Tobacco smoking and depressive symptoms in Chinese middle-aged and older adults: Handling missing values in panel data with multiple imputation

Xiahua Du

Inner Mongolia Medical University

Rina Wu

The Affiliated Hospital of Inner Mongolia Medical University

Lili Kang

Inner Mongolia Medical University

Longlong Zhao

Inner Mongolia Medical University

Changle Li (✉ lichangle2010@sina.com)

Inner Mongolia Medical University

Research Article

Keywords: smoking, depressive symptoms, multiple imputation, China

Posted Date: February 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1319368/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Tobacco smoking and depressive symptoms in Chinese middle-aged and older adults: Handling missing values in panel data with multiple imputation

Xiahua Du¹, Rina Wu², Lili Kang³, Longlong Zhao³, Changle Li^{3#}

1 College of Humanities Education, Inner Mongolia Medical University, Hohhot, China

2 The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China

3 School of Health Management, Inner Mongolia Medical University, Hohhot, China

#Correspondance to:

Changle Li, Ph.D, Associate Professor

School of Health Management, Inner Mongolia Medical University, Jinshan Economic Development Zone,
Hohhot, China 010110.

E-mail: lichangle2010@sina.com

Abstract

Background: The high co-occurrence of tobacco smoking and depression is a major public health concern during the novel coronavirus disease-2019 pandemic. However, no studies have dealt with missing values when assessing depression. Therefore, the present study aimed to examine the effect of tobacco smoking on depressive symptoms using a multiple imputation technique.

Methods: This research was a longitudinal study using data from four waves of the China Health and Retirement Longitudinal Study conducted between 2011 and 2018, and the final sample consisted of 74,381 observations across all four waves of data collection. The present study employed a multiple imputation technique to deal with missing values, and a fixed effects logistic regression model was used for the analysis.

Results: The results of fixed effects logistic regression showed that heavy smokers were 1.2 times more likely to suffer from depressive symptoms than those who never smoked. Compared to those who never smoked, for short-term and moderate-term quitters, the odds of suffering from depressive symptoms increased by 30% and 22%, respectively. The magnitudes of the odds ratios for the variables short-term quitters, moderate-term quitters, and long-term quitters decreased in absolute terms with increasing time-gaps since quitting.

Conclusions: The empirical findings suggested that among Chinese middle-aged and older adults, heavy smokers and short-term and moderate-term quitters were more likely to have depressive symptoms than those who never smoked. Moreover, former smokers reported that the probability of having depressive symptoms decreased with a longer duration since quitting. These results may have important implications that support the government in allocating more resources to smoking cessation programs to help middle-aged and older smokers.

Keywords: smoking, depressive symptoms, multiple imputation, China

Background

Tobacco smoking is one of the biggest public health threats, resulting in more than 7 million deaths a year worldwide [1]. In China, 50.5% of adult men were current smokers in 2018, although the figure was only 2.1% for adult women. The prevalence of current smoking was 30.2% among adults aged 45-64 years and 23.1% among adults aged 65 years or older, implying that there are more than 163 million middle-aged and older adults who smoke in China. Even though tobacco smoking has been proven to be a major cause of diseases such as cancers, heart diseases, and respiratory diseases, only 16.1% of current smokers in China

plan to quit smoking within twelve months [2]. Meanwhile, depression is currently becoming a significant public health problem, with more than 300 million people estimated to suffer from depression worldwide [3]. The prevalence of depressive symptoms among older adults was 20.0%. Depression in late life is associated with an elevated risk of morbidity and suicide and decreased physical and cognitive functioning [4,5].

The novel coronavirus disease 2019 pandemic has led to adverse changes in health behaviors such as smoking and physical activity and widespread mental disorders such as depression and anxiety [6,7]. The high co-occurrence of tobacco smoking and depression is a major public health concern during this unprecedented crisis. The reciprocal relationships between tobacco smoking and depression have been widely documented; for example, depression is associated with later smoking behavior, and smoking exposure is associated with later depression [8-12]. Several previous studies have demonstrated an association between tobacco smoking and depression: tobacco smoking increases the risk of depressive symptoms. Such an association has been shown across different age groups, such as adolescents [13], adults [14,15], middle-aged and older adults [16,17], and elderly people [18]. Conversely, few studies have shown that tobacco smoking reduces depressive symptoms [19,20]. Furthermore, other studies have shown that depression predicts the persistence of tobacco smoking [21-25].

A previous study found that, among Chinese middle-aged and older adults, smokers were less likely to develop depressive symptoms, which is inconsistent with previous findings in developed countries [20]. The main reason for this finding may be that there were 1,965 missing 10-item Center for Epidemiologic Studies Depression Scale (CESD-10) baseline scores and 2,888 missing CESD-10 follow-up scores among respondents in both interviews. Missing values are a common challenge in most social research studies, and

the problem is often pronounced in studies using self-rated instruments. Similarly, previous studies on smoking and depression may have encountered the problem of missing values [26,27]. Missing values reduce statistical power, cause bias in the estimation of the parameters, and lessen the representativeness of the samples [28].

