Journal of Gerontological Social Work

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/wger20

# Late Life Cognitive Function Trajectory Among the Chinese Oldest-Old Population—A Machine Learning Approach

Jierong Hu, Minzhi Ye & Juan Xi

JOURNAL OF GERONTOLOGICAL SOCIAL WORK
https://doi.org/10.1080/01634372.2024.2339982



# Late Life Cognitive Function Trajectory Among the Chinese Oldest-Old Population—A Machine Learning Approach

Jierong Hu[a], Minzhi Ye[b], and Juan Xi[c]

[a]Department of Innovative Social Work, City University of Macau, Macau, China; [b]School of Lifespan Development and Educational Science, Kent State University, Kent, USA; [c]Department of Sociology, Akron University, Akron, USA

**ABSTRACT**
Informed by the biopsychosocial framework, our study uses the Chinese Longitudinal Healthy Longevity Survey (CLHLS) dataset to examine cognitive function trajectories among the oldest-old (80+). Employing K-means clustering, we identified two latent groups: High Stability (HS) and Low Stability (LS). The HS group maintained satisfactory cognitive function, while the LS group exhibited consistently low function. Lasso regression revealed predictive factors, including socioeconomic status, biological conditions, mental health, lifestyle, psychological, and behavioral factors. This data-driven approach sheds light on cognitive aging patterns and informs policies for healthy aging. Our study pioneers non-parametric machine learning methods in this context.

**ARTICLE HISTORY**
Received 16 June 2023
Accepted 3 April 2024

**KEYWORDS**
Oldest-old; MMSE; longitudinal; K-Means; biopsychosocial framework

## Introduction

The total number of people with dementia in China is projected to reach 23.3 million by 2030 (Xu et al., 2017). Meanwhile, the Chinese oldest-old population is also increasing. The impact of cognitive decline on this population's health is profound and multifaceted, as it often leads to a decline in the ability to perform daily activities (Cloutier et al., 2021), increases the risk of accidents and injuries (Smith et al., 2021), and is strongly associated with other health and social problems (McWhirter et al., 2020) that complicate medical management, intervention, and caregiving plans (Anstey et al., 2020; Jia, Quan, et al., 2020; Ren et al., 2022). The challenges are a pressing concern as the anticipated annual costs of older adults with cognitive decline in China are estimated to reach more than $114 billion by 2030, imposing substantial burdens on families, communities, and society (Xu et al., 2017).

Although cognitive impairment mainly affects older adults, it is neither a normal part of the aging process nor inevitable, as many risk factors are modifiable (Livingston et al., 2020). For example, Veríssimo et al. (2022) revealed that some older adults maintain adequate cognitive functioning at

**CONTACT** Juan Xi jx@uakron.edu Department of Sociology, University of Akron, 302 Buchtel Common, Akron, OH 44325-1905, USA

the most advanced age. Understanding the protective and risk factors for the cognitive functioning of Chinese older adults is important. Most research on old-age cognitive function used samples from Western countries, and little was known about those in non-Western countries where life conditions are very different (Mukadam et al., 2019). Recently, more and more research has been paid to cognitive disorders in Chinese older adults (e.g., Zhou et al., 2021; C.-E. Zhu et al., 2022). However, most studies, with only a few exceptions (Jia, Du, et al., 2020; Mukadam et al., 2019), have mainly focused on antecedents of one or two life domains (Mao et al., 2020; Yin et al., 2017; C.-E. Zhu et al., 2022).

Guided by the biopsychosocial framework (Engel, 1977), in this study, we focus on the Chinese oldest-old population, aiming to identify different types of cognitive function trajectories and identify protective and risk factors across multiple life domains by using the Chinese Longitudinal Healthy Longevity Survey (CLHLS) data (Zeng et al., 2002).

## *Theoretical framework*

This study is guided by the biopsychosocial framework for health (Engel, 1977). Rather than narrowly focusing on pathophysiology or other biological factors in understanding health and disease, the biopsychosocial model considers any diseases or health conditions a product of processes across multiple domains of human life and can only be understood in its complex context of biological, personal, and social-cultural environments (Engel, 1977; Marmot & Wilkinson, 2005). When applied to cognitive functioning in old age, the biopsychosocial approach emphasizes that multiple factors across an individual's different life domains must be considered simultaneously. These can include an older adult's physical health conditions, biological history, psychological and behavioral characteristics, socioeconomic status during the life course, social activities and connections, and cultural and environmental context (Bolton & Gillett, 2019; McInnis-Dittrich, 2009).

The development in the literature supports the broad view of the biopsychosocial framework on cognitive functioning. For example, after reviewing a large number of studies, the 2020 Lancet Commission on Dementia highlighted 12 major risk factors across biological, psychological, and social domains of life: hypertension, diabetes, traumatic brain injury, obesity, hearing impairment, depression, smoking, excessive alcohol consumption, physical inactivity, less education, low social contact, and air pollution (Livingston et al., 2020).

As an illustration within the biopsychosocial framework, research on Chinese older adults underscores a range of biological risk or protective factors such as physical health (Tu et al., 2020; Zhao et al., 2021), physical function (Li & Li, 2022; Zhang et al., 2019), chronic diseases, advanced age, parental history of dementia, and prior medical issues (Jia, Du, et al., 2020). Psychological (Zhang et al., 2019) and lifestyle factors (Gao et al., 2017), such as depressive symptoms (Zhou et al., 2021; Zhu et al., 2022) and smoking and alcohol consumption (Li & Li, 2022), are significant contributors. Moreover, social factors like childhood adversity (Ma et al., 2021), being female, residing in rural areas, having a lower socioeconomic status, and being single, play a role in shaping cognitive function in older adults in China (Jia, Du, et al., 2020; J. Wang et al., 2020; Zhang et al., 2019).

Many risk and protective factors influencing cognitive decline are primarily observed in middle-aged and older adults (Q. Wang & Kang, 2019), while limited long-term data for the oldest-old (80+) presents challenges for comprehensive research (K. X. Ye et al., 2023). For example (Yang et al., 2021), studied the impact of residence place using cross-sectional data, which does not capture cognitive changes. Other studies, employing either shorter-term or long-term follow-up data, examined various factors like education (Gao et al., 2017), lifestyle (Li & Li, 2022), early cognitive reserve (Chen & Lu, 2020), and alcohol use (Han & Jia, 2021) but lacked a systematic approach considering a broad range of risk factors as suggested by the biopsychosocial model (Engel, 1977).

In addition, previous studies either used models assuming a single distribution of change trajectories (e.g., Hu et al., 2021), possibly missing out on unobserved multidimensional natural groupings, or relied on parametric statistical approaches (e.g., Yu et al., 2020) demanding sufficient prior knowledge and computational resources. The current study employs K-Means analysis, a model that offers flexibility in identifying unobserved change clusters without preconceived functional form assumptions, presenting a new perspective on data patterns and unseen groups (Verboon & Pat-El, 2022). By comprehensively considering biological, psychological, and social factors, we aim to identify the protective/risk factors associated with cognitive trajectories among the Chinese oldest-old population across social, psychological, and behavioral domains.

The present study aims to address the following research questions:

(1) Are there multiple qualitatively different cognitive functioning trajectories among the Chinese oldest-old population?
(2) If so, how many trajectory patterns exist, and how do they differ?
(3) Among the broad range of protective/risk factors suggested by the biopsychosocial framework and the existing literature, which factors are important in predicting the patterns of cognitive functioning trajectory a Chinese is likely to follow at their oldest age?

## Methods

### Data

We used the Chinese Longitudinal Healthy Longevity Survey (CLHLS) data (Zeng et al., 2002) to answer our research questions. The CLHLS provides a large random sample, following up on 9,093 Chinese oldest-old (aged 77 and above) population from 1998 to 2018. The panel was randomly selected in 1998 from about half of the counties and cities in 23 of China's 31 provinces. Interviews were conducted in person. If a respondent died between the two interviews, the next closest kin of the deceased respondent was interviewed as the proxy respondent, and there were no more follow-ups after that. Over time, the number of follow-up interviews rapidly decreased due to the high mortality rate for this age group (Lv et al., 2019). Only ten respondents survived at the time when the 2018 survey was conducted. Detailed information on the survey design has been published, and the data quality has been verified (e.g., Zeng et al., 2002). The CLHLS is currently the most widely used panel data on the health of the Chinese oldest-old population, with the largest sample size and most extended follow-up.

