Received: 4 December 2024 | Revised: 17 January 2025 | Accepted: 17 January 2025

DOI: 10.1002/alz.14624

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

A systematic exposure-wide framework leveraging machine learning to identify multidomain exposure factors and their joint influence on cognitive function: Evidence from a neurological cohort

Jingtao Wu[1,2,3] | Bowen Yin[4] | Rui Wen[4] | Huanting Pei[4] | Siqi Zhu[4] | Jiaxin Zhao[1,2] | Yanbing Li[1,2] | Ming Yang[1,2,3] | Yaoyu Hu[1,2,3] | Qun Xu[1,2,3] | Ang Li[1,2] | Yuxia Ma[4]

[1]Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing, China

[2]Center of Environmental and Health Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

[3]Medical Center for Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

[4]Department of Nutrition and Food Hygiene, School of Public Health, Hebei Medical University, Hebei Key Laboratory of Environment and Human Health, Shijiazhuang, China

Correspondence
Yuxia Ma, Department of Nutrition and Food Hygiene, School of Public Health, Hebei Medical University, Hebei Key Laboratory of Environment and Human Health, Shijiazhuang 050017, China.
Email: mayuxia@hebmu.edu.cn

Ang Li, Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing 100005, China.
Email: pumcleon@ibms.pumc.edu.cn

Funding information
National Key R&D Program of China, Precision Medicine Project–Cohort Study on Nervous System Diseases, Grant/Award Number: 2017YFC0907700; Community-based Cohort Study on Nervous System Diseases, Grant/Award Number: 2017YFC0907701; Noncommunicable Chronic Diseases-National Science and Technology Major Project, Grant/Award Number: 2023ZD0513200

Abstract

INTRODUCTION: Cognitive decline has become a growing public concern, yet large-scale exposure data identifying the contributing factors remain limited.

METHODS: We conducted an exposure-wide association study involving 1142 participants and 207 exposures, using machine learning to assess the relative contribution and joint effects of key factors. Cluster analysis and intervention simulation trials helped identify high-risk subpopulations and the potential benefits of targeted interventions.

RESULTS: In adjusted mixed models, the socioeconomic status domain emerged as the strongest predictor of longitudinal global cognitive score ($\beta = 2.91$, $p < 0.0001$, $q < 0.0001$), while the dietary domain also played an important role in memory function. The cluster analysis found that the "unfavorable lifestyle" dominated phenotype was associated with the poorest cognitive outcomes. Simulation trials indicated that cognitive scores could improve by shifting individuals from unfavorable to favorable phenotypes.

DISCUSSION: Cognitive health requires multidomain interventions, particularly in the socioeconomic and dietary fields, and necessitates collaboration between government and individuals.

KEYWORDS
cognitive function, exposure-wide association study, machine learning, multidomain intervention

Highlights

- The exposure-wide association study design, which assesses a broad range of exposures, is used to identify novel variables and understand their contributions to cognitive function.
- The findings from the multidomain analysis indicate that socioeconomic status is the most significant contributor to global cognitive function, while diet plays the largest role in memory function.
- Increasing the proportion of favorable phenotypes through multidomain interventions can significantly enhance public cognitive health.

1 | BACKGROUND

With populations aging and life expectancy rising, cognitive decline has become a growing public concern, affecting patients' quality of life.[1] Neurodegenerative diseases associated with aging have emerged as leading causes of death and disability.[2] Previous studies have found 60 years of chronological age to be a specific period of significant change for age-related diseases, emphasizing the need to promote early diagnosis and prevention strategies.[3] Therefore, focused efforts to 1 prevent and mitigate cognitive decline among people in early old age (60-70 years) are crucial, as this group is particularly vulnerable to age-related changes. Although effective therapies for dementia have recently been approved,[4] the concerns about efficacy and safety[5,6] are still debated. Meanwhile, as factors that influence cognition are increasingly discovered, cognitive impairment can be stabilized or even improved through positive lifestyle interventions.[7,8] Therefore, developing favorable lifestyle patterns is still the preferred way to maintain cognitive function and alleviate cognitive decline.

