional on a respondent being a smoker, chronic disease diagnoses fail to predict the number of cigarettes this respondent smokes per day. Moreover, while the diagnoses of all four chronic conditions examined (hypertension, diabetes, stroke, and myocardial infarction) are found to increase the likelihood of smoking cessation, the diagnosis of myocardial infarction has the largest impact, followed by diabetes diagnosis.

Data

The China Health and Nutrition Survey

The data analyzed in this study came from the China Health and Nutrition Survey (CHNS: <https://www.cpc.unc.edu/projects/china>), a panel survey project jointly designed and implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Health at the Chinese Center for Disease Control and Prevention. The first wave of the CHNS, conducted in 1989, drew a sample of over 16,000 individuals from about 3,800 households in nine Chinese provinces (namely, Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and

Guizhou) that vary substantially in geography, economic development, public resources, and health indicators, based on a stratified, random sampling framework. Follow-up waves were conducted in 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, and 2015. Three municipal cities (namely, Beijing, Shanghai, and Chongqing) were added to the project from 2011 (—locations of the participating provinces and municipal cities can be seen at https://www.cpc.unc.edu/projects/china/about/proj_desc/chinamap).

The CHNS collected rich information on survey respondents' socio-demographic characteristics, food consumption behavior, nutrition intakes, health status, employment status, income, etc. Most important for the present study, the CHNS started to collect information on respondents' smoking behavior in 1991 and chronic disease diagnoses (beyond the diagnosis of hypertension, which has been recorded since 1991) in 1997.

Since the present study was conducted based on the de-identified, publicly available CHNS data, it does not constitute human subjects research. Its institutional review board review was waived because there was no interaction with any individual, and no identifiable private information was used.

Sample Restrictions

For the purpose of our study, we applied several sample restrictions and data-cleaning methods to form our analytical sample: (1) We dropped data collected before 2004 to ensure that all observations (potentially) have information on both smoking and chronic disease diagnoses lagged for two time periods (—the reason why we need two-period lagged values is provided in the next section). (2) To exploit as much longitudinal information as possible, we kept only observations from the nine original CHNS provinces, excluding those municipal cities that joined the project in 2011. (3) We excluded females from the analysis because only a small proportion (6.18%) of female respondents in the data had ever smoked. (4) We further excluded (male) respondents under the age of 50 in 2004 because the incidences of chronic diseases are very low for these individuals.²⁵ (5) Missing information (eg, on education or disease diagnosis) was filled in with available information provided in other waves. For example, if a male respondent was diagnosed with hypertension in 2004, we can infer that the missing hypertension diagnosis status for this respondent in 2006 should be imputed as “had been diagnosed with hypertension”. These procedures resulted in an unbalanced panel of 10,202 individual-year observations for 2,986 individuals in 226 communities (ie, rural villages and urban

residential districts) in nine Chinese provinces, appearing at least in one wave between 2004 and 2015—the flowchart in Figure 1 details our data compilation process.

Method

Primary Approach: A Zero-Inflated Negative Binomial Regression Framework

Consider the following statistical model that links a male respondent's smoking behavior and its determinants:

$$Nsmk_{it} = \beta_0 + \beta_1 Nchn_{it} + \beta_2 Z_{it} + \varepsilon_{it} \quad (1)$$

where the outcome variable of interest, $Nsmk_{it}$, is the number of cigarettes respondent i smokes per day at time t . The explanatory variable of primary interest, $Nchn_{it}$, is the total number of chronic conditions which respondent i had ever been diagnosed with by time t . Z_{it} represents a set of other observed determinants of one's smoking behavior that vary at the individual, household, and com-

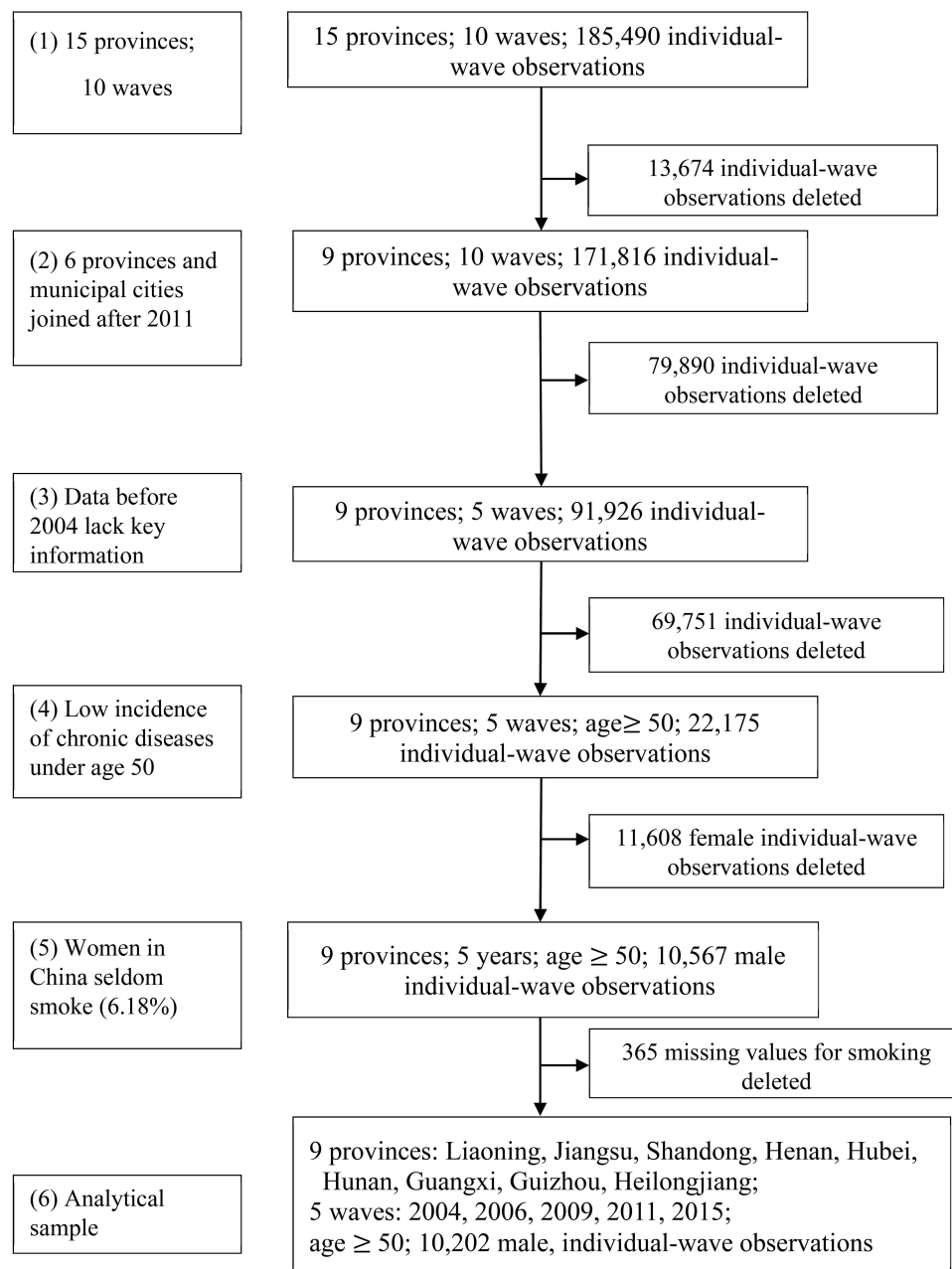


Figure 1 Flow charts for data collection and compilation.

munity levels (—detailed in the next paragraph). The error term ε_{it} captures the influence of unobserved factors.