Missingness mechanisms were first introduced by Rubin in 1976 [29]. Rubin distinguished three fundamental missing-values mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MCAR occurs when the missingness is unrelated to the observed and unobserved value for a unit. Under an MAR mechanism, the probability of a missing value for an item may depend on observed data but not on unobserved data. MNAR means that the probability of missingness depends on the underlying value of an item. Many researchers adopt a listwise deletion approach (complete-case analysis) to deal with missing values. This approach is based on the assumption of MCAR. However, this assumption is sometimes difficult to justify in practice [30]. Therefore, various imputation methods have been developed to compensate for missing values in survey data. The methods include random selection, preceding question, question mean, individual mean, single regression, and multiple imputation. Multiple imputation is the most accurate technique for dealing with missing values when assessing depressive symptoms [26].

Overall, the difficulty of estimating the effect of smoking status on depression has led to conflicting findings in the general population. Moreover, although the association between tobacco smoking and depression has been documented, thus far, there have been no studies dealing with missing values when assessing depression. To bridge these gaps, the present study is aimed to examine the effect of tobacco smoking on depressive symptoms using panel data with a multiple imputation technique.

Methods

Data Source

The data used in this study were obtained from the China Health and Retirement Longitudinal Study (CHARLS), launched by the National School of Development of Peking University. The CHARLS sample was drawn from 28 provinces, 150 counties/districts, and 450 villages/residential committees. Multistage stratified sampling with a probability proportional to size was used for the survey. More details on the sampling procedure and data collection process are available in the study by Zhao et al. [31]. The CHARLS questionnaires include sections on household information, demographic background, health status and physical functioning, health care and insurance, work, retirement, pension, income, expenditure, assets, housing characteristics, etc.

The CHARLS is a biennially longitudinal survey of Chinese families and individuals aged older than 45 years. In the first wave in 2011, 17,337 persons (older than 45 years old) were successfully interviewed; in the three waves of full sample follow-up surveys in 2013, 2015, and 2018, 18,248 persons, 20,083 persons, and 19,584 persons were successfully interviewed, respectively. The measured variables and their respective percentages of missing values are presented in Table 1. After eliminating the measured variables with missing values less than 1% (871 persons), the final sample consisted of a total of 74,381 persons for the data collection waves, 58,854 with no missing values and an additional 15,527 with missing values in at least one of the measured variables.

-----Table 1 about here-----

Measurements

Depressive symptoms

The CESD-10 was used to assess depressive symptoms in the CHARLS questionnaire. First, the CESD-10 was developed from the full-length version of the 20-item Center for Epidemiologic Studies Depression Scale (CESD-D), which was designed as a screening instrument for depressive symptoms in older adults [32]. Second, the CESD-10 indicated good predictive accuracy when compared to the CESD-D. Last, the CESD-10 has shown reasonable validity and reliability in the Chinese population and adequate validity across a range of ages in longitudinal studies [33,34].

In the CHARLS, each respondent was asked, ‘How often have you felt or behaved this way during the last week?’. The survey consists of 10 items (e.g., I felt depressed, I felt fearful, my sleep was restless, and I was happy), which can be rated on a 4-point Likert scale from 0 (less than 1 day) to 3 (5-7 days). The range of the CESD-10 total score is 0 to 30, and a cutoff score of 10 or higher indicates the presence of significant depressive symptoms [32]. Therefore, the variable of depressive symptoms was set as a dichotomous variable that equaled 1 if the individual self-rated CESD-10 score was equal to or greater than 10 and equaled 0 if otherwise.

Smoking status

In the CHARLS, each adult was asked, ‘Have you ever chewed tobacco, smoked a pipe, smoked self-rolled cigarettes, or smoked cigarettes/cigars?’ and ‘Do you still have a smoking habit or have you totally stopped smoking?’. According to these two questions, all adults were divided into three mutually exclusive groups: never smoked, current smokers, and former smokers. For further analysis, the present study categorized current smokers into three subgroups (light, moderate, and heavy smokers) based on their average cigarette consumption (‘In one day about how many cigarettes do you consume?’). Light smokers were current smokers who reported consuming from 1-10 cigarettes per day, moderate smokers were those who

consumed from 11-19 cigarettes per day, and heavy smokers were those who consumed 20 cigarettes or more per day. Former smokers were categorized into three subgroups (short-term, moderate-term, and long-term quitters) based on the total number of years since the respondents had quit smoking ('At what age did you totally quit smoking?'): short-term quitters were former smokers who had quit smoking ≤ 1 year age, and moderate-term quitters and long-term quitters were former smokers who had quit smoking 2-5 years and ≥ 6 years age, respectively [35].

Covariates

The analysis also considered the following three categories of variables as covariates to explain depressive symptoms: (1) current health-related factors, including self-rated health, functional limitations, and chronic conditions; (2) several demographic characteristics that may also affect depressive symptoms, such as sex, age, educational attainment, marital status, and rural residency; and (3) in addition to smoking, another health behavior factor included in the analysis was drinking. The definitions of the covariates are provided in Table 2.