Although CLHLS added new respondents at each survey year, this study focuses on the original panel of the Chinese oldest-old population selected in 1998. We excluded the respondents who did not provide information on cognitive function in baseline and follow-up surveys, leading the final sample size to 8,903. The CLHLS obtained ethics approval from Duke University and Peking University (IRB00001052–13074). Written informed consent was obtained from all participants.

Sample descriptive statistics measured at baseline (1998) were reported in Table 1. At the baseline survey (1998), our sample had a mean age of 92 (sd = 7.74) with an age range spanning 77 to 122. About 60% of the sample were female, and 38% were urban residents. Ninety-three percent of the sample reported Han ethnicity. Only 16% remained married, and 80% were widowed. About 10% of the sample lived alone, and 5% lived in institutions. The majority of the respondents lived in a household (85%). Among them, 14% lived with spouses, 62% lived with children, and 23% lived with relatives or other people. Two-thirds of the sample had no formal schooling. Less than 3% of the sample had 12 years or more of education. The average years of schooling was 1.79. Over half of the sample (56%) reported that they had experienced hunger in childhood.

### Measurement

*Cognitive function* was evaluated by the Chinese version of the Mini-Mental State Examination (MMSE). The MMSE is a classic cognitive function assessment tool with established high validity and reliability (Tan & Feng, 2022).

**Table 1.** Sample descriptive statistics at baseline survey and group profile comparison.

| | All ($n = 8{,}903$) | | | | | High-Stable Age-Trajectory Group ($n = 6.054$) | | Low-Stable Age-Trajectory Group ($n = 2{,}849$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Min | Max | Missing | Mean | S.D. | Mean | S.D. | Diff. | |
| **Biological Factors** | | | | | | | | | | | |
| Age at 1998 | 92.05 | 7.74 | 77 | 122 | 0 | 90.96 | 0.10 | 94.24 | 0.14 | −3.28 | *** |
| Mother's age or age at death | 68.12 | 14.66 | 18 | 113 | 0 | 67.97 | 0.28 | 68.74 | 0.46 | −0.77 | |
| Father's age or age at death | 64.06 | 12.43 | 20 | 114 | 0 | 64 | 0.24 | 64.38 | 0.44 | −0.38 | |
| Birth order | 2.42 | 1.66 | 1 | 14 | 238 | 2.46 | 0.02 | 2.34 | 0.03 | 0.12 | ** |
| Self-Rated Health | 2.61 | 0.84 | 0 | 4 | 591 | 2.69 | 0.01 | 2.40 | 0.02 | 0.29 | *** |
| Visual function | 2.46 | 0.87 | 0 | 3 | 131 | 2.65 | 0.01 | 2.08 | 0.02 | 0.57 | *** |
| Number of natuRal teeth | 5.15 | 7.40 | 0 | 32 | 66 | 5.64 | 0.10 | 4.16 | 0.13 | 1.48 | *** |
| Weight | 46.56 | 10.62 | 16 | 110 | 662 | 47.78 | 0.14 | 43.81 | 0.20 | 3.97 | *** |
| Activities of daily livings (ADL) | 7.50 | 2.68 | 6 | 18 | 35 | 6.88 | 0.03 | 8.74 | 0.06 | −1.86 | *** |
| Suffering from Parkinson's disease | 0.01 | 0.10 | 0 | 1 | 712 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | + |
| Suffering from ulcer | 0.04 | 0.18 | 0 | 1 | 657 | 0.04 | 0.00 | 0.03 | 0.00 | 0.01 | |
| Suffering from cancer | 0.01 | 0.07 | 0 | 1 | 741 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | |
| Suffering from glaucoma | 0.03 | 0.16 | 0 | 1 | 703 | 0.02 | 0.00 | 0.03 | 0.00 | −0.01 | ** |
| Suffering from cataract | 0.20 | 0.40 | 0 | 1 | 656 | 0.18 | 0.01 | 0.24 | 0.01 | −0.06 | *** |
| Suffering from tb | 0.01 | 0.10 | 0 | 1 | 689 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | |
| Suffering from asthma | 0.13 | 0.34 | 0 | 1 | 607 | 0.14 | 0.00 | 0.13 | 0.01 | 0.01 | |
| Suffering from stroke or CVD | 0.03 | 0.18 | 0 | 1 | 657 | 0.03 | 0.00 | 0.05 | 0.00 | −0.02 | *** |
| Suffering from heart diseases | 0.08 | 0.27 | 0 | 1 | 690 | 0.08 | 0.00 | 0.07 | 0.00 | 0.01 | |
| Suffering from diabetes | 0.01 | 0.09 | 0 | 1 | 724 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | + |
| Suffering from hypertension | 0.14 | 0.35 | 0 | 1 | 693 | 0.15 | 0.00 | 0.12 | 0.01 | 0.03 | *** |
| Suffering from otherdisease | 0.14 | 0.34 | 0 | 1 | 938 | 0.12 | 0.00 | 0.18 | 0.01 | −0.06 | *** |
| Comobidity | 0.61 | 0.91 | 0 | 11 | 0 | 0.59 | 0.01 | 0.66 | 0.02 | −0.07 | ** |
| **Psychological and Lifestyle Factors** | | | | | | | | | | | |
| Self-Rated Quality of Life | 2.89 | 0.73 | 0 | 4 | 596 | 2.93 | 0.01 | 2.79 | 0.01 | 0.14 | *** |
| Looking at bright side of things | 2.92 | 0.81 | 0 | 4 | 969 | 2.98 | 0.01 | 2.75 | 0.02 | 0.23 | *** |
| Keep things neat and clean | 3.03 | 0.74 | 0 | 4 | 870 | 3.08 | 0.01 | 2.92 | 0.02 | 0.16 | *** |
| Feel fearful (reverse coded) | 2.65 | 0.86 | 0 | 4 | 935 | 2.69 | 0.01 | 2.54 | 0.02 | 0.15 | *** |
| Feel lonely (reverse coded) | 2.56 | 0.89 | 0 | 4 | 965 | 2.61 | 0.01 | 2.43 | 0.02 | 0.18 | *** |
| Make own decisions | 2.53 | 1.03 | 0 | 4 | 1,058 | 2.6 | 0.01 | 2.36 | 0.02 | 0.24 | *** |
| Feel useless as aging (reverse coded) | 1.95 | 1.00 | 0 | 4 | 1,015 | 2.01 | 0.01 | 1.78 | 0.02 | 0.23 | *** |
| As happy as when younger | 2.31 | 1.06 | 0 | 4 | 1,079 | 2.39 | 0.01 | 2.11 | 0.02 | 0.28 | *** |
| Rice | 0.75 | 0.43 | 0 | 1 | 6 | 0.76 | 0.01 | 0.74 | 0.01 | 0.02 | + |
| Wheat | 0.19 | 0.39 | 0 | 1 | 6 | 0.19 | 0.01 | 0.20 | 0.01 | −0.01 | |
| Frequency eating Fruit/Vegietable | 3.50 | 1.46 | 0 | 6 | 9 | 3.64 | 0.02 | 3.21 | 0.03 | 0.43 | *** |
| Frequency eating meat | 1.11 | 0.70 | 0 | 2 | 61 | 1.14 | 0.01 | 1.05 | 0.01 | 0.09 | *** |
| Frequency eating fish | 0.81 | 0.64 | 0 | 2 | 93 | 0.84 | 0.01 | 0.75 | 0.01 | 0.09 | *** |
| Frequency eating eggs | 1.10 | 0.70 | 0 | 2 | 61 | 1.11 | 0.01 | 1.07 | 0.01 | 0.04 | ** |
| Frequency eating bean/bean products | 1.05 | 0.65 | 0 | 2 | 41 | 1.08 | 0.01 | 0.98 | 0.01 | 0.10 | *** |
| Frequency drinking tea | 0.70 | 0.86 | 0 | 2 | 382 | 0.76 | 0.01 | 0.57 | 0.02 | 0.19 | *** |
| Frequency eating candy/sugar | 0.95 | 0.76 | 0 | 2 | 76 | 0.95 | 0.01 | 0.95 | 0.01 | 0.00 | |
| Ever smoked | 0.32 | 0.47 | 0 | 1 | 4 | 0.36 | 0.01 | 0.25 | 0.01 | 0.11 | *** |
| Ever drunk alcohol | 0.35 | 0.48 | 0 | 1 | 7 | 0.37 | 0.01 | 0.31 | 0.01 | 0.06 | *** |
| Physical exercise or not | 0.27 | 0.45 | 0 | 1 | 8 | 0.33 | 0.01 | 0.16 | 0.01 | 0.17 | *** |
| Frequency doing household chores | 0.67 | 0.84 | 0 | 2 | 8 | 0.79 | 0.01 | 0.45 | 0.01 | 0.34 | *** |
| Frequency reading newspaper | 0.27 | 0.63 | 0 | 2 | 14 | 0.35 | 0.01 | 0.12 | 0.01 | 0.23 | *** |
| Frequency petting pets | 0.25 | 0.59 | 0 | 2 | 15 | 0.28 | 0.01 | 0.19 | 0.01 | 0.09 | *** |
| Frequency playing majiang | 0.16 | 0.46 | 0 | 2 | 12 | 0.21 | 0.01 | 0.07 | 0.01 | 0.14 | *** |
| Frequency watching TV | 0.77 | 0.83 | 0 | 2 | 9 | 0.91 | 0.01 | 0.49 | 0.01 | 0.42 | *** |
| Frequency practicing religion | 0.20 | 0.49 | 0 | 2 | 8 | 0.23 | 0.01 | 0.16 | 0.01 | 0.07 | *** |
| **Social Factors** | | | | | | | | | | | |
| Urban | 0.38 | 0.49 | 0 | 1 | 0 | 0.41 | 0.01 | 0.32 | 0.01 | 0.09 | *** |
| Female | 0.60 | 0.49 | 0 | 1 | 0 | 0.54 | 0.01 | 0.72 | 0.01 | −0.18 | *** |
| Education | 1.79 | 3.58 | 0 | 26 | 43 | 2.16 | 0.05 | 1.05 | 0.05 | 1.11 | *** |