The covariates in \mathbf{Z} are chosen based on health economic theory and empirical findings in previous studies. First, basic economic theory predicts that individuals' consumption behavior is affected by their income, the price of the commodity being consumed (ie, cigarettes in our case), and the prices of related goods (eg, alcoholic beverages and other foods).²⁶ Due to the lack of complete price information in the CHNS data, we include in \mathbf{Z} the full set of community and survey-wave fixed effects (dummy variables) to capture price variations. The inclusion of these fixed effects also effectively absorbs the influence of unobserved factors, such as the local disease environment and medical conditions, that vary at the community level. Second, smoking has been modeled as a life-course behavior in the context of China.²⁷ We follow Kenkel et al (2009)²⁷ and include one's age (in years), age squared, years of education, and working status (a dummy for "currently working"), as covariates in \mathbf{Z} . We also follow Yen et al (2010)²⁸ and include one's marital status and whether one drinks alcoholic beverages as covariates. Third, to the extent that individuals' smoking behavior may change in response to health shocks²⁸ and health insurance coverage,²⁹ we include one's self-reported overall health status (a dummy variable for feeling "good" or "very good") and whether one has health insurance coverage as covariates. Finally, we include a dummy variable for whether other household members smoke to capture the household environment and within-household peer effects in \mathbf{Z} .

Note that the outcome variable of interest, $Nsmk_{it}$, can take only non-negative integers. The count-data nature of this variable naturally suggests a Poisson regression approach to estimating Equation (1).³⁰ Yet in practice, the equal-dispersion property dictated by Poisson distributions, ie, $Var(Nsmk|Z) = E(Nsmk|Z)$, is often violated, and the Negative Binomial model is usually adopted to address the potential over-dispersion problem.³⁰ To further account for the fact that many respondents never smoked, which results in a clustering of zeros in $Nsmk_{it}$, we adopt a Zero-inflated Negative Binomial (ZINB) regression framework to estimate Equation (1)—the "zero-inflation" part is estimated by a Logit model predicting "certain zeros".

Finally, since our data involve repeated observations of a subset of respondents over time, we adjust standard

errors for intra-class clustering at the community level, as suggested by Liang and Zeger (1986)³¹ and Cameron and Trivedi (2005).³⁰ All estimations in this study were performed in Stata SE 14.0.

Methods for Robustness Checks

There are also several problems that may result in misleading estimates of model parameters, especially β_1 , which captures the effects of chronic disease diagnoses on smoking. First, smoking can do serious harm to one's health,^{2-4,6-11,13} thereby creating potential reverse causality from $Nsmk_{it}$ to $Nchn_{it}$, which in turn will bias the estimate of β_1 . Yet, this may not be a major concern in our setting because $Nchn_{it}$, ie, the number of chronic conditions ever diagnosed, is likely to be recorded before $Nsmk_{it}$ was observed. Still, as a robustness check, we replace $Nchn_{it}$ with its lagged measure observed in the previous period, $Nchn_{i,t-1}$,

$$Nsmk_{it} = \beta_0 + \beta_1 Nchn_{i,t-1} + \beta_2 Z_{it} + \varepsilon_{it} \quad (2)$$

in some analyses reported below. The concern over reverse causality does apply to one's overall health status, whether one drinks, and whether one's family members smoke. Since whether a person smokes at time t cannot affect the values of these explanatory variables realized at time $t-1$, we replace the current values of these variables with their lagged values in Equation (2) in robustness checks.

The second issue concerns the potential influence of unobserved confounding factors. To address this issue, we include the lagged measure of the outcome variable, $Nsmk_{i,t-1}$, as an additional covariate in the model as a check (—if a balanced panel dataset were available, a more rigorous approach is to control for individual-level fixed effects). To the extent that many unobserved confounders, such as genetic fitness, have persistent effects on one's health status and smoking behavior, the lagged outcome measure, $Nsmk_{i,t-1}$, can sensibly capture the influence of these confounders. This variable also helps to capture the addictive nature of smoking.

It is worth pointing out that here we invoke the assumption that conditional on Z_{it} , the lagged outcome measure $Nsmk_{i,t-1}$ is uncorrelated with ε_{it} , the disturbance at time t . This assumption will fail if there exist unobserved time-invariant confounders at the individual level that create a correlation between $Nsmk_{i,t-1}$ and ε_{it} . This correlation will bias not only the estimated coefficient of $Nsmk_{i,t-1}$ but also other coefficients through the correlations among explanatory variables.³² If a balanced panel

dataset with more waves were available, one possible solution is to apply the system-GMM (Generalized Method of Moments) method developed by Anderson and Hsiao (1981)³³ and Arellano and Bond (1991),³⁴ which transforms all variables in Equation (2) into their first-differenced form and then uses 3-period lagged and earlier values of these variables as instrumental variables for (the first-differenced form of) $Nsmk_{i,t-1}$.³² Unfortunately, our unbalanced panel (with relatively few waves) does not permit us to apply this method. As an alternative, we assess how the potential correlation between $Nsmk_{i,t-1}$ and ε_{it} may affect our estimate of β_1 by replacing $Nsmk_{i,t-1}$ with $Nsmk_{i,t-2}$. The rationale is that compared with $Nsmk_{i,t-1}$, $Nsmk_{i,t-2}$ is equally likely to be correlated with ε_{it} through unobserved time-invariant confounders but less likely to be correlated with other explanatory variables (since it is lagged for more periods). Thus, the results obtained using $Nsmk_{i,t-2}$ as a predictor for one's current smoking behavior are less likely to be biased. This prediction allows us to assess whether our estimate of β_1 is seriously biased by comparing the results obtained with $Nsmk_{i,t-1}$ and those obtained with $Nsmk_{i,t-2}$.

A final check for robustness is to see whether the estimated effects of chronic disease diagnoses vary significantly across subsamples defined by the values of individuals' socio-economic characteristics and those of statistically significant predictors in the model. This heterogeneity analysis allows us to assess whether our results are driven by particular subgroups or reflect a general pattern.

Results

Descriptive Findings

Table 1 depicts the profile of male respondents aged 50 or above in our sample. About two-thirds of the respondents were rural residents. Roughly 88% have a spouse. The average respondent has nearly seven years of education. Despite the average age of 65, 42% of these respondents were still working at the time of the survey, with the average respondent earning an annual income slightly higher than 30,000 yuan (approximately 4,594 USD under the 2020 price). Presumably due to the rapid expansion of China's health insurance sector since the early 2000s,^{25,35-37} nearly 81% of the respondents were covered by health insurance.

Regarding respondents' smoking behavior, 48% were smoking at the time of the survey. On average, smokers

(—defined as those who were smoking at the time of the survey) consumed nearly 16 cigarettes per day, yielding an average of 7.6 cigarettes smoked per day in the sample. The average number of chronic diseases diagnosed (—hypertension, diabetes, myocardial infarction, and stroke combined) of respondents was 0.46, with a minimum of zero and a maximum of four. Among these four chronic conditions, hypertension has the highest prevalence, 32%, followed by diabetes (8%), stroke (7%), and myocardial infarction (4%).

Quick comparisons between smokers (N = 1,780) and non-smokers (N = 2,063) reveal that the incidences of chronic diseases are in general higher among non-smokers. Yet, it bears mentioning that these comparisons were made without netting out the influence of other confounding factors. For example, chronic diseases might have urged some former smokers to quit smoking—the lagged values of cigarette consumption also suggest possible smoking cessation among current non-smokers. Table 1 also shows non-trivial differences in socio-economic characteristics, such as education, working status, and income, between smokers and non-smokers. A rigorous regression framework is therefore adopted to control for the influence of potential confounding factors.