-----Table 2 about here-----

Multiple imputation of missing values

During the CHARLS investigation, the respondents were not required to answer any question in the cognition and depression section that they do not want to answer, and the interviewers went on to the next question. Moreover, the CESD-10 must be answered by the respondents themselves and cannot be answered by other family members. As a result, the proportion of missing values for the CESD-10 items was approximately 10% (see Table 1). In addition, the respondents who smoked filtered or unfiltered cigarettes answered the question about average cigarette consumption. The respondents who smoked a

pipe, self-rolled cigarettes, cigars, or water cigarettes skipped the average cigarette consumption question.

Therefore, missing values accounted for approximately 12% of the values for the variable of smoking status. The present study adopted the MAR assumption and employed a multiple imputation technique to deal with missing values.

Multiple imputation is a simulation-based statistical technique that allows researchers to increase the availability of data points, thus reducing biases associated with the deletion of observations due to missing values [36]. Multiple imputation has three elemental phases: imputation, analysis, and pooling. In the imputation phase, m copies of the dataset are created, with the missing values replaced by imputed values using an appropriate model. Rubin suggested that $m=5$ should be sufficient to obtain valid inference [37]. Therefore, 5 copies were created in this study to reduce the sampling error due to imputations. The imputation process was carried out based on multiple imputation by chained equations (MICE). MICE is a practical approach to impute missing data for multiple variables based on a set of univariate imputation models [38]. The present study selected conditional models based on the type of variables. MICE allows the use of logistic and Poisson regression models to impute binary variables, such as rural residency, and count variables, such as chronic conditions. Moreover, ordered logistic and multinomial logistic regression models can impute ordered categorical variables such as the CESD-10 items, educational attainment, and self-rated health, and unordered categorical variables, such as smoking status. Multiple imputation should include variables associated with the probability of missing values [39]. The variables listed in Table 2 were used in the imputation models.

In the analysis phase, each of the 5 completed datasets was analyzed using a desired statistical method. The results obtained from 5 the completed datasets were combined into a single multiple-imputation result

in the pooling phase. The single parameter estimate is calculated as the mean of the m ($= 5$) parameter estimates of $\hat{Q}$:

$$\bar{Q} = m^{-1} \sum_{i=1}^m \hat{Q}_i$$

The estimated variance of this MI estimate is calculated based on Rubin's rules as expressed below:

$$T = \bar{U} + (1 + \frac{1}{m})B$$

where $\bar{U} = m^{-1} \sum_{i=1}^m U_i$ is the estimated within imputation variance and U_i is the estimated variance from imputed data. $B = (m - 1)^{-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2$ is the between imputation variance [30].

Statistical analysis

Depressive symptoms may be both an antecedent and a consequence of tobacco smoking [10]. Panel data, also referred to as longitudinal data in epidemiology, are a dataset in which observations of multiple subjects are collected over time. The benefit of panel data is that it is possible to control for the unknown or unmeasured determinants of depressive symptoms that are constant over time. Furthermore, since panel data cover repeated measures of tobacco smoking and depression symptoms, it is much easier to estimate whether changes in tobacco smoking precede changes in depressive symptoms or vice versa [40]. Based on a four-wave unbalanced panel dataset, the current study estimated the effect of tobacco smoking on depressive symptoms and used a logistic regression model. The logistic regression model built a latent regression and was defined as follows:

$$y_{it}^* = x'_{it}\alpha + \beta_1 Hsmoker_{it} + \beta_2 Msmoker_{it} + \beta_3 Lsmoker_{it} + \beta_4 STsmoker_{it} + \beta_5 MTsmoker_{it} + \beta_6 LTsmoker_{it} + \varepsilon_{it}, \quad i = 1, \dots, n, t = 1, \dots, T_i$$

y_{it}^* is an unobserved latent variable linked to the observed binary response variable (with or without depressive symptoms). x'_{it} is the vector of the demographic characteristics and health status of an individual. $Hsmoker_{it}$, $Msmoker$, $Lsmoker_{it}$, $STsmoker_{it}$, $MTsmoker_{it}$, and $LTsmoker_{it}$ are dummy variables that were defined as heavy smokers, moderate smokers, light smokers, short-term quitters, moderate-term quitters, and long-term quitters. μ_i is the unobserved and individual-specific heterogeneity, and ε_{it} is a time-varying error term.

A logistic regression model was performed to analyze the impact of tobacco smoking on depressive symptoms among middle-aged and older adults in China. The first step in the analysis, pooled logistic regression, was a starting point. After that, this study treated the data as a panel structure and made a choice between the fixed effects and random effects logistic model. In this study, a possible unobserved variable was attitudes toward smoking, which was correlated with the time-varying explanatory variables (tobacco smoking) in the model. With such correlated heterogeneity, a fixed effects logistic model should be preferred over a random effects logistic model; however, when estimating a fixed effects logistic model, many pieces of information are lost. A random effects logistic model was also presented in this study [40,41]. The results are presented as odds ratios (ORs) along with 95% confidence intervals (CIs). All statistical analyses were conducted employing the Stata 15 statistical software package.