(*Continued*)

**Table 1.** (Continued).

| | All (*n* = 8,903) | | | | | High-Stable Age-Trajectory Group (*n* = 6.054) | | Low-Stable Age-Trajectory Group (*n* = 2,849) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Min | Max | Missing | Mean | S.D. | Mean | S.D. | Diff. | |
| Han Ethnicity | 0.93 | 0.26 | 0 | 1 | 16 | 0.93 | 0.00 | 0.93 | 0.00 | 0.00 | |
| Occupation before age 60 | | | | | 0 | | | | | | |
| Governmental/Managerial | 0.03 | 0.17 | 0 | 1 | | 0.04 | 0.00 | 0.01 | 0.00 | 0.03 | *** |
| Professional/Technical | 0.04 | 0.21 | 0 | 1 | | 0.06 | 0.00 | 0.02 | 0.00 | 0.04 | *** |
| Farmer | 0.55 | 0.50 | 0 | 1 | | 0.53 | 0.01 | 0.59 | 0.01 | −0.06 | *** |
| Manufactural/commercial/Service worker | 0.16 | 0.37 | 0 | 1 | | 0.18 | 0.01 | 0.12 | 0.01 | 0.06 | *** |
| Housemaker | 0.19 | 0.39 | 0 | 1 | | 0.17 | 0.00 | 0.24 | 0.01 | −0.07 | *** |
| Other | 0.02 | 0.14 | 0 | 1 | | 0.02 | 0.14 | 0.02 | 0.14 | 0.00 | |
| Main Financial Support | | | | | 2 | | | | | | |
| Pension | 0.16 | 0.37 | 0 | 1 | | 0.19 | 0.01 | 0.09 | 0.01 | 0.10 | *** |
| Family support | 0.66 | 0.47 | 0 | 1 | | 0.63 | 0.01 | 0.72 | 0.01 | −0.09 | *** |
| Other | 0.18 | 0.38 | 0 | 1 | | | | | | 0.00 | |
| Adequate medical care | 0.97 | 0.18 | 0 | 1 | 16 | 0.97 | 0.00 | 0.96 | 0.00 | 0.01 | |
| Adequate medical care as a child | 0.83 | 0.37 | 0 | 1 | 60 | 0.84 | 0.00 | 0.83 | 0.01 | 0.01 | |
| Childhood hunger | 0.56 | 0.50 | 0 | 1 | 71 | 0.54 | 0.01 | 0.60 | 0.01 | −0.06 | *** |
| Current Marrital Status | | | | | 4 | | | | | | |
| Married | 0.16 | 0.37 | 0 | 1 | | 0.2 | 0.01 | 0.10 | 0.01 | 0.10 | *** |
| Widowed | 0.80 | 0.40 | 0 | 1 | | 0.77 | 0.01 | 0.87 | 0.01 | −0.10 | *** |
| Other | 0.04 | 0.20 | 0 | 1 | | 0.03 | 0.00 | 0.03 | 0.00 | 0.00 | |
| # marriage life time | 1.16 | 0.47 | 0 | 5 | 9 | 1.18 | 0.01 | 1.13 | 0.01 | 0.05 | *** |
| Age at first marriage | 20.45 | 5.19 | 10 | 87 | 585 | 20.72 | 0.07 | 19.88 | 0.09 | 0.84 | *** |
| Number of children | 4.70 | 2.81 | 0 | 17 | 116 | 4.72 | 0.04 | 4.65 | 0.05 | 0.07 | |
| Living Arrangement | | | | | 0 | | | | | | |
| Alone | 0.10 | 0.30 | 0 | 1 | | 0.11 | 0.00 | 0.08 | 0.01 | 0.03 | *** |
| In a household | 0.85 | 0.36 | 0 | 1 | | 0.84 | 0.00 | 0.87 | 0.01 | −0.03 | *** |
| In institutions | 0.05 | 0.22 | 0 | 1 | | 0.05 | 0.00 | 0.05 | 0.00 | 0.00 | |
| People live with | | | | | 0 | | | | | | |
| live with spouse | 0.14 | 0.35 | 0 | 1 | | 0.17 | 0.00 | 0.08 | 0.01 | 0.09 | *** |
| live with children | 0.62 | 0.48 | 0 | 1 | | 0.60 | 0.01 | 0.67 | 0.01 | −0.07 | *** |
| live with Other people | 0.23 | 0.42 | 0 | 1 | | 0.23 | 0.42 | 0.25 | 0.43 | −0.02 | |
| Number of people live together | 3.15 | 2.11 | 0 | 34 | 15 | 3.09 | 0.03 | 3.28 | 0.05 | −0.19 | ** |

*$p < .05$; **$p < .01$; ***$p < .001$.

The MMSE includes the following domains of cognitive function: orientation, registration, attention, calculation, recall, and language. The answers to most of the questions were recorded as correct/wrong/not able to answer. We coded "correct" as 1 and combined "wrong" and "not able to answer" into 0. One question asks the respondents to list as many kinds of edible food as possible. The answer was recorded as the number of food items listed. We capped the answer at 7. The summary MMSE scores ranged from 0 to 30, with higher scores indicating better cognitive functioning.

In this study, we included 73 possible cognitive functioning predictors mentioned in the literature and available in the dataset. All of them were measured in the 1998 baseline survey. All the categorical variables were dummy-coded.

### *For biological factors, we included measures for physical health conditions and family history*

These included age at baseline survey (in years), self-rated health (0 = poor to 4 = excellent); activities of daily living (ADL; summary score from the six items measuring capacity to bath, dress oneself, use the toilet, move around, feed oneself, and continence); visual function (0 = blind, 1 = cannot see the picture presented, 2 = can see the picture, but cannot distinguish the break in the circle; 3 = can see and distinguish the break in the circle); weight (in kilograms); the number of natural teeth; suffering from each of the following diseases/conditions (yes = 1 for each condition): Parkinson's disease; ulcer; cancer; glaucoma; cataract; TB; asthma; stroke; heart diseases; diabetes; hypertension; and other diseases. We also added up the number of conditions as a measure of comorbidity. Family history was measured by the mother's current age or age at death, the father's current age or age at death, and the respondent's birth order. Most variables have a low level of missing values. There are substantial amounts of missing values for mothers' and fathers' ages. We used mean imputation for these two variables' missing values and listwise deletion for the rest.