Main Estimation Results

To assess more rigorously how chronic disease diagnoses affect the number of cigarettes Chinese males smoke per day while accounting for the fact that many of them never smoked, we estimated Equation (1) using a Zero-inflated Negative Binomial (ZINB) regression model—Vuong's (1989) likelihood ratio tests³⁸ suggest that the ZINB model outperformed all alternative models, including the Zero-inflated Poisson model, the standard Poisson, and the standard Negative Binomial model. Table 2, columns (1)-(2), presents the coefficient estimates of the ZINB model. The results suggest that the number of chronic diseases ever diagnosed by medical doctors ($Nchn_{it}$) has a statistically significant predictive power for whether a Chinese male respondent smokes (column 1: the “zero-inflation” part). Since the Logit model for the “zero-inflation” part predicts whether a male respondent's cigarette consumption is a “certain zero”, the positive coefficient of $Nchn_{it}$ implies that chronic disease diagnoses reduce the incidence of smoking. However, once conditional on a respondent being a smoker, the number of chronic conditions diagnosed has virtually no explanatory power for the number of cigarettes smoked per day (column 2: the Negative Binomial model).

Table 1 Summary Statistics of All Individual-Wave Observations from 2004 to 2015

Variables	Full Sample		Smokers		Non-Smokers	
	Mean (Proportion of 1 for Dummy Variables)	Std. Dev.	Mean (Proportion of 1 for Dummy Variables)	Std. Dev.	Mean (Proportion of 1 for Dummy Variables)	Std. Dev.
Whether smoking	0.48	0.50	1.00	0.00	0.00	0.00
Number of cigarettes smoked per day (current period)	7.65	10.32	15.93	9.48	0.00	0.00
Number of cigarettes smoked per day (1-period lagged)	8.01	10.39	14.29	10.48	2.58	6.53
Number of cigarettes smoked per day (2-period lagged)	8.92	10.54	14.25	10.40	4.13	8.07
Chronic disease diagnosed						
Hypertension (current period)	0.32	0.47	0.26	0.44	0.37	0.48
Diabetes (current period)	0.07	0.26	0.05	0.22	0.10	0.30
Stroke (current period)	0.07	0.26	0.05	0.21	0.09	0.28
Myocardial infarction (current period)	0.04	0.19	0.02	0.14	0.05	0.22
Hypertension diagnosed (1-period lagged)	0.24	0.43	0.19	0.39	0.30	0.46
Diabetes diagnosed (1-period lagged)	0.05	0.22	0.03	0.18	0.07	0.25
Stroke diagnosed (1-period lagged)	0.05	0.21	0.03	0.18	0.06	0.23
Myocardial infarction (1-period lagged)	0.02	0.15	0.01	0.11	0.03	0.18
Number of chronic diseases diagnosed (current period)	0.46	0.71	0.34	0.60	0.57	0.79
Number of chronic diseases diagnosed (1-period lagged)	0.36	0.64	0.26	0.53	0.45	0.71
Self-reported health status (current) (=1, if "good" or "very good")	0.91	0.29	0.92	0.27	0.90	0.30
Self-reported health status (1-period lagged)	0.92	0.28	0.93	0.26	0.91	0.29
Drinking (=1 if currently drinking)	0.50	0.50	0.63	0.48	0.38	0.49
Drinking (1-period lagged)	0.54	0.50	0.64	0.48	0.45	0.50
Others smoke in the household (=1 if yes)	0.33	0.47	0.38	0.48	0.28	0.45
Others smoke in the household (1-period lagged)	0.35	0.48	0.39	0.49	0.31	0.46
Age (years)	65.27	8.86	63.61	8.24	66.81	9.14
Age ² (years)/100	43.39	11.99	41.14	10.91	45.47	12.56
Education (years)	6.82	4.14	6.52	3.88	7.09	4.35
Annual income (yuan)	31,138	44,800	29,368	47,636	32,777	41,924
Has spouse (=1, if having a spouse)	0.88	0.32	0.89	0.31	0.87	0.34
Health insurance (=1, if having medical insurance coverage)	0.81	0.39	0.78	0.42	0.84	0.37
Working (=1, if currently working in the labor market)	0.42	0.49	0.50	0.50	0.34	0.47
Rural (=1, if a respondent resides in a rural area)	0.66	0.48	0.70	0.42	0.62	0.49
Number of individuals	2,986		1,780		2,063	
Number of wave-individual observations	10,202		4,902		5,300	

Notes: Data from 2004, 2006, 2009, 2011, and 2015 waves of the CHNS are used in the table; Information on "cancer diagnosis" is not used in the analysis below because it was not available until 2011; Smokers are defined as individuals who reported "currently smoking" at the time of the survey; non-smokers are defined as individuals who reported "currently not smoking" at the time of the survey.

In other words, the ZINB model reduces to a Logit model: we can fit the relationship of interest (Equation 1) reasonably well by a Logit model predicting the likelihood of smoking cessation (—accordingly, we coded the outcome variable to be = 1 if a respondent is “currently not smoking” and = 0 otherwise). Thus, column (3) of Table 2 reports Logit estimates of model parameters, ignoring the distribution within the range of positive numbers of cigarettes smoked. The estimates are indeed very close to the “zero-inflation” part of the ZINB model (Table 2, column 1). Since coefficient estimates in Logit models are in general not equal to the marginal effects of changing one unit of the corresponding explanatory variables,³² for ease of interpretation, column (4) of Table 2 reports the corresponding marginal effects (evaluated at sample means of the explanatory variables). The estimates suggest that an additional chronic disease diagnosed is associated with a 4.8 percentage-point increase in the likelihood of smoking cessation, an effect statistically significant at the 1% level. Further excluding explanatory variables that are not statistically significant at the 10% level, columns (5) and (6) of Table 2 show estimated effects of chronic disease diagnoses that are similar to their counterparts reported in columns (3) and (4).

To further explore the diagnosis of which chronic disease has the largest effect, Table 3, columns (1)–(8), examines the role of the diagnosis of each of the four chronic diseases of interest (ie, hypertension, diabetes, stroke, and myocardial infarction), also based on a ZINB model. While the diagnoses of all four chronic conditions are found to raise the likelihood of smoking cessation, they (—except the case of stroke diagnosis—) have no significant impact on the number of cigarettes smoked, conditional on a male respondent being a smoker. The diagnosis of myocardial infarction has the largest on smoking cessation (column 7), followed by diabetes (column 3), stroke (column 5), and hypertension diagnosis (column 1). Columns (9) and (10) of Table 3 further include all four chronic diseases in the same model; the estimates reveal a similar ordering pattern discussed above. Finally, Logit estimates of marginal effects (column 12) suggest that the diagnosis of myocardial infarction, diabetes, stroke, and hypertension raises the likelihood of smoking cessation, respectively, by 10.5, 7.4, 3.6, and 3.4 percentage points. These estimates suggest that while chronic disease diagnoses do serve to reduce cigarette consumption in China, their effects are relatively small.

The estimated coefficients on other explanatory variables are also informative. For example, column (4) of Table 2 shows that the age profile of smoking has an inverted-U shape, suggesting that Chinese males tend to reduce their cigarette consumption after reaching a certain age. Consistent with previous findings,²⁷ income is positively associated with smoking cessation. Also, consistent with intuition, if other household members smoke, a respondent is less likely to quit smoking. If a respondent drinks, he is also less likely to quit smoking.

Robustness Checks

The above findings suggest that chronic disease diagnoses help urge Chinese males to quit smoking, although with relatively small effects. But as discussed in Method, there are still threats that might invalidate these findings. This section, thus, performs a series of robustness checks to strengthen our findings. First, we assess the possibility that our findings are driven by reverse causality from respondents’ smoking behavior to their health. More specifically, we replace the “current” values of a set of explanatory variables, including the number of chronic diseases diagnosed, the diagnosis of each specific chronic condition, self-reported overall health status, whether one drinks, and whether one’s other household members smoke, with their one-period lagged values. The new estimates, reported in columns (1)–(4) of Tables 4 and 5, are comparable to their counterparts reported in Tables 2 and 3, lending support to our original estimates.