Results

A descriptive summary of all variables over time is displayed in Table 3. The total sample size was 74,381 respondents, with 51.76% of the respondents being female and 32.61% aged 65 years or older. In addition, 77.88% of the respondents reported living in rural areas, and approximately 34% of the respondents completed at least middle school. The proportions of respondents with depressive symptoms were 37.55%

in Wave 1, 32.26% in Wave 2, 33.68% in Wave 3, and 38.50% in Wave 4, implying that the proportions of respondents with depressive symptoms first decreased and then increased. Approximately one in four respondents has been current smokers over the years, and the proportions of light, moderate, and heavy smokers were approximately 9%, 2%, and 14%, respectively. Moreover, approximately 10% of the respondents were former smokers from 2011 to 2018. Among former smokers, the proportions of short-term, moderate-term, and long-term quitters were approximately 2%, 2%, and 6%, respectively.

-----Table 3 about here-----

The results of the logistic regression analysis are shown in Table 4 as ORs. Column III of Table 4 presents the effect of tobacco smoking on depressive symptoms using the fixed effects logistic model. The results revealed that smoking status was associated with depressive symptoms. Heavy smokers were 1.2 times more likely to suffer from depressive symptoms than those who never smoked (OR=1.20; 95% CI: 1.05, 1.37). Compared to those who never smoked, for short-term and moderate-term quitters, the odds of suffering from depressive symptoms increased 30% and 22%, respectively (OR=1.30; 95% CI: 1.04, 1.63, OR=1.22, 95% CI: 1.02, 1.47). In particular, long-term quitters had an increased likelihood of suffering from depressive symptoms than those who never smoked, but the difference was not statistically significant. The magnitudes of the ORs for the variables short-term quitters, moderate-term quitters, and long-term quitters decreased in absolute terms with increasing time-gaps since quitting. Therefore, among former smokers, the probability of suffering from depressive symptoms decreased with increasing duration of since quitting smoking.

Irrespective of the estimation method, heavy smokers showed an increased probability of suffering from depressive symptoms compared to those who had never smoked. Short-term quitters showed an

increased probability of suffering from depressive symptoms compared to those who had never smoked (see Columns I-III of Table 4).

-----Table 4 about here-----

Discussion

To the best of our knowledge, this is the first research study that examined the effect of tobacco smoking on depressive symptoms among individuals aged 45 years or older in China using a nationally representative survey with a multiple imputation technique. The results indicated that among Chinese middle-aged and older adults, heavy smokers and short-term and moderate-term quitters were more likely to have depressive symptoms than those who never smoked after controlling for other relevant variables. Our findings are consistent with the findings for European middle-aged and older adults [16] and American middle-aged adults [17]. Therefore, local communities and primary care facilities should consider promoting health education programs for middle-aged and older smokers and improving their understanding of the hazards of tobacco smoking.

This type of analysis, however, does not identify pathways between tobacco smoking and depressive symptoms. One possible explanation is that smoking or chewing tobacco releases nicotine affecting an individual's neurocircuitry, which increases their susceptibility to depression [13,15]. Another possible explanation is from a self-medication model, suggesting that smokers use nicotine to alleviate depressed mood [9,18,42].

This study found that short-term and moderate-term quitters were more likely to have depressive symptoms than those who never smoked regardless of the time elapsed since quitting, but the likelihood of having depressive symptoms declined with increasing time gaps since stopping smoking. Although the

causality between smoking cessation and depression cannot be established without a longitudinal follow-up design, the decline in the prevalence of depressive symptoms after one year of quitting implies that smoking cessation and depressive symptoms are related. These results are comparable to other findings in the literature [43-45]. In this study, former smokers reported that the probability of having depressive symptoms decreased with a longer duration since quitting, and hence, the earlier a smoker stops smoking, the greater the impact of smoking cessation on not being depressed. In other words, smoking cessation is an effective way to reduce middle-aged and older adults' risk of depression. Therefore, the government should consider allocating more resources to smoking cessation programs to help adults and adolescent smokers quit as early as possible and/or to remain nonsmokers. It is also essential to tailor smoking cessation programs for middle-aged and older adults and help them quit smoking and prevent relapse.

Extensive health-related surveys, such as the CHARLS, provide numerous data regarding health-related behaviors and health outcomes. However, almost every analysis faces the annoying problem of missing data. Multiple imputation recovers a fully observed sample size. More importantly, multiple imputation restores the natural variability of the missing values. Recovering information and restoring variability may reduce bias or increase precision, which results in a valid statistical inference from multiple imputation [28,46]. It will be essential to employ multiple imputation in future analyses, when survey items on smoking status and depressive symptoms have a large number of missing values.