### *For psychological and lifestyle factors, we considered mental health, perceived quality of life, and lifestyle*

The survey used the following items to measure mental health: look on the bright side of things; keep things neat and clean; make their own decisions; feel useless with age; as happy as a younger age; feel fearful or anxious; feel lonely and isolated. Responses to each item were coded as 0 = never, 1 = seldom, 2 = sometimes, 3 = often, and 4 = always. The summary scale of these items had low reliability (Cronbach's alpha = .6). We decided to use individual items in the analysis. The survey also measured self-rated quality of life (0 = very bad to 4 = very good). Diet and lifestyle variables included: staple food rice (yes = 1), wheat (yes = 1), other(yes = 1); the frequency of eating fruits or vegetables (rarely or never = 0 to 6 = both almost every day); the frequency of consuming each of the following food items (rarely or never = 0 to 3 = almost every day for each food item): meat; fish; eggs; bean/bean products; tea; and sugary food or candy; ever smoked (yes = 1); ever drunk alcohol (yes = 1); exercise (yes = 1); and the frequency doing each of the following activities (never = 0 to 2 = almost every day for each activity): doing chores; reading newspapers; petting pets; playing Majiang; watching TV; and practicing religion.

*Social factors* included *socioeconomic variables* such as education (years of schooling); the main occupation before 60 (governmental/institutional/managerial positions, farmer, professional/technical worker, manufacturer/commercial/service worker, housemaker, other = ref.); the main source of financial support in 1998 (pension, family support, other = ref.); adequate medical service in 1998 (yes = 1); adequate medical service in childhood (yes = 1);

often went to bed hungry in childhood (yes = 1). We also considered living arrangements and marital status because they indicate an older adult's immediate social environment. These included types of living arrangements (in a household, living alone, in an institution, other = ref.); the number of people living together (count); and people living with (spouse, child, other = ref.); marital status (married, widowed, and other = ref.); the number of marriages in a lifetime (count); age at first marriage (in years); and the number of Children in a lifetime (count). Social demographic variables included sex (female = 1), ethnicity (Han = 1, else = 0), and urban/rural residence (urban = 1, rural = 0). These demographic factors function as social stratification mechanisms in Chinese society, shaping an individual's access to resources (Bian, 2002).

### *Analytic plan*

First, we used the K-means algorithm to identify potential latent groups with different cognitive function trajectories using the longitudinal repeated measures of cognitive function in the 1998 baseline survey and the 2000, 2002, 2005, 2008, and 2011 follow-up surveys. The last two waves of data (2014 and 2018) were not used in this analysis because of the high sparsity.

K-means is one of the latest techniques demonstrated as a preferred method for trajectory clustering analysis (Den Teuling et al., 2023; Verboon & Pat-El, 2022). For example, Verboon and Pat-El (2002) found that the non-parametric K-means algorithm consistently performed well in recovering known clustering structures within longitudinal trajectories. This approach differs from previous studies comparing cognitive functions across different observed statuses, such as sex and race/ethnicity, with a fixed number of groups (e.g., Hu et al., 2021). Similarly, Den Teuling et al. (2023) highlighted K-means' practical equivalence to group-based trajectory modeling and its computational efficiency, making it an effective choice for large datasets. Both studies affirm that K-means is an accurate and reliable approach for identifying distinct clusters based on longitudinal trajectories, making it a suitable method for trajectory clustering analysis. By considering unobserved patterns in the data beyond demographic or socioeconomic differences, we aim to capture more complex underlying structures in the population that are not directly observable.

The K-means clustering analysis was run in RStudio (2022.12.0). To depict respondents' trajectory of cognitive function in old age, we used two sets of time variables: age and survey wave. We first used survey waves as time variables to derive trajectories. The survey wave trajectories would depict changes in each respondent's MMSE during the study time from baseline to their last follow-up observation. We then repeated the same analysis using age as the time variable for a more intuitive presentation of cognitive function changes in old age. The age trajectories describe MMSE changes for each respondent from their baseline age to when they were last observed. We also checked cohort effects in sensitivity analysis, but the results did not show a significant difference (results available upon request). Due to the exploratory nature of the analysis, we used the results from both approaches as cross-validation.

Next, we compared the profiles of the detected groups on a wide range of biological, psychological/behavioral, and social characteristics. Finally, we used the Lasso regression analysis (STATA 17) to find a set of important predictors that can be used to predict latent group membership. All predictors included in the Lasso regression analysis were measured in 1998. We chose to use the Lasso regression analysis instead of the regular logistic regression due to the large number of potential predictors and their correlations, which can lead to overfitting and multicollinearity issues (Hastie et al., 2009). The predictors selected by the Lasso regression model can be used across different samples to predict cognitive functioning in old age, aiding in developing suitable programs for older adults in fields like social work.

**Table 2.** Sample descriptive statistics of cognitive function and group comparisons.

| | All (*n* = 8,903) | | | | | | High-Stable Age-Trajectory Group (*n* = 6,054) | | | Low-Stable Age-Trajectory Group (*n* = 2,849) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std. | Min | Max | n | Missing | Mean | S.D. | n | Mean | S.D. | n | Diff. | |
| MMSE_98 | 21.11 | 8.93 | 0 | 30 | 8,694 | 209 | 24.92 | 0.07 | 5,795 | 13.51 | 0.19 | 2,899 | 11.41 | *** |
| MMSE_00 | 20.36 | 9.55 | 0 | 30 | 4,662 | 4,241 | 24.05 | 0.11 | 3,352 | 10.92 | 0.28 | 1,310 | 13.13 | *** |
| MMSE_02 | 22.04 | 7.70 | 0 | 30 | 1,966 | 6,937 | 24.06 | 0.15 | 1,603 | 13.13 | 0.44 | 363 | 10.93 | *** |
| MMSE_05 | 21.88 | 7.88 | 0 | 30 | 748 | 8,155 | 23.23 | 0.26 | 640 | 13.87 | 0.93 | 108 | 9.36 | *** |
| MMSE_08 | 20.36 | 8.60 | 0 | 30 | 241 | 8,662 | 21.74 | 0.52 | 209 | 11.34 | 1.67 | 32 | 10.40 | *** |
| MMSE_11 | 20.97 | 7.55 | 0 | 30 | 88 | 8,815 | 21.85 | 0.77 | 78 | 14.10 | 3.15 | 10 | 7.75 | *** |
| MMSE_14 | 23.23 | 7.03 | 0 | 30 | 30 | 8,873 | 23.23 | 7.03 | 30 | – | – | | – | |
| MMSE_18 | 20.43 | 11.97 | 1 | 30 | 7 | 8,896 | 20.43 | 11.97 | 7 | – | – | | – | |

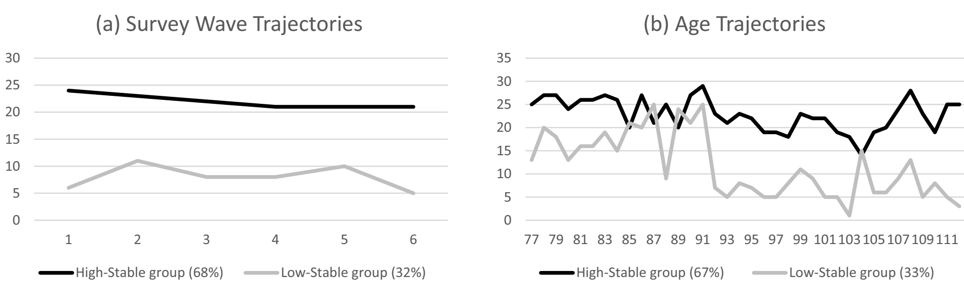


**Figure 1.** Cognitive function trajectory groups identified by K-Means Algorithm.

**Table 3.** Parameters of K-Mean results for both survey wave trajectories and age trajectories.

| | Survey wave trajctories | | Age trajectories | |
|---|---|---|---|---|
| | 2-groups | 3-groups | 2-groups | 3-groups |
| Caliniski | 5531 | 1743 | 16023 | 4236 |
| Group size | 68% – 32% | 44% – 31% – 25% | 67% – 33% | 49% – 28% – 23% |

## Results

The left panel of Table 2 reported the sample statistics of MMSE during the study period. The sample mean MMSE was 21.11 in the baseline survey, which fluctuated around this value during the study period. The sample size shrank rapidly during the study due to the high old-age mortality rates. The mean MMSE in the later waves reflected the positive selection effect.