In real-life scenarios, the human body is simultaneously exposed to multiple factors, and the concept of the exposome was proposed to emphasize systematically exploring complex and heterogeneous exposure–health associations.[9] Given that cognitive decline is influenced by multifactorial exposures, health multidomain interventions are recommended.[10] But knowledge of multidomain strategies that could mitigate cognitive function remain relatively limited. Current studies focus mainly on validating a limited set of hypothesized modifiable factors for cognition function, such as sleep duration.[11] Without broader investigation, additional factors may be overlooked or unknown. Investigating a wide range of factors could help confirm existing relationships and identify potentially novel prevention targets. Systematically identifying multiple factors across multidomain is of great importance but its application in cognitive function research remains limited.

A growing body of evidence has identified various modifiable risk factors for cognitive health, necessitating a deeper understanding of their relative and overall contribution and joint effects. Meanwhile, data-driven methods like machine learning (ML) have been proposed to objectively identify novel variables and reveal how these variables contribute to the development of health outcomes.[12] Compared to conventional statistical methods, ML algorithms are better suited for processing high-dimensional datasets and identifying distinctive factors associated with cognitive impairment that were previously difficult to quantify.[13] This knowledge can help guide targeted interventions based on key factors, especially when health care resources are limited.

In addition, as understanding of cognitive intervention strategies continues to deepen, it is important to focus the application of these strategies on individuals at the highest risk or those likely to benefit the most from the intervention. Thus, there is an urgent need to determine high-risk subpopulations for cognitive impairment. Differential characteristics of exposure patterns enable us to effectively target these vulnerable populations. However, the knowledge of these strategies has not yet been fully applied to the evaluation of cognitive interventions.

An exposure-wide association study (EWAS), similar to genome- or epigenome-wide association studies, has been proposed as a hypothesis-free strategy for investigating the association between multiple exposures and specific health outcomes. In previous EWAS research, over 100 exposure variables have been used to create a detailed profile of all possible factors that could contribute to health outcomes, such as depression,[14] which provided a more comprehensive and novel understanding of disease etiology. On the other hand, to strengthen conclusions about which high-risk subpopulations may be high-priority intervention targets, an alternative unsupervised ML approach can be used to identify groups of individuals sharing specific exposure-factor profiles. Moreover, we extend the Monte Carlo simulation approach to evaluate the improvement of cognitive functioning, assuming that exposure patterns were intervened.

In this study, utilizing data from the Community Cohort Study of Nervous System Diseases (CCSNSD), we conducted an EWAS design to comprehensively identify variables influencing cognitive function. Then, by taking advantage of the ML algorithms, we assessed both the

individual and joint contributions of the exposure domain to longitudinal changes in cognitive function. Finally, we quantified the potential population-level cognitive benefits of multidomain interventions using Monte Carlo simulations.

RESEARCH IN CONTEXT

1. **Systematic review**: Previous studies have focused mainly on the effects of a limited number of factors on cognitive function. Recent research, however, suggests that multidimensional lifestyle interventions may better reflect real-world conditions. Despite this, there is a lack of large-scale exposure data to systematically identify the contributing factors and quantify their relative contributions.
2. **Interpretation**: Multiple exposure domains were associated significantly with cognitive function, with socioeconomic status and diet emerging as the most significant contributors. The "unfavorable lifestyle" phenotype was linked to the poorest cognitive outcomes, whereas precise interventions targeting this group showed potential for improving cognitive health.
3. **Future directions**: More attention should be placed on multidomain and precise interventions for individuals with an "unfavorable lifestyle." Further research is needed to validate these findings and assess the benefits after the application of the intervention.

2 | METHODS

2.1 | Study design and population

This study analyzed the data from CCSNSD, conducted by the National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention. The research aimed to identify potential risk factors for three nervous system diseases: epilepsy in individuals >1 year, as well as Alzheimer's disease (AD) and Parkinson's disease (PD) in individuals 55 years of age or older. Participants without these diseases were enrolled at baseline in 2018, through a multistage stratified random-sampling approach in the Hebei, Zhejiang, Shaanxi, and Hunan provinces of China. In this study, available data were acquired from 4855 participants recruited in Hebei Province. Two cities and two counties within the province were selected randomly for sampling. Individuals 55 years of age or older and those without nervous system diseases (epilepsy, AD, and PD) were included, and the participants who lacked an assessment of cognitive function assessment were excluded (Figure 1). Finally, 1142 participants were collected at baseline, with 1083 (94.8%) completing follow-up between 2020 and 2021. All participants provided written informed consent and the study received approval from the Ethics Review Committee of the National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention (No. 2017-020).