Another threat, also discussed in Method, is the potential existence of unobserved confounding factors such as genetic fitness, which has been shown to affect one’s health behavior and health outcomes significantly.^{39–41} As a check, we include $Nsmk_{i,t-1}$, the 1-period lagged value of the outcome variable of interest in the model to capture the influence of unobserved factors. As revealed in column (5) of Tables 4 and 5, the inclusion of $Nsmk_{i,t-1}$ hardly alters our original findings reported in Tables 2 and 3. Yet as discussed in Method, the inclusion of $Nsmk_{i,t-1}$ may induce a correlation between $Nsmk_{i,t-1}$ and ε_{it} (through unobserved time-invariant confounders), thereby biasing the estimated effects of chronic disease diagnoses. We thus assess how the potential correlation between $Nsmk_{i,t-1}$ and ε_{it} may affect our estimate of β_1 by replacing $Nsmk_{i,t-1}$ with $Nsmk_{i,t-2}$. As discussed in Method, compared with $Nsmk_{i,t-1}$, $Nsmk_{i,t-2}$ is equally likely to be correlated with the unobserved time-invariant confounders

Table 2 Zero-Inflated Negative Binomial and Logit Estimates of Associations Between the Number of Chronic Diseases Diagnosed and Daily Cigarette Consumption (Chinese Males Aged 50 or Above), Main Results

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variables	Number of Cigarettes Smoked per Day		Smoking Cessation		Smoking Cessation	
Models	Zero-Inflated Negative Binomial		Logit	Logit	Logit	Logit
	Inflation Model (Logit)	Negative Binomial Part				
Reported Estimates	Coefficients	Coefficients	Coefficients	Marginal Effects	Coefficients	Marginal Effects
Number of chronic diseases diagnosed (current)	0.242***	0.002	0.243***	0.048***	0.266***	0.053***
	(0.055)	(0.022)	(0.054)	(0.011)	(0.054)	(0.011)
Self-reported health status (current)	-0.139	0.054	-0.127	-0.025		
	(0.099)	(0.040)	(0.098)	(0.019)		
Drinking (dummy, = 1 if yes)	-0.928***	0.028	-0.942***	-0.188***	-0.964***	-0.193***
	(0.066)	(0.022)	(0.066)	(0.012)	(0.065)	(0.012)
Other smokers in the household (dummy, = 1 if yes)	-0.328***	0.029	-0.327***	-0.065***	-0.319***	-0.064***
	(0.086)	(0.026)	(0.085)	(0.017)	(0.084)	(0.017)
Age (years)	-0.072	0.027	-0.063	-0.013	-0.053	-0.010
	(0.055)	(0.021)	(0.055)	(0.011)	(0.053)	(0.011)
Age (years) squared/100	0.080*	-0.032**	0.072*	0.014*	0.066*	0.013*
	(0.041)	(0.016)	(0.041)	(0.008)	(0.039)	(0.008)
Years of education (years)	0.016	-0.004	0.015	0.003		
	(0.012)	(0.004)	(0.012)	(0.002)		
Log annual income (yuan)	0.062*	0.021*	0.062*	0.012**	0.059*	0.012*
	(0.032)	(0.011)	(0.032)	(0.006)	(0.031)	(0.006)
Has spouse (dummy, = 1 if yes)	-0.113	-0.005	-0.123	-0.025		
	(0.128)	(0.046)	(0.127)	(0.025)		
Has insurance coverage (dummy, = 1 if yes)	0.006	0.061**	0.037	0.007		
	(0.083)	(0.031)	(0.081)	(0.016)		
Working (dummy, = 1 if yes)	-0.081	0.052**	-0.090	-0.018		
	(0.075)	(0.024)	(0.075)	(0.015)		
Survey-wave fixed effects	yes	yes	yes	yes	yes	yes
Community fixed effects	yes	yes	yes	yes	yes	yes
Constant	1.216	2.025***	0.976		0.484	
	(1.907)	(0.713)	(1.907)		(1.790)	
Number of individual-wave observations	9,108		9,108		9,288	
Log pseudolikelihood	-20,200.59		-5301.40		-5416.478	

Notes: Data from 2004, 2006, 2009, 2011, and 2015 waves of the CHNS are used in the analysis. In the “zero-inflation” part of the Zero-inflated Negative Binomial model, the dependent variable is defined as = 1 if “zero-inflated” and = 0 if otherwise. In the Logit models predicting “smoking cessation”, the dependent variable is defined as = 1 if a respondent is “not smoking” and = 0 if otherwise; Standard errors are reported in parentheses, adjusted for clustering at the community level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 Zero-Inflated Negative Binomial and Logit Estimates of Associations Between Chronic Disease Diagnoses and Daily Number of Cigarette Consumption (Chinese Males Aged 50 or Above)

Outcome Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
	Number of Cigarettes Smoked per Day		Number of Cigarettes Smoked per Day		Number of Cigarettes Smoked per Day		Number of Cigarettes Smoked per Day		Number of Cigarettes Smoked per Day		Smoking Cessation		
	Zero-inflated Negative Binomial	Inflation Model (Logit)	Neg-Bin Part	Zero-Inflated Negative Binomial	Inflation Model (Logit)	Neg-Bin Part	Zero-Inflated Negative Binomial	Inflation Model (Logit)	Neg-Bin Part	Zero-Inflated Negative Binomial	Inflation Model (Logit)	Neg-Bin Part	Logit
Reported Estimates	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Marginal Effects
Diagnosis:													
Hypertension	0.236***	0.006							0.175*	0.004	0.173*	0.034*	
	(0.088)	(0.029)							(0.091)	(0.030)	(0.090)	(0.018)	
Diabetes			0.423***	-0.066					0.357**	-0.067	0.370**	0.074**	
			(0.144)	(0.073)					(0.148)	(0.073)	(0.146)	(0.029)	
Stroke					0.320*	0.108**			0.200	0.128**	0.182	0.036	
					(0.176)	(0.055)			(0.185)	(0.058)	(0.183)	(0.036)	
Myocardial infarction							0.593***	-0.133	0.487**	-0.154	0.529**	0.105**	
							(0.224)	(0.128)	(0.231)	(0.122)	(0.231)	(0.046)	
Overall health status	-0.190*	0.056	-0.197**	0.047	-0.175*	0.057	-0.202**	0.046	-0.140	0.061	-0.130	-0.026	
	(0.098)	(0.040)	(0.098)	(0.041)	(0.099)	(0.041)	(0.098)	(0.041)	(0.099)	(0.040)	(0.098)	(0.019)	
Drinking	-0.942***	0.028	-0.936***	0.026	-0.933***	0.029	-0.933***	0.025	-0.924***	0.026	-0.938***	-0.187***	
	(0.066)	(0.022)	(0.066)	(0.022)	(0.065)	(0.022)	(0.066)	(0.022)	(0.066)	(0.022)	(0.066)	(0.012)	
Other smokers	-0.331***	0.029	-0.327***	0.028	-0.332***	0.028	-0.327***	0.028	-0.326***	0.027	-0.324***	-0.064***	
	(0.086)	(0.026)	(0.086)	(0.026)	(0.086)	(0.026)	(0.086)	(0.026)	(0.087)	(0.025)	(0.086)	(0.017)	

(Continued)

Table 3 (Continued).