Although the current study employs a national survey with multiple imputation to analyze the effect of smoking status on depressive symptoms in Chinese middle-aged and older adults, several limitations should be emphasized. First, since the CESD-10 is a self-reported screening instrument for symptoms of depression and not a diagnostic tool, our analysis may have resulted in an underestimation of the

association between tobacco smoking and depressive symptoms. Second, the data were obtained via surveys, and thus the limitations of all self-reported data exist, such as recall bias and the unreliability of responses when respondents are under pressure. Third, when the respondents' unchanging depressive symptoms across all four waves did not contribute to the likelihood, the results of the fixed effects model could be less precise and have larger standard errors.

Conclusions

The purpose of this study was to empirically ascertain the effect of tobacco smoking on depressive symptoms among individuals aged 45 years or older in China using multiple imputed datasets. The empirical findings suggest that among Chinese middle-aged and older adults, heavy smokers, and short-term and moderate-term quitters were more likely to have depressive symptoms than those who never smoked. Moreover, former smokers reported that the probability of having depressive symptoms decreased with a longer duration since quitting. These results may have important implications that support the government in allocating more resources to smoking cessation programs to help middle-aged and older smokers.

Abbreviations

CESD-10:10-item Center for Epidemiologic Studies Depression Scale; MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random; CHARLS: China Health and

Retirement Longitudinal Study; CESD-D: 20-item Center for Epidemiologic Studies Depression Scale;
MICE: multiple imputation by chained equations; OR: odds ratio; CI: confidence interval.

Declarations

Ethics approval and consent to participate

For this research, we used a publicly available secondary dataset with all individual identifiers removed prior to making the dataset publicly available. No ethical approval was required due to the type and nature of the dataset used.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Peking University Open Research Data Platform repository: <https://opendata.pku.edu.cn/dataverse/CHARLS..>

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Inner Mongolia Autonomous Region Natural Science Fund (2021BS07006). The funding bodies played no role in the design of the study, the collection, analysis, and interpretation of the data or the in writing of the manuscript.

Authors' contributions

XD and CL collaboratively designed the study, developed the methodology, and interpreted the results. CL led the data analysis and wrote the manuscript. RW, LK and LZ made important contributions to the revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Not Applicable

References

1. World Health Organization. Tobacco. 2021. <https://www.who.int/news-room/fact-sheets/detail/tobacco>. Accessed 12 Jan 2022.
2. World Health Organization. Global Adult Tobacco Survey (GATS). Fact sheet China 2018. 2019. <https://www.who.int/docs/default-source/wpro---documents/countries/china/2018-gats-china-factsheet-cn-en.pdf>. Accessed 10 Jan 2022.
3. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. *J Psychiatr Res*. 2020; 126:134-40.
4. Blazer DG. Depression in late life: review and commentary. *J Gerontol A Biol Sci Med Sci*. 2003;58:249-65.
5. Tang T, Jiang J, Tang X. Prevalence of depressive symptoms among older adults in mainland China: A systematic review and meta-analysis. *J Affect Disord*. 2021;293:379-90.
6. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, Anxiety and Stress during COVID-19: Associations with Changes in Physical Activity, Sleep, Tobacco and Alcohol Use in Australian Adults. *Int J Environ Res Public Health*. 2020;17:4065.

7. Tzu-Hsuan Chen D. The psychosocial impact of the COVID-19 pandemic on changes in smoking behavior: Evidence from a nationwide survey in the UK. *Tob Prev Cessat.* 2020;6:59.
8. Audrain-McGovern J, Rodriguez D, Kassel JD. Adolescent smoking and depression: evidence for self-medication and peer smoking mediation. *Addiction.* 2009;104:1743-56
9. Chaiton MO, Cohen JE, O'Loughlin J, Rehm J. A systematic review of longitudinal studies on the association between depression and smoking in adolescents. *BMC Public Health.* 2009;9:356.
10. Liew HP, Gardner S. The interrelationship between smoking and depression in Indonesia. *Health Policy Technol.* 2016; 5:26-31.
11. An R, Xiang X. Smoking, heavy drinking, and depression among U.S. middle-aged and older adults. *Prev Med.* 2015;81:295-302.
12. Wootton RE, Richmond RC, Stuijtzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med.* 2020;50:2435-43.
13. Boden JM, Fergusson DM, Horwood LJ. Cigarette smoking and depression: tests of causal linkages using a longitudinal birth cohort. *Br J Psychiatry.* 2010;196:440-6.
14. Kang E, Lee J. A longitudinal study on the causal association between smoking and depression. *J Prev Med Public Health.* 2010;43:193-204.
15. Wiesbeck GA, Kuhl HC, Yaldizli O, Wurst FM; WHO/ISBRA Study Group on Biological State and Trait Markers of Alcohol Use and Dependence. Tobacco smoking and depression--results from the WHO/ISBRA study. *Neuropsychobiology.* 2008;57:26-31.
16. Werneck AO, Peralta M, Tesler R, Marques A. Cross-sectional and prospective associations of lifestyle risk behaviors clustering with elevated depressive symptoms among middle-aged and older adults. *Maturitas.* 2022;155:8-13.
17. Johnson EO, Breslau N. Is the association of smoking and depression a recent phenomenon? *Nicotine Tob Res.* 2006;8:257-62.
18. Lam TH, Li ZB, Ho SY, Chan WM, Ho KS, Li MP, et al. Smoking and depressive symptoms in Chinese elderly in Hong Kong. *Acta Psychiatr Scand.* 2004;110:195-200.