### *K-Means clustering analysis*

The results of the K-means analysis are reported in Figure 1. Among the multiple results generated by the K-means algorithm, we decided on the two-group model for several reasons (Table 3). We compared the Calinski-Harabasz (CH) index (Liu et al., 2010). The two-group model in both analyzes (i.e., survey wave trajectories and age trajectories) had the highest CH index compared to other models (Table 3). The size of each cluster was also adequate. Moreover, the characteristics of the two clusters were interpretable and theoretically meaningful.

Figure 1(a) shows the mean MMSE trajectory of the two groups identified by the K-means algorithm when the survey wave is the time axis. The groups were labeled High-Stable (HS) and Low-Stable (LS). The HS group contains those with relatively high cognitive scores from 1998 to 2011. Although there was some slow and slight decline in the trajectory of the HS group, it remained high and stable overall. About 68% of the respondents were in the HS category.

On the other hand, the LS group was those who had relatively low MMSE scores over time. Although their scores fluctuated between 1998 and 2011, they remained relatively stable at a low level. About 32% of the respondents were in the LS group.

Figure 1(b) shows the mean MMSE trajectory of the two groups identified by the K-means algorithm when participants' age is the time axis. Similar to Figure 1(a), there was an HS group (67% of the sample) and an LS group (33% of the sample) with similar trends.

Comparing the two sets of K-means clustering analyzes, one using age as the time variable and the other using survey waves, we found the results quite similar. They both arrived at a two-group solution with similar patterns (high-stable and low-stable trajectories). The age trajectories were less smooth than the wave trajectories because of the finer unit of the time variable.

Comparing the group membership classification, we found that 84% of the participants were in the same group in both analyzes ($r = 0.62$). Because of the high level of agreement in the results of both analyzes and because the age trajectories are more intuitive and more frequently reported in the literature (Zhang et al., 2019), in the next portion of the analysis, we focused on comparing group profiles for HS vs. LS age-trajectory groups and identifying important protective/risk factors for being in the high-stable age-trajectory group.

### *Group-profile comparison*

The right panels of Table 2 reported the sample mean MMSE for the two age-trajectory clusters over the survey waves. The HS group started high, and there was a gradual decline in the follow-up surveys. However, the means were consistently above 20, a critical point in the MMSE scale with clinical significance (Yin et al., 2017; Zhou et al., 2021). On the other hand, the LS group was 10 points lower in the baseline survey and fluctuated around that level. Their mean MMSE scores were consistently well below 20.

Comparing the biological, psychological, lifestyle, and social characteristics of the two groups in the 1998 baseline survey (Table 1), starting with biological factors, the HS group was about three years younger. They reported better self-rated health. They had better visual function, more natural teeth, higher body weight, and fewer limitations in daily activities. They were less likely to suffer from glaucoma, cataracts, stroke or CVD, and other diseases. Although they were more likely to have hypertension, their number of diseases at the baseline survey was lower than in the LS group.

Concerning psychological and lifestyle factors, the HS group reported a higher quality of life in 1998. They were more likely to look at the bright side of things, to keep things neat and clean, to make their own decisions, or to feel happy. They were less likely to feel lonely, fearful, or useless as they aged. They more frequently ate fruits, vegetables, meat, fish, eggs, and beans or bean products. They drank tea more frequently and were more likely to have smoked or used alcohol. They were more likely to exercise. They did chores, read newspapers, petted pets, played Majiang, watched TV, and practiced religion more often than the LS group.

As to the social characteristics, demographically, the HS group consisted of fewer women, was less likely to live in a rural area, and had more education. Compared to the LS group, the HS group was more likely to be in occupations with more resources in Chinese society before 60, such as governmental/managerial/institutional jobs, professional/technical occupations, and manufacturer/commercial/service jobs (Bian, 2002). They were less likely to be farmers or housemakers, positions with less prestige and resources in

Chinese culture (Bian, 2002). They were more likely to be supported by a pension and less likely to depend on their family. They were also less likely to have experienced childhood hunger. The comparison of social and economic indicators implied that the HS group was in an advantaged social position relative to the LS group.

Regarding living arrangements and marital status, the HS group was more likely to remain married, less likely to be widowed, had more marriages in their lifetime, and were older when they first married. They were more likely to live alone or with their spouse but less likely to live with their children. They were less likely to live with many people in the same place. In sum, compared to the LS group, the HS group had higher social and economic status at baseline and earlier in life. They had healthier diets and lifestyles, better quality of life and mental health, and better physical conditions.

**Table 4.** Ranked predictors of cognitive function trajectory groups.

| Results from Lasso Regression Analysis with Adaptive Cross-Validation Methods | | | |
|---|---|---|---|
| Predictors measured at Baseline Survey | Penalized Coef. | Standardized Coef. | OR |
| **High-Stable Age-Trajectory Group = 1** | | | |
| Activities of daily livings (ADL) | −0.14 | −0.25 | 0.87 |
| Visual function | 0.28 | 0.21 | 1.33 |
| Age at 1998 | 0.02 | 0.15 | 1.02 |
| Frequency doing chores | 0.15 | 0.13 | 1.16 |
| Female | −0.22 | −0.11 | 0.80 |
| Frequency watching TV | 0.12 | 0.10 | 1.13 |
| Frequency eating fruit/vegietable | 0.07 | 0.10 | 1.07 |
| Governmental/Managerial occupation | 0.51 | 0.09 | 1.67 |
| Weight | 0.01 | 0.09 | 1.01 |
| Education | 0.02 | 0.09 | 1.02 |
| Looking at bright side of things | 0.10 | 0.08 | 1.11 |
| Suffering from diabetes | −0.74 | −0.07 | 0.48 |
| Main financial support from family | −0.14 | −0.07 | 0.87 |
| Frequency eating bean/bean products | 0.10 | 0.06 | 1.10 |
| Frequency playing Majiang | 0.13 | 0.06 | 1.14 |
| Frequency drinking tea | 0.06 | 0.06 | 1.07 |
| Physical exercise or not | 0.11 | 0.05 | 1.11 |
| Frequency eating eggs | −0.07 | −0.05 | 0.93 |
| Frequency eating fish | 0.07 | 0.04 | 1.07 |
| Adequate medical care | 0.28 | 0.04 | 1.32 |
| Widowed | −0.09 | −0.04 | 0.91 |
| Frequency practicing religion | 0.07 | 0.04 | 1.08 |
| Feel lonely | −0.04 | −0.03 | 0.96 |
| Self-rated quality of life | 0.04 | 0.03 | 1.05 |
| Number of children | 0.01 | 0.03 | 1.01 |
| Urban | 0.06 | 0.03 | 1.06 |
| Childhood hunger | −0.06 | −0.03 | 0.94 |
| Feel fearful | −0.03 | −0.03 | 0.97 |
| Ever smoked | 0.04 | 0.02 | 1.04 |
| Self-rated health | 0.04 | 0.02 | 1.05 |
| Feel useless | −0.02 | −0.02 | 0.98 |
| Mother's age or age at death | 0.00 | −0.01 | 1.00 |
| Frequency eating meat | 0.01 | 0.01 | 1.01 |
| Eating rice | 0.02 | 0.01 | 1.02 |
| *Lamda* | *0.01* | | |
| *Deviance Ratio* | *0.08* | | |

### *Lasso regression analysis for protective/risk factors*

To answer the third research question about protective factors for being in the high-stable cognitive function trajectory group in old age or the risk factors for being in the adverse trajectory group, we conducted a Lasso regression analysis with adaptive cross-validation (Table 4). The selected predictors were ranked according to their standardized coefficients to indicate their relative importance.

The presence of more limitations in daily activity at baseline, among all biopsychosocial factors considered in the study, was the most potent risk factor for falling in the adverse trajectory group in subsequent years. On the other hand, better visual function protected the Chinese oldest-old population and helped them maintain cognitive functioning. Although, on average, the HS group was younger, after controlling for other predictors, older ages at baseline increased the odds of being in the HS group. Other important biological protective factors include better self-rated health and higher body weight at baseline. On the other hand, diabetes increases the risk of falling in the adverse trajectory group.

Protective factors in the psychological domain include a higher self-rated quality of life and a positive life orientation at baseline. The frequencies of feeling lonely, fearful, and useless at baseline were important risk factors for an adverse trajectory of cognitive function in the oldest years.