Outcome Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Number of Cigarettes Smoked per Day	Smoking Cessation	Smoking Cessation
Model	Zero-inflated Negative Binomial	Zero-inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Zero-Inflated Negative Binomial	Logit	Logit
Reported Estimates	Inflation Model (Logit)	Neg-Bin Part	Inflation Model (Logit)	Neg-Bin Part	Inflation Model (Logit)	Neg-Bin Part	Inflation Model (Logit)	Neg-Bin Part	Inflation Model (Logit)	Neg-Bin Part	Coef.	Marginal Effects
Age	-0.068 (0.055)	0.026 (0.021)	-0.064 (0.055)	0.026 (0.021)	-0.060 (0.055)	0.027 (0.021)	-0.057 (0.055)	0.027 (0.021)	-0.071 (0.055)	0.026 (0.021)	-0.062 (0.055)	-0.012 (0.011)
Age squared/100	0.077* (0.041)	-0.032** (0.016)	0.076* (0.041)	-0.032** (0.016)	0.072* (0.041)	-0.032** (0.016)	0.070* (0.041)	-0.032** (0.016)	0.079* (0.041)	-0.032** (0.016)	0.071* (0.041)	0.014* (0.008)
Years of education	0.015 (0.012)	-0.004 (0.004)	0.016 (0.012)	-0.004 (0.004)	0.016 (0.012)	-0.005 (0.004)	0.016 (0.012)	-0.005 (0.004)	0.016 (0.012)	-0.004 (0.004)	0.015 (0.012)	0.003 (0.002)
Log annual income	0.062* (0.032)	0.021* (0.011)	0.066** (0.032)	0.023** (0.011)	0.064** (0.032)	0.022** (0.011)	0.065** (0.032)	0.022** (0.011)	0.063* (0.032)	0.021* (0.011)	0.063** (0.032)	0.013** (0.006)
Has spouse	-0.102 (0.129)	-0.006 (0.046)	-0.109 (0.128)	-0.009 (0.045)	-0.098 (0.129)	-0.012 (0.045)	-0.106 (0.128)	-0.008 (0.045)	-0.120 (0.128)	-0.001 (0.046)	-0.131 (0.126)	-0.026 (0.025)
Has insurance	0.004 (0.082)	0.062** (0.031)	0.018 (0.083)	0.061** (0.031)	0.013 (0.082)	0.063** (0.031)	0.007 (0.083)	0.062** (0.031)	0.008 (0.083)	0.061** (0.031)	0.039 (0.081)	0.008 (0.016)
Working	-0.090 (0.075)	0.054** (0.023)	-0.092 (0.074)	0.050** (0.023)	-0.095 (0.074)	0.052** (0.023)	-0.091 (0.074)	0.049** (0.023)	-0.080 (0.075)	0.050** (0.023)	-0.089 (0.075)	-0.018 (0.015)
Survey-wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Community fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.213 (1.911)	2.036*** (0.710)	0.993 (1.908)	2.039*** (0.714)	0.868 (1.909)	2.029*** (0.713)	0.754 (1.913)	2.048*** (0.719)	1.140 (1.905)	2.047*** (0.711)	0.895 (1.905)	
Number of individual-wave observations	9,127		9,122		9,130		9,128		9,108		9,108	
Log pseudolikelihood	-20,243.23		-20,245.99		-20,271.85		-20,259.41		-20,191.33		-5298.34	

Notes: Data from 2004, 2006, 2009, 2011, and 2015 waves of the CHNS are used in the analysis; In the “zero-inflation” part of the Zero-inflated Negative Binomial model, the dependent variable is defined as = 1 if “zero-inflated”, and = 0 if otherwise; In the Logit models predicting “smoking cessation”, the dependent variable is defined as = 1 if a respondent is “not smoking” and = 0 if otherwise; The number of individual-wave observations varies somewhat across columns due to missing values in the different explanatory variables included; Standard errors are reported in parentheses, adjusted for clustering at the community level; **p* < 0.1, ***p* < 0.05, ****p* < 0.01

in ϵ_{it} but less likely to be correlated with other explanatory variables. Thus, the effects of chronic disease diagnoses estimated using $Nsmk_{i,t-2}$ as a predictor for one’s current smoking behavior are less likely to be biased. Yet as shown in column (6) of Tables 4 and 5, the estimated effects of chronic disease diagnoses remain comparable regardless of whether $Nsmk_{i,t-1}$ or $Nsmk_{i,t-2}$ is included. Such a similarity in estimates lends further support to our original estimates reported in Table 2 and 3.

Heterogeneity

To further see if our main findings are driven by some particular subgroups of individuals, we repeated our estimation of the preferred models (columns 3 and 4 of Table 2) for various subsamples defined by the values of a set of important covariates. Two sets of covariates are considered. The first includes respondents’ socio-economic characteristics, ie, residing locations (“rural areas” versus “urban areas”), years of schooling (“above median” versus “below median”), working status (“working” versus “not working”), and income (above median versus below median), which may affect individuals’ consumption habit. The second set includes statistically significant predictors of smoking decisions (Table 2, columns 3–4), ie, age (“above median” versus “below median”), whether one drinks (“yes” versus “no”), and whether other household members smoke (“yes” versus “no”). Table 6 reports the results. Each panel of the table reports the estimated Logit coefficient on the number of chronic diseases diagnosed (left column) and the corresponding marginal effect evaluated at the sample means of explanatory variables (right column) for a particular subgroup of interest. The subsample estimates are quite comparable to those full-sample estimates reported above, suggesting that our main findings of the effects of chronic disease diagnoses are not driven by some particular subset of individuals. Rather, they reflect a general pattern.

Discussion

Our analysis above suggests that chronic disease diagnoses help reduce smoking incidence in China. But the effects are small, depicting a picture somewhat different than those found in non-Chinese cohorts. For example, Keenan (2009)¹⁵ found that American adults diagnosed with chronic diseases were 3.2 times more likely to quit smoking than individuals without new diagnoses. Kim and Sambou (2019)¹⁶ found in South Korea that the odds of

Table 4 Zero-Inflated Negative Binomial and Logit Estimates of Associations Between the Number of Chronic Diseases Diagnosed and Daily Cigarette Consumption (Chinese Males Aged 50 or Above), Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	Number of Cigarettes Smoked per Day		Smoking Cessation		Smoking Cessation	
Model	Zero-Inflated Negative Binomial		Logit	Logit	Logit	Logit
	Inflation Model (Logit)	Negative Binomial Part				
Reported Estimates	Coefficient	Coefficient	Coefficient	Marginal Effects	Marginal Effects	Marginal Effects
Number of chronic diseases diagnosed (1-period lagged)	0.261***	-0.009	0.260***	0.053***	0.033***	0.039***
	(0.063)	(0.029)	(0.063)	(0.013)	(0.009)	(0.011)
Daily number of cigarettes smoked (1-period lagged)					-0.022***	
					(0.001)	
Daily number of cigarettes smoked (2-period lagged)						-0.018***
						(0.001)
Self-reported health status (1-period lagged)	-0.141	0.014	-0.124	-0.025	-0.012	-0.049**
	(0.097)	(0.043)	(0.098)	(0.020)	(0.017)	(0.020)
Drinking (1-period lagged)	-0.641***	0.046*	-0.639***	-0.130***	-0.043***	-0.086***
	(0.069)	(0.025)	(0.069)	(0.014)	(0.011)	(0.012)
Others smoke in the household (1-period lagged)	-0.261***	0.027	-0.258***	-0.053***	-0.020*	-0.026*
	(0.089)	(0.026)	(0.088)	(0.018)	(0.012)	(0.015)
Age	-0.021	0.013	-0.010	-0.002	-0.000	-0.000
	(0.052)	(0.022)	(0.052)	(0.011)	(0.008)	(0.010)
Age squared/100	0.042	-0.022	0.033	0.007	0.002	0.002
	(0.038)	(0.017)	(0.038)	(0.008)	(0.006)	(0.007)
Years of education	0.020	-0.005	0.019	0.004	0.001	0.003
	(0.013)	(0.004)	(0.013)	(0.003)	(0.002)	(0.002)
Log annual income	0.041	0.017	0.038	0.008	0.003	0.006
	(0.032)	(0.011)	(0.031)	(0.006)	(0.005)	(0.006)
Has spouse	-0.088	-0.030	-0.092	-0.019	-0.006	-0.014
	(0.140)	(0.050)	(0.138)	(0.028)	(0.019)	(0.023)
Has insurance coverage	0.033	0.071**	0.062	0.013	0.010	0.026
	(0.088)	(0.033)	(0.085)	(0.017)	(0.014)	(0.017)
Working	-0.117	0.043*	-0.116	-0.024	-0.028**	-0.023
	(0.076)	(0.024)	(0.077)	(0.016)	(0.013)	(0.015)
Survey-wave fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