2.2 | Exposure assessment

This exposure information, including socioeconomic status (SES), demographic factors (DM), diet, activities of daily living (ADLs), medical history, physical activities, and psychosocial factors, was collected by trained investigators.

2.2.1 | SES domain

SES data were collected through a questionnaire designed by the research team, which included information on educational level, household income, living region (rural/urban), employment status, and job type.

2.2.2 | DM domain

Demographic information such as sex (male/female), age (years), smoking status, drinking status, average sleep duration, and height, weight, and body mass index (BMI) was collected by trained investigators. Age was determined by subtracting the birth date from the survey date. Average sleep duration was categorized as short (<6 h), optimal (6–8 h), or long (>8 h) based on established evidence.[15] Height was measured using a stadiometer (Seca, Germany) with a precision of 0.1 cm, and participants were asked to remove their shoes. Weight was recorded using an electronic scale with a precision of 0.1 kg, and participants were asked to remove their shoes and to minimize clothing. Hip and waist measurements were taken using an inelastic measuring tape with a precision of 0.1 cm. BMI was calculated by dividing body weight in kilograms by height in meters squared (kg/m^2) and was categorized into four groups: underweight (<18.5 kg/m^2), normal (18.5–23.9 kg/m^2), overweight (24.0–27.9 kg/m^2), and obese (≥28.0 kg/m^2), based on Chinese adult criteria.[16] All physiological measurements were measured by a trained investigator in a quiet, well-ventilated room with a controlled, consistent temperature, following standardized procedures.

2.2.3 | Diet domain

Dietary information was assessed using a validated semiquantitative food frequency questionnaire covering 65 food items (Table S1).[17] Participants were asked about their dietary habits over the previous 12 months, including the frequency and amount of each food item consumed. Previous studies[18,19] have highlighted the importance of dietary components as factors in the onset of dementia and cognitive decline. The diet components included nine healthy food groups