Protective lifestyle variables include doing chores, watching TV, and eating fruits and vegetables. These variables were highly ranked as protective factors. Frequently drinking tea and eating beans or bean products, fish, meat, and rice were also beneficial. Physical exercise and frequently practicing religion or playing Majiang, a traditional Chinese game, were also protective. Surprisingly, smoking was found to be a beneficial factor. Another surprising finding was the consumption of eggs. The HS group ate more eggs at baseline, but after controlling for other factors, the high frequency of eating eggs became a risk factor for an adverse outcome.

Protective social factors include holding governmental or managerial positions, having more education, living in urban areas, having adequate medical care, and having more children. On the other hand, being females, depending on family for financial support, widowhood, and childhood hunger were risk factors increasing the odds of being in the adverse trajectory of cognitive function in the oldest years.

## Discussion and conclusion

The current study is among the first to utilize a non-parametric machine learning-based clustering method (K-means) to uncover the latent clusters of cognitive function trajectories among the Chinese oldest-old population.

The study identified two latent clusters by analyzing the 13-year follow-up data from CLHLS. Furthermore, the study employed a Lasso regression analysis to identify a set of protective or risk factors, including biological conditions, psychological health, lifestyle, and sociodemographic background, that predict latent group membership, offering valuable insights for developing policies, programs, and services to promote active and healthy aging.

### *Two distinct cognitive subgroups with stable fluctuations among the Chinese oldest-old population*

The study revealed the presence of two distinct latent groups (i.e., HS and LS) in the cognitive functioning trajectories of the Chinese oldest-old population. The HS group shows stability at a healthy level, indicating that most older adults maintained satisfactory cognitive function even in their advanced age. Conversely, the LS group, comprising approximately one-third of the sample, exhibited consistently low cognitive function at baseline, and displayed a minimal decline afterward.

The result challenges previous views that cognitive decline is an inevitable consequence of aging. In contrast to the longstanding belief that biological age determines changes in cognitive function among older adults (Hu et al., 2021), our results demonstrate that cognitive decline in old age is not inevitable. Specifically, the HS group maintained cognitive function well into their 90s and 100s, while the LS group displayed a minimal decline in their MMSE scores with increasing age. In fact, after controlling for other risk factors, older age became a protective factor. These results support an alternative interpretation of age-related changes in cognitive function, suggesting that cognitive decline may primarily reflect abnormal pathological and death-related processes rather than normal age-related aging (Wilson et al., 2020). Recent brain science research highlights that structural changes in the aging brain may not directly lead to cognitive decline, as the aging brain can compensate for structural losses by modulating functional areas and utilizing unrelated brain regions (Sullivan & Pfefferbaum, 2006). The results of our study underscore the stability of cognitive function across different groups of older adults and highlight the significance of protective and risk factors, including social engagement, nutrition, physical fitness, and positive attitudes toward life, in shaping cognitive function among older individuals.

### *Exploring individual differences reveals heterogeneity in cognitive function within homogeneous life course cohorts*

The Chinese oldest-old population in our sample has experienced tremendous social-historical changes in their lifetime, from the two World Wars, the Civil War, the establishment of the People's Republic of China (PRC), and the dramatic transformation from a central-planned economy to a market economy (Gao et al., 2017). Built upon the central-planned economy, from the establishment of the PRC till the turn of the century, the social structure had allocated more financial, educational, and medical resources to urban residents, government officials, professional and technical occupations, manufacture workers for state-owned companies over farmers and some other occupations (Bian, 2002). The macro social forces were reflected in their gendered life experiences, childhood resource deprivation, educational levels, earlier occupations, places of residence, living arrangements, access to medical care, and financial resources, which, we found, influenced their cognitive functioning in old age.

Our study also revealed that, alongside these social-economic-historical forces, psychological and behavioral factors, including mental health, dietary structure, exercise, and participation in leisure activities, were also associated with the types of cognitive function trajectory that they follow. These identified risk/protective factors align with previous findings (Gao et al., 2017; Jia, Du, et al., 2020; Li & Li, 2022; Tu et al., 2020; Yu et al., 2020; C.-E. Zhu et al., 2022). While an optimistic life orientation helps maintain a healthy cognitive function at the oldest ages, emotional disorders such as fear and loneliness are risk factors for cognitive impairment (Aschwanden et al., 2020; Dos Santos et al., 2018). In addition, the latent HS group demonstrated higher consumption of fruits, vegetables, meat, and fish (K. X. Ye et al., 2023), and the LS group exhibited reduced exercise and engagement in fewer leisure activities (Gao et al., 2017).

It is worth noting that the LS group is less likely to smoke, and smoke was found to be a protective factor in our study. The finding contrasts with the 2020 Lancet Commission on Dementia (Livingston et al., 2020, p. 428), where smoking was emphasized as a major risk factor in later life. However, in the context of Chinese oldest-old adults, literature presents conflicting views on smoking, identifying it as both a protective (Y. C.-E. Zhu et al., 2022) and a risk factor (Jia, Du, et al., 2020; Li & Li, 2022). The LS group consumed less tobacco, which may also be due to their limited financial resources and living arrangements because they are more likely to live with adult children (M. Ye et al., 2017), and smoking is less acceptable among younger generations in China. More research is needed to clarify the role of smoking for this special cohort of Chinese adults.

Biological factors were very important, too. Among the more than 70 predictive factors included in the analysis, the limit of the ability to do daily activities (ADL) was the strongest risk factor for being in the adverse cognitive function trajectory group. Once older adults are limited with ADL, it is a sign of physical health impairment and a stressor leading to depression or anxiety and reduced participation in social activities. More research is needed to examine the relationship between physical health and cognitive impairment, especially the underlying mechanisms. Other health conditions, such as lower visual ability and diabetes, predict an adverse trajectory in old age (Zhao et al., 2021). In conclusion, the current study shows that older Chinese adults' cognitive impairment trajectory is significantly associated with many modifiable risk factors in psychological and behavioral/lifestyle domains.

### *A Machine-learning based analysis*

This study also contributes to the literature by its exploratory orientation and data-driven methodology. In a well-established research field, it seems paradoxical to conduct exploratory research. However, most deductive theory-driven research is only partially confirmatory (Berk, 2020). Many have used exploration in model adjustment and covariate selection. For example, to determine the functional form of the cognitive function trajectories in old age, most researchers would try different functional forms, such as linear, quadratic, and even cubic trajectories, and select the form that best fits the data (Gao et al., 2017). This partial exploratory approach has its limitations. It does not allow maximum flexibility in examining all possible functional forms and selecting the ones with the best fit. Our fully exploratory approach provided us with this flexibility, and we can move beyond a prespecified functional form and allow the K-Means algorithm to find the best form systematically. We used different algorithms to analyze the trajectories, one defined by age and the other by survey wave and got similar results. With a large representative sample, we have confidence in our results.

In traditional theory-driven research, key independent variables are often determined by theory, but the selection of covariates is not. The selection of covariates, especially when the literature suggests many covariates, is often an under-the-table activity (Berk, 2020). Even when there are explanations for which covariates to include in the analysis, there is usually no justification for covariates to be excluded. Our data-driven approach uses algorithms to consider all possible covariates and make selections explicitly and systematically. As machine learning become more and more mature, their systematic use can complement traditional theory-driven research (Williams, 2011).

### *Practical implications*

Because risk and protective factors exist in life's biological, psychological, and social domains, interventions should be multidimensional (McInnis-Dittrich, 2009; Zittel et al., 2002). This research suggests the need to test and implement multiple strategies simultaneously working on multiple life domains (Livingston et al., 2020) and many international multidisciplinary interventions for delaying dementia and disabilities that China has joined (Kivipelto et al., 2020). For example, in rural

China, a multidomain intervention group received guidelines for a healthy lifestyle, dietary advice, physical activity promotion, personalized leisure activities, cognitive training, and intensive management of major vascular risk factors (Kivipelto et al., 2020).