19. Magee W, Clarke P. The effect of smoking on depressive symptoms. *Addict Behav.* 2021;112:106641.
20. Cheng HG, Chen S, McBride O, Phillips MR. Prospective relationship of depressive symptoms, drinking, and tobacco smoking among middle-aged and elderly community-dwelling adults: Results from the China Health and Retirement Longitudinal Study (CHARLS). *J Affect Disord.* 2016;195:136-43.
21. Holahan CK, Holahan CJ, Powers DA, Hayes RB, Marti CN, Ockene JK. Depressive symptoms and smoking in middle-aged and older women. *Nicotine Tob Res.* 2011;13:722-31.
22. Fergusson DM, Goodwin RD, Horwood LJ. Major depression and cigarette smoking: results of a 21-year longitudinal study. *Psychol Med.* 2003;33:1357-67.
23. Weinberger AH, Kashan RS, Shpigel DM, Esan H, Taha F, Lee CJ, et al. Depression and cigarette smoking behavior: A critical review of population-based studies. *Am J Drug Alcohol Abuse.* 2017;43:416-31.
24. Stubbs B, Vancampfort D, Firth J, Solmi M, Siddiqi N, Smith L, Carvalho AF, Koyanagi A. Association between depression and smoking: A global perspective from 48 low- and middle-income countries. *J Psychiatr Res.* 2018;103:142-9.
25. Acartürk CZ, Nierkens V, Agyemang C, Stronks K. Depressive symptoms and smoking among young Turkish and Moroccan ethnic minority groups in The Netherlands: a cross-sectional study. *Subst Abuse Treat Prev Policy.* 2011;6:5.
26. Shrive FM, Stuart H, Quan H, Ghali WA. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Med Res Methodol.* 2006;6:57.
27. Andresen EM, Byers K, Friary J, Kosloski K, Montgomery R. Performance of the 10-item Center for Epidemiologic Studies Depression scale for caregiving research. *SAGE Open Med.* 2013;1:2050312113514576.
28. Kang H. The prevention and handling of the missing data. *Korean J Anesthesiol.* 2013;64:402-6.
29. Rubin, D.B., Inference and missing data. *Biometrika*, 1976; 63:581-92.
30. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol.* 2017;9:157-66.

31. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol*. 2014;43:61-8.
32. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994;10:77-84.
33. Cheng ST, Chan AC. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. *Int J Geriatr Psychiatry*. 2005;20:465-70.
34. Ossip-Klein D, Rothenberg B, Andresen E, Depression screening measures. In: Andresen E, Rothenberg B, Zimmer J editors. *Assessing the health status of older adults*. New York: Springer Publishing Co; 1997. p.180-244.
35. Canada government. Tobacco use statistics. 2008. <https://www.canada.ca/en/health-canada/services/health-concerns/tobacco/research/tobacco-use-statistics/terminology.html>. Accessed 20 Dec 2021.
36. Penn DA. Estimating missing values from the general social survey: An application of multiple imputation. *Soc Sci Quart*. 2007; 88: 573-84.
37. Rubin, D.B., *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons;1987.
38. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30:377-99.
39. Spratt M, Carpenter J, Sterne JA, Carlin JB, Heron J, Henderson J, et al. Strategies for multiple imputation in longitudinal studies. *Am J Epidemiol*. 2010;172:478-87.
40. Andreß HJ, Golsch K, and Schmidt AW. *Applied panel data analysis for economic and social surveys*. Berlin:Springer Science & Business Media; 2013.
41. Greene WH. *Econometric analysis* 7th edition. International edition. New Jersey: Prentice Hall; 2012.
42. Fluharty M, Taylor AE, Grabski M, Munafò MR. The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine Tob Res*. 2017;19:3-13.
43. McClave AK, Dube SR, Strine TW, Kroenke K, Caraballo RS, Mokdad AH. Associations between smoking cessation and anxiety and depression among U.S. adults. *Addict Behav*. 2009;34:491-7.

44. Piper ME, Rodock M, Cook JW, Schlam TR, Fiore MC, Baker TB. Psychiatric diagnoses among quitters versus continuing smokers 3 years after their quit day. *Drug Alcohol Depend.* 2013;128:148-54.
45. Prochaska JJ, Hall SM, Tsoh JY, Eisendrath S, Rossi JS, Redding CA, et al. Treating tobacco dependence in clinically depressed smokers: effect of smoking cessation on mental health functioning. *Am J Public Health.* 2008;98:446-8.
46. Lee KJ, Carlin JB. Recovery of information from multiple imputation: a simulation study. *Emerg Themes Epidemiol.* 2012;9:3.