(Continued)

Table 4 (Continued).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	Number of Cigarettes Smoked per Day		Smoking Cessation		Smoking Cessation	
Model	Zero-Inflated Negative Binomial		Logit	Logit	Logit	Logit
	Inflation Model (Logit)	Negative Binomial Part				
Reported Estimates	Coefficient	Coefficient	Coefficient	Marginal Effects	Marginal Effects	Marginal Effects
Community fixed effects	yes	yes	yes	yes	yes	yes
Constant	-0.657	2.535***	-0.991			
	(1.799)	(0.764)	(1.803)			
Number of individual-wave observations	8,175		8,175		8,013	6,869
Log pseudolikelihood	-18,293.08		-4844.24		-3732.57	-3501.87

Notes: Data from 2004, 2006, 2009, 2011, and 2015 waves of the CHNS are used in the analysis; In the Zero-inflation part of the Zero-inflated Negative Binomial model, the dependent variable is defined as =1 if zero-inflated and =0 if otherwise; In the Logit models predicting "smoking cessation", the dependent variable is defined as = 1 if a respondent is "not smoking" and = 0 if otherwise; The number of individual-wave observations varies somewhat across columns due to missing values in the different explanatory variables included; Standard errors are reported in parentheses, adjusted for clustering at the community level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

quitting smoking increased by 111.8% when individuals were diagnosed with a chronic disease. They also found that, as the number of chronic diseases diagnosed increased by one, the odds of quitting smoking increased by 50.6%, and the number of cigarettes smoked per day decreased by 12.9%.

There are several reasons for these differences between previous and our findings. First, perhaps due to their relatively lower income level, Chinese individuals may not pay as much attention to their health as their wealthier American and Korean counterparts do. Put differently, under the assumption that health is a normal good (the consumption of which increases as income grows), Chinese individuals may have a lower demand for it given their low income. Thus, chronic disease diagnoses do not serve as an effective walk-up call for them. Second, China has a very large smoking population (exceeding 300 million), which may have generated substantial peer effects even for smokers diagnosed with chronic diseases. Smokers diagnosed with chronic conditions may underestimate the harmful effects of smoking when seeing so many smokers (with or without chronic conditions) around them.

The small effects of chronic disease diagnoses we found imply that many Chinese elders continue to smoke even after being aware of their chronic conditions. This is

concerning, as continued smoking is very likely to exacerbate their chronic conditions. Thus, effective measures should be taken to urge chronic disease patients to quit smoking or reduce the adverse effects of continued smoking on their health. For example, the government might need to tax cigarette production more heavily to help curb cigarette consumption. Medical professionals may also need to help smoking patients change their diet to include more foods enriched with antioxidants to reduce the harmful effects of smoking.

With the above-discussed analyses and findings, our study contributes to the literature in two ways. First, to the best of our knowledge, this study is among the very few that have examined how chronic disease diagnoses affect individuals' smoking behavior in China. Thus, it provides new information that may help inform China's healthcare policy. Second, it complements previous studies of the effects of chronic-disease diagnoses on individuals' food-consumption behavior⁴² and labor-market outcomes in China,⁴³ thereby helping to depict a fuller picture of the behavioral impacts of chronic disease diagnoses in China.

Whereas our findings are informative and policy-relevant, we note a number of limitations in this study. First, due to data limitations, we do not include the diagnosis of cancer in our study. The finding of a relatively small effect of chronic disease diagnoses on smoking

Table 5 Zero-Inflated Negative Binomial and Logit Estimates of Associations Between the Diagnosis of Specific Chronic Disease and Daily Number of Cigarettes Smoked (Chinese Males Aged 50 or Above), Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	Number of Cigarettes Smoked per Day		Smoking Cessation		Smoking Cessation	
Model	Zero-Inflated Negative Binomial		Logit	Logit	Logit	Logit
	Inflation Model (Logit)	Negative Binomial Model				
Reported Estimates	Coefficient	Coefficient	Coefficient	Marginal Effects	Marginal Effects	Marginal Effects
Hypertension diagnosed (1-period lagged)	0.250***	0.001	0.241**	0.049***	0.028**	0.034**
	(0.095)	(0.034)	(0.094)	(0.019)	(0.013)	(0.016)
Diabetes diagnosed (1-period lagged)	0.325**	-0.118	0.344**	0.070**	0.046**	0.032
	(0.156)	(0.107)	(0.154)	(0.031)	(0.023)	(0.028)
Stroke diagnosed (1-period lagged)	0.144	0.107	0.124	0.025	0.026	0.040
	(0.228)	(0.077)	(0.227)	(0.046)	(0.032)	(0.039)
Myocardial infarction diagnosed (1-period lagged)	0.429*	-0.154	0.475*	0.097*	0.058	0.078
	(0.258)	(0.130)	(0.260)	(0.053)	(0.044)	(0.050)
Number of cigarettes smoked per day (1-period lagged)					-0.022***	
					(0.001)	
Number of cigarettes smoked per day (2-period lagged)						-0.018***
						(0.001)
Self-reported health status (1-period lagged)	-0.147	0.022	-0.131	-0.027	-0.012	-0.049**
	(0.097)	(0.043)	(0.097)	(0.020)	(0.017)	(0.020)
Drinking (1-period lagged)	-0.641***	0.045*	-0.637***	-0.130***	-0.043***	-0.086***
	(0.069)	(0.025)	(0.070)	(0.014)	(0.011)	(0.012)
Other smokers in the household (1-period lagged)	-0.258***	0.026	-0.256***	-0.052***	-0.020*	-0.026*
	(0.089)	(0.026)	(0.088)	(0.018)	(0.012)	(0.015)
Age	-0.021	0.013	-0.010	-0.002	-0.000	0.000
	(0.052)	(0.022)	(0.052)	(0.011)	(0.008)	(0.010)
Age squared/100	0.042	-0.022	0.033	0.007	0.002	0.002
	(0.038)	(0.017)	(0.038)	(0.008)	(0.006)	(0.007)
Years of education	0.020	-0.005	0.019	0.004	0.001	0.003
	(0.013)	(0.004)	(0.013)	(0.003)	(0.002)	(0.002)
Log annual income (yuan)	0.041	0.016	0.038	0.008	0.003	0.006
	(0.032)	(0.011)	(0.031)	(0.006)	(0.005)	(0.006)

(Continued)

Table 5 (Continued).

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Variable	Number of Cigarettes Smoked per Day		Smoking Cessation		Smoking Cessation	
Model	Zero-Inflated Negative Binomial		Logit	Logit	Logit	Logit
	Inflation Model (Logit)	Negative Binomial Model				
Reported Estimates	Coefficient	Coefficient	Coefficient	Marginal Effects	Marginal Effects	Marginal Effects
Has spouse	-0.089	-0.028	-0.094	-0.019	-0.006	-0.014
	(0.139)	(0.050)	(0.138)	(0.028)	(0.019)	(0.023)
Has insurance coverage	0.034	0.067**	0.063	0.013	0.010	0.026
	(0.088)	(0.033)	(0.086)	(0.017)	(0.014)	(0.018)
Working	-0.116	0.042*	-0.115	-0.023	-0.027**	-0.023
	(0.077)	(0.024)	(0.077)	(0.016)	(0.013)	(0.015)
Survey-wave fixed effects	yes	yes	yes	yes	yes	yes
Community fixed effects	yes	yes	yes	yes	yes	yes
Constant	-0.662	2.551***	-1.000			
	(1.805)	(0.762)	(1.808)			
Number of individual-wave observations	8,175		8,175		8,013	6,869
Log pseudolikelihood	-18,286.97		-4842.87		-3732.01	-3501.25

Notes: Data from 2004, 2006, 2009, 2011, and 2015 waves of the CHNS are used in the analysis; In the Zero-inflation part of the Zero-inflated Negative Binomial model, the dependent variable is defined as = 1 if “zero-inflated” and =0 if otherwise. In the Logit models predicting “smoking cessation”, the dependent variable is defined as = 1 if a respondent is “not smoking” and = 0 if otherwise; Standard errors are reported in parentheses, adjusted for clustering at the community level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

cessation may be partly due to our inability to incorporate cancer diagnosis in the analysis. Yet, given the lower incidence of cancer relative to other chronic diseases we examined, we believe that the inclusion of cancer diagnosis in the analysis will not overturn our main findings. Of course, future studies with more information on cancer diagnosis are certainly desirable.