Furthermore, multidisciplinary and integrated interventions should be promoted to more at-risk populations (i.e., LS), emphasizing the need for early prevention and detection. The identification of LS and their significant predictors (e.g., females, less educated individuals, rural dwellers, widows, and those without pension accounts) is an important new contribution to the rich literature and practice in the field, as these social statuses often are associated with fewer resources (Jia, Quan, et al., 2020; Ren et al., 2022; M. Ye et al., 2016). When disadvantaged status overlaps, individuals are at even higher risk (Bian, 2002). Social workers can directly utilize the results of this study to identify risk groups in the community and provide services based on their needs. For example, social workers can conduct targeted health education in local communities, addressing cardiovascular health and mental well-being issues. Mental health counseling services focusing on the needs of the oldest old adults should be available at the community level. They can also work with local government agencies to provide more accessible health care and financial assistance to the oldest old population. On the other hand, social policy and social work programs should also consider protective factors, such as healthy diets, exercise, and social engagements, promoting healthy and active aging for all older adults. Social workers are encouraged to advocate to local communities for developing exercise programs and leisure activities that engage older adults.

### *Limitations*

The study is not without limitations. First, we examined risk factors measured in the baseline survey, which took place in 1998, when the research field was less developed than it is today. Many risk factors, such as genetic factors, childhood intelligence, hearing function, sleep, and pain, were not intentionally measured. Future studies should analyze a more complete list of risk factors, such as hearing function (Livingston et al., 2020). Second, the Lasso regression analysis only considers linear relationships. The relationship between the predictors and the types of old-age cognitive function trajectory may be more complex, with many nonlinear relationships and interaction effects. Decision trees and ensemble methods should be considered in future research to allow high complicities of the predictive model. Third, our exploratory analysis leaves us little ground for causal inference or examining causal mechanisms.

## Conclusion

In summary, guided by the biopsychosocial framework (Engel, 1977), the current study identified two clusters of cognitive function trajectories among the oldest Chinese population and the risk and protective factors associated with them. The findings emphasize the importance of modifiable factors such as diet and lifestyle in influencing cognitive function in the oldest old. In future studies, social welfare policies, social work programs, and community services must recognize individuals belonging to low-stability (LS) groups and provide them with appropriate resources and support to promote healthy and active aging.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The author(s) reported there is no funding associated with the work featured in this article.

## References

Anstey, K. J., Peters, R., Zheng, L., Barnes, D. E., Brayne, C., Brodaty, H., Chalmers, J., Clare, L., Dixon, R. A., Dodge, H., Lautenschlager, N. T., Middleton, L. E., Qiu, C., Rees, G., Shahar, S., & Yaffe, K. (2020). Future directions for dementia risk reduction and prevention research: An international research network on dementia prevention consensus. *Journal of Alzheimer's Disease*, *78*(1), 3–12. https://doi.org/10.3233/JAD-200674

Aschwanden, D., Aichele, S., Ghisletta, P., Terracciano, A., Kliegel, M., Sutin, A. R., Brown, J., & Allemand, M. (2020). Predicting cognitive impairment and dementia: A machine learning approach. *Journal of Alzheimer's Disease: JAD*, *75*(3), 717–728. https://doi.org/10.3233/JAD-190967

Berk, R. A. (2020). *Statistical learning from a regression perspective*. Springer Nature.

Bian, Y. (2002). Chinese social stratification and social mobility. *Annual Review of Sociology*, *28*(1), 91–116. https://doi.org/10.1146/annurev.soc.28.110601.140823

Bolton, D., & Gillett, G. (2019). *The biopsychosocial model of health and disease: New philosophical and scientific developments*. Springer Nature. https://doi.org/10.1007/978-3-030-11899-0

Chen, L., & Lu, B. (2020). Cognitive reserve regulates the association between hearing difficulties and incident cognitive impairment evidence from a longitudinal study in China. *International Psychogeriatrics*, *32*(5), 635–643. https://doi.org/10.1017/S1041610219001662

Cloutier, S., Chertkow, H., Kergoat, M.-J., Gélinas, I., Gauthier, S., & Belleville, S. (2021). Trajectories of decline on instrumental activities of daily living prior to dementia in persons with mild cognitive impairment. *International Journal of Geriatric Psychiatry*, *36*(2), 314–323. https://doi.org/10.1002/gps.5426

Den Teuling, N. G. P., Pauws, S. C., & van den Heuvel, E. R. (2023). A comparison of methods for clustering longitudinal data with slowly changing trends. *Communications in Statistics – Simulation & Computation*, *52*(3), 621–648. https://doi.org/10.1080/03610918.2020.1861464

Dos Santos, S. B., Rocha, G. P., Fernandez, L. L., De Padua, A. C., & Reppold, C. T. (2018). Association of lower spiritual well-being, social support, self-esteem, subjective well-being, optimism and hope scores with mild cognitive impairment and mild dementia. *Frontiers in Psychology*, *9*, 371. https://doi.org/10.3389/fpsyg.2018.00371

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, *196*(4286), 129–136. https://doi.org/10.1126/science.847460

Gao, M., Kuang, W., Qiu, P., Wang, H., Lv, X., & Yang, M. (2017). The time trends of cognitive impairment incidence among older Chinese people in the community: Based on the CLHLS cohorts from 1998 to 2014. *Age and Ageing*, *46*(5), 787–793. https://doi.org/10.1093/ageing/afx038

Han, L., & Jia, J. (2021). Long-term effects of alcohol consumption on cognitive function in seniors: A cohort study in China. *BMC Geriatrics*, *21*(1), 699. https://doi.org/10.1186/s12877-021-02606-y

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

Hu, X., Gu, S., Zhen, X., Sun, X., Gu, Y., & Dong, H. (2021). Trends in cognitive function among Chinese elderly from 1998 to 2018: An age-period-cohort analysis. *Frontiers in Public Health*, *9*, 753671. https://doi.org/10.3389/fpubh.2021.753671

Jia, L., Du, Y., Chu, L., Zhang, Z., Li, F., Lyu, D., Li, Y., Li, Y., Zhu, M., Jiao, H., Song, Y., Shi, Y., Zhang, H., Gong, M., Wei, C., Tang, Y., Fang, B., Guo, D., Wang, F., & Qiu, Q. (2020). Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: A cross-sectional study. *The Lancet Public Health*, *5*(12), e661–e671. https://doi.org/10.1016/S2468-2667(20)30185-7

Jia, L., Quan, M., Fu, Y., Zhao, T., Li, Y., Wei, C., Tang, Y., Qin, Q., Wang, F., Qiao, Y., Shi, S., Wang, Y.-J., Du, Y., Zhang, J., Zhang, J., Luo, B., Qu, Q., Zhou, C., Gauthier, S., & Jia, J. (2020). Dementia in China: Epidemiology, clinical management, and research advances. *The Lancet Neurology*, *19*(1), 81–92. https://doi.org/10.1016/S1474-4422(19)30290-X

Kivipelto, M., Mangialasche, F., Snyder, H. M., Allegri, R., Andrieu, S., Arai, H., Baker, L., Belleville, S., Brodaty, H., Brucki, S. M., Calandri, I., Caramelli, P., Chen, C., Chertkow, H., Chew, E., Choi, S. H., Chowdhary, N., Crivelli, L., Torre, R. D. L., & Carrillo, M. C. (2020). World-Wide FINGERS Network: A global approach to risk reduction and prevention of dementia. *Alzheimer's & Dementia*, *16*(7), 1078–1094. https://doi.org/10.1002/alz.12123

Li, G., & Li, K. (2022). Turning point of cognitive decline for Chinese older adults from a longitudinal analysis: Protective factors and risk factors. *Healthcare*, *10*(11), 2304. Article 11. https://doi.org/10.3390/healthcare10112304

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. *2010 IEEE International Conference on Data Mining* (pp. 911–916). https://doi.org/10.1109/ICDM.2010.35

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., & Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the lancet commission. *The Lancet*, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Lv, X., Li, W., Ma, Y., Chen, H., Zeng, Y., Yu, X., Hofman, A., & Wang, H. (2019). Cognitive decline and mortality among community-dwelling Chinese older people. *BMC Medicine*, *17*(1), 63. https://doi.org/10.1186/s12916-019-1295-8

Mao, C., Li, Z.-H., Lv, Y.-B., Gao, X., Kraus, V. B., Zhou, J.-H., Wu, X.-B., Shi, W.-Y., Li, F.-R., Liu, S.-M., Yin, Z.-X., Zeng, Y., Shi, X.-M., & Newman, A. (2020). Specific leisure activities and cognitive functions among the oldest-old: The Chinese longitudinal healthy longevity survey. *Journals of Gerontology: Series A*, *75*(4), 739–746. https://doi.org/10.1093/gerona/glz086

Marmot, M., & Wilkinson, R. (2005). *Social determinants of health*. OUP Oxford.