Table 1 Frequencies and percentages of missing values for the measured variables

Variables	Complete Response, N	Complete %	Missing Response, N	Missing %
<i>Depressive symptoms</i>				
CESD-10-01	67,454	89.64	7,798	10.36
CESD-10-02	66,661	88.58	8,591	11.42
CESD-10-03	67,362	89.52	7,890	10.48
CESD-10-04	67,413	89.58	7,839	10.42
CESD-10-05	65,694	87.30	9,558	12.70
CESD-10-06	68,008	90.37	7,244	9.63
CESD-10-07	68,081	90.47	7,171	9.53
CESD-10-08	67,726	90.00	7,526	10.00
CESD-10-09	67,704	89.97	7,548	10.03
CESD-10-10	67,376	89.53	7,876	10.47
<i>Smoking status</i>	66,228	88.01	9,024	11.99
<i>Sex</i>	75,203	99.93	49	0.07
<i>Age</i>	74,830	99.44	422	0.56
<i>Educational attainment</i>	73,461	97.62	1,791	2.38
<i>Marital status</i>	75,145	99.86	107	0.14
<i>Rural residency</i>	72,725	96.64	2,527	3.36
<i>Self-rated health</i>	72,089	95.80	3,163	4.20
<i>Functional limitations</i>	74,781	99.37	471	0.63
<i>Chronic conditions</i>	72,614	96.49	2,638	3.51
<i>Drinking</i>	74,607	99.14	645	0.86

Note: 'Don't know', 'Refused' and 'Blank' equal missing.

Table 2: Definitions of the covariates used in the empirical models

Variable	Description
<i>Sex</i>	1 if the individual was male; 0 for females
<i>Age</i>	Continuous variable, actual age in years
<i>Educational attainment</i>	
Illiteracy	1 if the individual was illiterate; 0 if otherwise
Elementary school	1 if the individual attended elementary school; 0 if otherwise
Middle school	1 if the individual graduated from middle school; 0 if otherwise
High school	1 if the individual graduated from high school; 0 if otherwise
Above three-year of college	1 if the individual graduated from an above three-year college; 0 if otherwise
<i>Marital status</i>	
Never married	1 if the individual was never married; 0 if otherwise
Married	1 if the individual was married; 0 if otherwise
Separated or divorced	1 if the individual was separated or divorced; 0 if otherwise
Widowed	1 if the individual was widowed; 0 if otherwise
<i>Rural residency</i>	1 if the individual was rural resident; 0 if otherwise
<i>Self-rated health</i>	
Poor	1 if the individual reported their health status to be poor or very poor; 0 if otherwise
Fair	1 if the individual reported their health status to be fair; 0 if otherwise
Good	1 if the individual reported their health status to be good or very good; 0 if otherwise
<i>Functionally limitations</i>	
None	1 if the individual could eat, toilet, dress, bathe/shower, get in/out of bed, and walk without difficulty; 0 if otherwise
Mild	1 if the individual had one or two difficulties in eating, toileting, dressing, bathing/showering, getting in/out of bed, or walking; 0 if otherwise
Moderate	1 if the individual had three or four difficulties in eating, toileting, dressing, bathing/showering, getting in/out of bed, or walking; 0 if otherwise
Severe	1 if the individual had five to six difficulties in eating, toileting, dressing, bathing/showering, getting in/out of bed, or walking; 0 if otherwise
<i>Chronic conditions</i>	The number of chronic conditions diagnosed by a doctor.
<i>Drinking</i>	1 if the individual drank at least once a month; 0 if otherwise

Table 3 Description of variables in three waves (percentage)

Variable	Wave1 (2011) N=17,108	Wave2 (2013) N=17,940	Wave3 (2015) N=19,832	Wave4 (2018) N=19,501	All waves N=74,381
<i>Depressive symptoms</i>					
Yes	37.55	32.26	33.68	38.50	35.49
No	62.45	67.74	66.32	61.50	64.51
<i>Smoking status</i>					
Never smoked	64.20	67.50	60.50	60.19	62.96
Short-term quitters	2.17	2.10	1.46	3.33	2.27
Moderate-term quitters	2.51	1.55	1.45	4.30	2.46
Long-term quitters	5.23	5.63	5.90	8.12	6.26
Light smokers	8.65	8.95	11.17	8.34	9.31
Moderate smokers	1.57	1.61	2.46	1.88	1.90
Heavy smokers	15.67	12.66	17.06	13.84	14.84
<i>Sex</i>					
Male	48.68	48.31	48.47	47.56	48.24
Female	51.32	51.69	51.53	52.44	51.76
<i>Age</i>					
45-64 years	72.48	69.37	67.49	60.99	67.39
65-84 years	26.34	29.16	30.96	36.89	31.02
≥85 years	1.18	1.47	1.55	2.12	1.59
<i>Educational attainment</i>					
Illiterate	27.42	26.68	24.88	22.78	25.35
Elementary school	39.16	39.43	40.44	42.78	40.51
Middle school	20.67	20.85	21.48	21.80	21.23
High school	10.31	10.64	10.49	10.51	10.49
Above three-year college	2.44	2.40	2.71	2.13	2.42
<i>Marital status</i>					
Never married	0.89	0.85	0.78	0.61	0.78
Married	87.08	86.78	86.40	85.03	86.29
Separated or divorced	1.32	1.20	1.22	1.58	1.33
Widowed	10.71	11.17	11.60	12.78	11.60
<i>Rural residency</i>					
Yes	77.48	76.57	78.02	79.28	77.88
No	22.52	23.43	21.98	20.72	22.12