Second, also due to data limitations, we have no information on CHNS respondents’ smoking behavior or their chronic conditions before 1997. Thus, we are unable to trace out their entire smoking history. Some non-smokers with chronic conditions may have stopped smoking before the 1997 survey, which may be part of the reason for the relatively small impact of chronic disease diagnoses on smoking cessation. As the CHNS continues to follow younger cohorts, interested researchers would be able to

observe more respondents’ entire smoking history and thus extend our analysis in meaningful ways.

Third, because very few females in China smoke and very few young individuals in China suffer from chronic conditions, we do not include these two groups of respondents in the analysis. Thus, our study is not generalizable to female and younger individuals in other countries. Future studies targeting these subgroups in understudied countries and regions are expected to be fruitful.

Finally, our method to control for unobserved confounding factors using the lagged outcome measure $Nsmk_{i,t-1}$ may lead to bias in our estimates, because $Nsmk_{i,t-1}$ and the error term ε_{it} might be correlated through time-invariant unobservable factors. Unfortunately, our unbalanced panel dataset (with relatively few waves) does not permit us to apply the

Table 6 Heterogeneity in the Estimated Impacts of the Number of Chronic Diseases Diagnosed

Outcome Variables	(1)	(2)	(3)	(4)
	Smoking Cessation		Smoking Cessation	
Model	Logit	Logit	Logit	Logit
Reported Estimates	Coefficient	Marginal Effects	Coefficient	Marginal Effects
	A. Rural areas		B. Urban areas	
No. chronic diseases diagnosed (current)	0.237***	0.048***	0.249***	0.047***
	(0.068)	(0.014)	(0.089)	(0.016)
Control variables included?	Yes	Yes	Yes	Yes
No. wave-individual observations	5,908	5,908	3,198	3,198
Log pseudolikelihood	-3493.551	-3493.551	-1776.697	-1776.697
	C. Education < Median		D. Education ≥ Median	
No. chronic diseases diagnosed (current)	0.213**	0.041**	0.297***	0.058***
	(0.094)	(0.018)	(0.070)	(0.013)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	3,172	3,172	5,787	5,787
Log pseudolikelihood	-1778.249	-1778.249	-3315.027	-3315.027
	E. Household income < Median		F. Household income ≥ Median	
Number of chronic diseases diagnosed (current)	0.245***	0.049***	0.243***	0.046***
	(0.082)	(0.016)	(0.071)	(0.013)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	4,293	4,293	4,738	4,738
Log pseudolikelihood	-2493.648	-2493.648	-2641.521	-2641.521
	G. Age < Median		H. Age ≥ Median	
Number of chronic diseases diagnosed (current)	0.228**	0.045**	0.302***	0.056***
	(0.098)	(0.019)	(0.067)	(0.012)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	4,455	4,455	4,506	4,506
Log pseudolikelihood	-2568.866	-2568.866	-2489.208	-2489.208

(Continued)

Table 6 (Continued).

Outcome Variables	(1)	(2)	(3)	(4)
	Smoking Cessation		Smoking Cessation	
Model	Logit	Logit	Logit	Logit
Reported Estimates	Coefficient	Marginal Effects	Coefficient	Marginal Effects
	I. Not drinking		J. Drinking	
Number of chronic diseases diagnosed (current)	0.257***	0.049***	0.251***	0.049***
	(0.082)	(0.015)	(0.087)	(0.017)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	4,408	4,408	4,560	4,560
Log pseudolikelihood	-2478.716	-2478.716	-2618.317	-2618.317
	K. Not working		L. Working	
Number of chronic diseases diagnosed (current)	0.210***	0.040***	0.320***	0.064***
	(0.063)	(0.012)	(0.109)	(0.022)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	5,227	5,227	3,760	3,760
Log pseudolikelihood	-2925.402	-2925.402	-2185.604	-2185.604
	M. No other smokers in the household		N. Other smokers in the household	
Number of chronic diseases diagnosed (current)	0.240***	0.046***	0.321***	0.061***
	(0.067)	(0.013)	(0.096)	(0.018)
Control variables included?	Yes	Yes	Yes	Yes
Number of wave-individual observations	6,200	6,200	2,737	2,737
Log pseudolikelihood	-3485.350	-3485.350	-1538.012	-3485.350

Notes: Each panel of the table reports the estimated Logit coefficient on the number of chronic diseases diagnosed (left column) and the corresponding marginal effect evaluated at the sample means of explanatory variables (right column) for a particular subgroup of interest. "Control variables" include all the variables reported in column (1) of Table 2; In the Logit models predicting "smoking cessation", the dependent variable is defined as = 1 if a respondent is "not smoking" and = 0 if otherwise; Standard errors are reported in parentheses, adjusted for clustering at the community level; ***p<0.01, **p<0.05.

system-GMM method developed by Anderson and Hsiao³³ and Arellano and Bond.³⁴ Alternatively, we assess how the potential correlation between $Nsmk_{i,t-1}$ and ε_{it} may affect our estimate of β_1 by replacing $Nsmk_{i,t-1}$ (the 1-period lagged outcome measure) with $Nsmk_{i,t-2}$ (the 2-period

lagged outcome measure) and found similar estimates of chronic disease diagnoses. However, there is no guarantee that this approach completely addresses the problem; thus, future studies with balanced panel data that can facilitate the system-GMM method are highly desirable.

Despite these limitations, we believe that our analysis has provided important findings to inform healthcare policy in China.

Conclusion

The present study, analyzing an unbalanced panel dataset on 2,986 Chinese males over the age of 50, discovered that while chronic disease diagnoses helped reduce the number of cigarettes one smokes per day, their effects are quite small. In other words, many Chinese males continued to smoke even after knowing about their chronic conditions. Relevant policies and measures are thus needed to urge chronic disease patients in China to quit smoking or reduce the adverse effects that continued smoking might impose on their health. For example, the government might need to tax cigarette production more heavily to help curb cigarette consumption. Medical professionals may also need to encourage smoking patients to switch to a diet that includes more foods enriched with antioxidants to reduce the harmful effects of smoking.

Acknowledgments

The authors thank Qiran Zhao, Zhihao Zheng, and Chen Zhu for their helpful comments on earlier versions of this paper. This work was supported by the 2115 Talent Development Program of China Agricultural University. Financial support from the Social Science Foundation of Beijing Municipality (grant number: 19JDYJB029) is also gratefully acknowledged. This research was conducted based upon the data from the China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention; the Carolina Population Center, the University of North Carolina at Chapel Hill; the National Institutes of Health (NIH; R01-HD30880, DK056350, and R01-HD38700); and the Fogarty International Center, NIH, for financial support for the CHNS data collection and analysis files since 1989.

Disclosure

The authors report no conflicts of interest in this work.