Ma, J., Yang, Y., Wan, Y., Shen, C., Qiu, P., & Liu, S. Y. (2021). The influence of childhood adversities on mid to late cognitive function: From the perspective of life course. *PLoS One*, *16*(8), e0256297. https://doi.org/10.1371/journal.pone.0256297

McInnis-Dittrich, K. (2009). *Social work with older adults: A biopsychosocial approach to assessment and intervention* (3rd ed.). Allyn & Bacon.

McWhirter, L., Ritchie, C., Stone, J., & Carson, A. (2020). Functional cognitive disorders: A systematic review. *The Lancet Psychiatry*, *7*(2), 191–207. https://doi.org/10.1016/S2215-0366(19)30405-5

Mukadam, N., Sommerlad, A., Huntley, J., & Livingston, G. (2019). Population attributable fractions for risk factors for dementia in low-income and middle-income countries: An analysis using cross-sectional survey data. *The Lancet Global Health*, *7*(5), e596–e603. https://doi.org/10.1016/S2214-109X(19)30074-9

Ren, R., Qi, J., Lin, S., Liu, X., Yin, P., Wang, Z., Tang, R., Wang, J., Huang, Q., Li, J., Xie, X., Hu, Y., Cui, S., Zhu, Y., Yu, X., Wang, P., Zhu, Y., Wang, Y., Huang, Y., & Wang, G. (2022). The China Alzheimer Report 2022. *General Psychiatry*, *35*(1), e100751. https://doi.org/10.1136/gpsych-2022-100751

Smith, L., Jacob, L., Kostev, K., Butler, L., Barnett, Y., Pfeifer, B., Soysal, P., Grabovac, I., López-Sánchez, G. F., Veronese, N., Yang, L., Oh, H., & Koyanagi, A. (2021). Mild cognitive impairment is associated with fall-related injury among adults aged ≥65 years in low- and middle-income countries. *Experimental Gerontology*, *146*, 111222. https://doi.org/10.1016/j.exger.2020.111222

Sullivan, E. V., & Pfefferbaum, A. (2006). Diffusion tensor imaging and aging. *Neuroscience & Biobehavioral Reviews*, *30*(6), 749–761. https://doi.org/10.1016/j.neubiorev.2006.06.002

Tan, T. K., & Feng, Q. (2022). Validity and reliability of mini-mental state examination in older adults in China: Inline mini-mental state examination with cognitive functions. *International Journal of Population Studies*, *8*(1), 1–16. https://doi.org/10.18063/ijps.v8i1.1285

Tu, L., Lv, X., Yuan, C., Zhang, M., Fan, Z., Xu, X., Zeng, Y., Yu, X., & Wang, H. (2020). Trajectories of cognitive function and their determinants in older people: 12 years of follow-up in the Chinese longitudinal healthy longevity survey. *International Psychogeriatrics*, *32*(6), 765–775. https://doi.org/10.1017/S1041610220000538

Verboon, P., & Pat-El, R. (2022). Clustering longitudinal data using R: A Monte Carlo study. *Methodology*, *18*(2), 144–163. Article 2. https://doi.org/10.5964/meth.7143

Veríssimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., & Ullman, M. T. (2022). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nature Human Behaviour*, *6*(1), 97–110. Article 1. https://doi.org/10.1038/s41562-021-01169-7

Wang, Q., & Kang, W. (2019). Childhood socioeconomic circumstances, social status, and health in older age: Are they related in China? *Advances in Life Course Research*, *42*, 100289. https://doi.org/10.1016/j.alcr.2019.100289

Wang, J., Xiao, L. D., Wang, K., Luo, Y., Li, X., & Lennon, M. (2020). Cognitive impairment and associated factors in rural elderly in North China. *Journal of Alzheimer's Disease*, *77*(3), 1241–1253. https://doi.org/10.3233/JAD-200404

Williams, G. (2011). *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer Science & Business Media.

Wilson, R. S., Wang, T., Yu, L., Bennett, D. A., & Boyle, P. A. (2020). Normative cognitive decline in old age. *Annals of Neurology*, *87*(6), 816–829. https://doi.org/10.1002/ana.25711

Xu, J., Wang, J., Wimo, A., Fratiglioni, L., & Qiu, C. (2017). The economic burden of dementia in China, 1990–2030: Implications for health policy. *Bulletin of the World Health Organization*, *95*(1), 18–26. https://doi.org/10.2471/BLT.15.167726

Yang, L., Cheng, J., & Wang, H. (2021). Place of residence and cognitive function in older adults in China: The mediating role of social participation. *International Journal of Environmental Research and Public Health*, *19*(1), 13. https://doi.org/10.3390/ijerph19010013

Ye, M., Chen, L., & Kahana, E. (2016). Mealtime interactions and life satisfaction among older adults in Shanghai. *Journal of Aging and Health*, *29*(4), 620–639. https://doi.org/10.1177/0898264316641080

Ye, M., Chen, Y., & Peng, Y. (2017). A new era in living arrangements: Determinants of quality of life among Chinese older adults. In M.-C. Tsai & W. Chen (Eds.), *Family, Work and Wellbeing in Asia* (pp. 43–64). Springer. https://doi.org/10.1007/978-981-10-4313-0_3.

Ye, K. X., Sun, L., Wang, L., Khoo, A. L. Y., Lim, K. X., Lu, G., Yu, L., Li, C., Maier, A. B., & Feng, L. (2023). The role of lifestyle factors in cognitive health and dementia in oldest-old: A systematic review. *Neuroscience & Biobehavioral Reviews*, *152*, 105286. https://doi.org/10.1016/j.neubiorev.2023.105286

Yin, Z., Fei, Z., Qiu, C., Brasher, M. S., Kraus, V. B., Zhao, W., Shi, X., & Zeng, Y. (2017). Dietary diversity and cognitive function among elderly people: A population-based study. *The Journal of Nutrition, Health & Aging*, *21*(10), 1089–1094. https://doi.org/10.1007/s12603-017-0912-5

Yu, J., Qiushi, F., Jintai, Y., Yi, Z., & Lei, F. (2020). Late-life cognitive trajectories and their associated lifestyle factors. *Journal of Alzheimer's Disease: JAD*, *73*(4), 1555–1563. https://doi.org/10.3233/JAD-191058

Zeng, Y., Vaupel, J. W., Xiao, Z., Chunyuan, Z., & Yuzhi, L. (2002). Sociodemographic and health profiles of the oldest old in China. *Population and Development Review*, *28*(2), 251–273. https://doi.org/10.1111/j.1728-4457.2002.00251.x

Zhang, Q., Wu, Y., Han, T., & Liu, E. (2019). Changes in cognitive function and risk factors for cognitive impairment of the elderly in China: 2005–2014. *International Journal of Environmental Research and Public Health*, *16*(16), 2847. https://doi.org/10.3390/ijerph16162847

Zhao, X., Zhou, Y., Wei, K., Bai, X., Zhang, J., Zhou, M., & Sun, X. (2021). Associations of sensory impairment and cognitive function in middle-aged and older Chinese population: The China health and retirement longitudinal study. *Journal of Global Health*, *11*, 08008. https://doi.org/10.7189/jogh.11.08008

Zhou, L., Ma, X., & Wang, W. (2021). Relationship between cognitive performance and depressive symptoms in Chinese older adults: The China Health and Retirement Longitudinal Study (CHARLS). *Journal of Affective Disorders*, *281*, 454–458. https://doi.org/10.1016/j.jad.2020.12.059

Zhu, Y., Li, C., Xie, W., Zhong, B., Wu, Y., & Blumenthal, J. A. (2022). Trajectories of depressive symptoms and subsequent cognitive decline in older adults: A pooled analysis of two longitudinal cohorts. *Age and Ageing*, *51*(1), afab191. https://doi.org/10.1093/ageing/afab191

Zhu, C.-E., Zhou, L., & Zhang, X. (2022). Effects of leisure activities on the cognitive ability of older adults: A latent variable growth model analysis. *Frontiers in Psychology*, *13*, 838878. https://doi.org/10.3389/fpsyg.2022.838878

Zittel, K. M., Lawrence, S., & Wodarski, J. S. (2002). Biopsychosocial model of health and healing. *Journal of Human Behavior in the Social Environment*, *5*(1), 19–33. https://doi.org/10.1300/J137v05n01_02