<i>Self-rated health</i>						
Poor	29.71	27.21	25.73	26.60	27.23	
Fair	46.54	49.11	50.30	48.67	48.72	
Good	23.75	23.68	23.97	24.73	24.05	
<i>Functional limitations</i>						
None	83.38	82.48	80.43	81.45	81.87	
Mild	11.38	12.74	14.00	12.52	12.71	
Moderate	3.27	3.24	3.67	3.76	3.50	
Severe	1.97	1.54	1.90	2.27	1.92	
<i>Chronic conditions</i>						
Yes	67.26	66.96	61.45	79.35	68.81	
No	32.74	33.04	38.55	20.65	31.19	
<i>Drinking</i>						
Yes	33.26	34.61	35.48	33.76	34.31	
No	66.74	65.39	64.52	66.24	65.69	

Table 4 Logistic regression analysis of depressive symptoms

	Pooled logistic (I) ORs (95% CI)	Random effects logistic (II) ORs (95% CI)	Fixed effects logistic (III) ORs (95% CI)
<i>Smoking status</i>			
Never smoked (ref.)	1.00	1.00	1.00
Short-term quitters	1.12 (0.97-1.28)	1.19 (1.01-1.40)	1.30 (1.04-1.63)
Moderate-term quitters	1.05 (0.93-1.19)	1.11 (0.95-1.29)	1.22 (1.02-1.47)
Long-term quitters	0.93 (0.82-1.05)	0.95 (0.82-1.10)	1.05 (0.88-1.26)
Light smokers	1.14 (1.04-1.26)	1.17 (1.04-1.30)	1.10 (0.96-1.27)
Moderate smokers	1.03 (0.88-1.21)	1.06 (0.89-1.27)	1.04 (0.83-1.29)
Heavy smokers	1.14 (1.06-1.23)	1.20 (1.09-1.31)	1.20 (1.05-1.37)
<i>Sex</i>			
Male	0.61 (0.57-0.66)	0.54 (0.49-0.59)	0.88 (0.54-1.43)
Female (ref.)	1.00	1.00	1.00
<i>Age</i>	0.99 (0.98-0.99)	0.99 (0.98-0.99)	1.01 (1.00-1.02)
<i>Educational attainment</i>			
Illiteracy (ref.)	1.00	1.00	1.00
Elementary school	0.93 (0.88-0.98)	0.92 (0.86-0.98)	1.11 (0.96-1.28)
Middle school	0.70 (0.66-0.75)	0.65 (0.60-0.71)	0.98 (0.78-1.25)
High school	0.59 (0.54-0.65)	0.51 (0.46-0.57)	0.82 (0.59-1.13)
Above three-year college	0.53 (0.44-0.64)	0.45 (0.36-0.55)	0.89 (0.50-1.60)
<i>Marital status</i>			
Never married (ref.)	1.00	1.00	1.00
Married	0.64 (0.51-0.80)	0.57 (0.43-0.76)	1.17 (0.50-2.78)
Separated or divorced	1.23 (0.94-1.61)	1.26 (0.90-1.76)	1.90 (0.78-4.64)
Widowed	0.92 (0.73-1.16)	0.91 (0.68-1.21)	1.96 (0.84-4.54)
<i>Rural residency</i>			
Yes	1.52 (1.43-1.61)	1.67 (1.55-1.79)	1.08 (0.88-1.32)
No (ref.)	1.00	1.00	1.00
<i>Self-rated health</i>			
Poor (ref.)	1.00	1.00	1.00
Fair	0.42 (0.40-0.44)	0.37 (0.35-0.39)	0.52 (0.48-0.55)
Good	0.21 (0.20-0.22)	0.18 (0.16-0.19)	0.35 (0.32-0.38)
<i>Functional limitations</i>			
None (ref.)	1.00	1.00	1.00
Mild	2.07 (1.96-2.18)	2.24 (2.10-2.39)	1.61 (1.49-1.74)

Moderate	3.63 (3.16-4.18)	4.49 (3.77-5.33)	2.91 (2.38-3.55)
Severe	4.98 (4.14-5.99)	7.28 (5.75-9.22)	4.97 (3.54-6.99)
<i>Chronic conditions</i>			
Yes	1.15 (1.14-1.17)	1.20 (1.18-1.22)	1.09 (1.06-1.13)
No (ref.)	1.00	1.00	1.00
<i>Drinking</i>			
Yes	1.01 (0.97-1.06)	1.01 (0.95-1.06)	1.02 (0.94-1.11)
No (ref.)	1.00	1.00	1.00
Observations	74,381	74,381	35,466

Bold indicates statistical significance, $p < 0.05$.