References

- World Health Organization. Global youth tobacco survey (GYTS), Timor-Leste 2019. 2021. Available from: <https://www.who.int/publications/i/item/9789290228394>. Accessed July 9, 2021.
- Hamel SB, Craig BJ. The effects of cigarette smoking on periodontal disease. *Probe*. 1997;31(6):204–206.
- Hecht SS. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst*. 1999;91(14):1194–1210. doi:10.1093/jnci/91.14.1194
- Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol*. 2004;43(10):1731–1737. doi:10.1016/j.jacc.2003.12.047
- Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *Br Med J*. 2004;328(7455):1519–1528. doi:10.1136/bmj.38142.554479.AE
- Kamholz SL. Pulmonary and cardiovascular consequences of smoking. *Clin Occup Environ Med*. 2004;88(6):1415–1430.
- Shankar A, Klein R, Klein BEK. The association among smoking, heavy drinking, and chronic kidney disease. *Am J Epidemiol*. 2006;164(3):263–271. doi:10.1093/aje/kwj173
- Benfante R, Reed D, Frank J. Does cigarette smoking have an independent effect on coronary heart disease incidence in the elderly? *Am J Public Health*. 2011;81(7):897–899. doi:10.2105/AJPH.81.7.897
- Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *Lancet*. 2011;378(9799):1297–1305. doi:10.1016/S0140-6736(11)60781-2
- Lee YCA, Hashibe M. Tobacco, alcohol, and cancer in low and high income countries. *Annals of Global Health*. 2014;80(5):378–383. doi:10.1016/j.aogh.2014.09.010
- Lubin JH, Couper D, Lutsey PL, Woodward M, Yatsuya H, Huxley RR. Risk of cardiovascular disease from cumulative cigarette use and the impact of smoking intensity. *Epidemiology*. 2016;27(3):395–404. doi:10.1097/EDE.0000000000000437
- Hernandez SL, Banks HE, Bailey AE, Bachman MJ, Kane J, Hartos JL. Relationships among chewing tobacco, cigarette smoking, and chronic health conditions in males 18–44 years of age. *J Prim Prev*. 2017;38(5):505–514. doi:10.1007/s10935-017-0485-4
- Reitsma MB, Fullman N, Ng M, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet*. 2017;389(10082):1885–1906. doi:10.1016/S0140-6736(17)30819-X
- Bünnings C. Does new health information affect health behavior? The effect of health events on smoking cessation. *Appl Econ*. 2013;49(10):1–14.
- Keenan PS. Smoking and weight change after new health diagnoses in older adults. *Arch Intern Med*. 2009;169(3):237–242. doi:10.1001/archinternmed.2008.557
- Kim DH, Sambou MO. Changes in smoking and drinking behaviors after the incidence of chronic diseases. *Addict Res Theory*. 2019;27(5):405–411. doi:10.1080/16066359.2018.1538411
- Richards M, Marti J. Heterogeneity in the smoking response to health shocks by out-of-pocket spending risk. *Health Econ Policy Law*. 2014;9(4):343–357. doi:10.1017/S1744133114000152
- Sundmacher L. The effect of health shocks on smoking and obesity. *Eur J Health Econ*. 2011;13(4):451–460. doi:10.1007/s10198-011-0316-0
- Shilian S, Jing W, Cui C, Xinchun XC. Analysis of epidemiological trends in chronic diseases of Chinese residents. *Aging Med*. 2020;3(4):226–233. doi:10.1002/agm2.12134
- National Center for Cardiovascular Disease. *Cardiovascular Disease Report in China 2018*. Beijing: Encyclopedia of China Press; 2018:1–224.
- Torre LA, Bray F, Siegel RL, Ferlay J, Tieuvent JL, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108. doi:10.3322/caac.21262
- Wang LM, Gao P, Zhang M, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *J Am Med Assoc*. 2017;317(24):2515–2523. doi:10.1001/jama.2017.7596

23. Wang WZ, Jiang B, Sun HX, et al. Prevalence, incidence and mortality of stroke in China: results from a nationwide population-based survey of 480,687 adults. *Circulation*. 2017;135(8):759–771. doi:10.1161/CIRCULATIONAHA.116.025250
24. Wang ZW, Chen Z, Zhang LF, et al. Status of hypertension in China: results from the China hypertension survey, 2012–2015. *Circulation*. 2018;137(22):2344–2356. doi:10.1161/CIRCULATIONAHA.117.032380
25. Chen Q, Chu X, Wang S, Zhang B. A triple-difference approach to re-evaluating the impact of China's new cooperative medical scheme on incidences of chronic diseases among older adults in rural communities. *Risk Manag Healthc Policy*. 2020;13:643–659. doi:10.2147/RMHP.S244021
26. Phelps CE. *Health Economics*. 4th ed. London, UK: Pearson; 2010.
27. Kenkel D, Lillard DR, Liu F. An analysis of life-course smoking behaviour in China. *Health Econ*. 2009;18:S147–S156. doi:10.1002/hec.1507
28. Yen ST, Shaw WD, Yuan Y. Cigarette smoking and self-reported health in China. *Chin Econ Rev*. 2010;21(4):532–543. doi:10.1016/j.chieco.2010.05.002
29. Marti J, Richards MR. Smoking response to health and medical spending changes and the role of Insurance. *Health Econ*. 2017;26(3):305–320. doi:10.1002/hec.3309
30. Cameron AC, Trivedi PK. *Microeconometrics: Methods and Applications*. Cambridge, UK: Cambridge university press; 2005.
31. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:675–697. doi:10.1093/biomet/73.1.13
32. Greene W. *Econometric Analysis*. 8th ed. London, UK: Pearson; 2018.
33. Anderson T, Hsiao C. Estimation of dynamic models with error components. *J Am Stat Assoc*. 1981;76:598–606. doi:10.1080/01621459.1981.10477691
34. Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev Econ Stud*. 1991;58:277–297. doi:10.2307/2297968
35. Cheng L, Liu H, Zhang Y, Shen K, Zeng Y. The impact of health insurance on health outcomes and spending of the elderly: evidence from China's new cooperative medical scheme. *Health Econ*. 2015;24(6):672–691. doi:10.1002/hec.3053
36. Li J, Shi L, Liang H, et al. Urban-rural disparities in health care utilization among Chinese adults from 1993 to 2011. *BMC Health Serv Res*. 2018;18(1):102. doi:10.1186/s12913-018-2905-4
37. Yip W, Hsiao W. China's health care reform: a tentative assessment. *Chin Econ Rev*. 2009;20(4):613–619. doi:10.1016/j.chieco.2009.08.003
38. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*. 1989;57:307–333. doi:10.2307/1912557
39. Waters EA, Muff J, Hamilton JG. Multifactorial beliefs about the role of genetics and behavior in common health conditions: prevalence and associations with participant characteristics and engagement in health behaviors. *Genet Med*. 2014;16(12):913–921. doi:10.1038/gim.2014.49
40. Zhang S, Chen Q, Zhang B. Understanding healthcare utilization in China through the Andersen behavioral model: review of evidence from the China health and nutrition survey. *Risk Manag Healthc Policy*. 2019;12:209–224. doi:10.2147/RMHP.S218661
41. Zhu C, Chen Q, Si W, Li Y, Chen G, Zhao Q. Alcohol use and depression: a Mendelian randomization study from China. *Front Genet*. 2020;11:585351. doi:10.3389/fgene.2020.585351
42. Zhao M, Konishi Y, Glewwe P. Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption. *J Health Econ*. 2013;32(2):367–385. doi:10.1016/j.jhealeco.2012.11.007
43. Liu X, Zhu C. Will knowing diabetes affect labor income? Evidence from a natural experiment. *Econ Lett*. 2014;124(1):74–78. doi:10.1016/j.econlet.2014.04.019

Risk Management and Healthcare Policy

Dovepress

Publish your work in this journal

Risk Management and Healthcare Policy is an international, peer-reviewed, open access journal focusing on all aspects of public health, policy, and preventative measures to promote good health and improve morbidity and mortality in the population. The journal welcomes submitted papers covering original research, basic science, clinical & epidemiological studies, reviews and evaluations,

guidelines, expert opinion and commentary, case reports and extended reports. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/risk-management-and-healthcare-policy-journal>