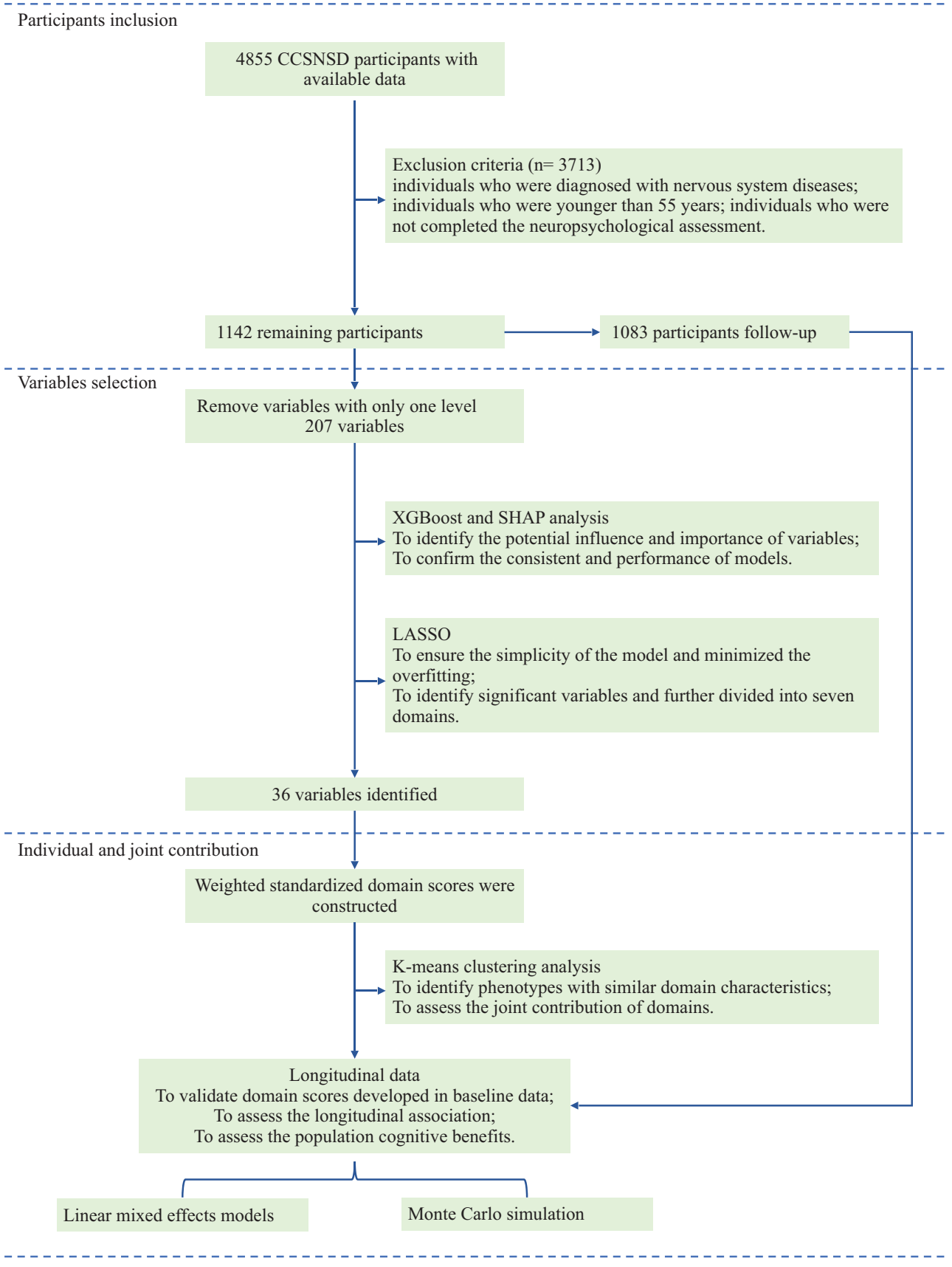


FIGURE 1 Overview of analytic design.

(green leafy vegetables, other vegetables, nuts, berries, beans, whole grains, seafood, poultry, and wine) and three unhealthy food groups (red meats, pastries and sweets, and fried and fast foods). These categories were scored according to the MIND (Mediterranean-DASH Intervention for Neurodegenerative Delay) Diet, which combines elements from the DASH (Dietary Approaches to Stop Hypertension) and Mediterranean diets, adjusted to include foods linked to a slower rate of cognitive decline.[18,20] Fruit and dairy scores, not included in the MIND Diet, were assessed using DASH Diet criteria.[21]

2.2.4 | ADL domain

ADLs were measured using the Activities of Daily Living scale developed by Elena Yu and William Liu at the University of Illinois at Chicago.[22,23] This scale consists of 20 items, including eight components of basic activities of daily living (BADLs) and 12 instrumental activities of daily living (IADLs), scored from 1 to 4, with higher scores indicating worse ability (1 = can do it myself, 2 = have some difficulty doing but can still do it by myself, 3 = need help to do it, 4 = cannot do it at all). The Chinese version of the Activities of Daily Living scale has been verified and recommended for older Chinese people.[24] The 20 individual ADL items as separate variables were included in further analyses.

2.2.5 | Medical history domain

Detailed medical history, including information on diseases or predisposing factors for three nervous system diseases (epilepsy, AD, and PD) and other comorbidities (such as hypertension, diabetes, cardiovascular disease, cancer, and chronic respiratory disease) was collected according to China's diagnosis and treatment guidelines.[25] These diseases were defined based on information from self-reported physician diagnoses or data linkage to the disease registry.

2.2.6 | Physical activities domain

Physical activities were measured using the China Health and Nutrition Survey (CHNS) questionnaire, reported previously in other studies.[26,27] Participants reported the average time spent each day on activities such as physical exercise, sedentary behavior, domestic work, and commuting.

2.2.7 | Psychosocial domain

Psychosocial factors were evaluated using the Geriatric Depression Scale (GDS-30), which consisted of 30 items[28] and has been validated in a large Chinese aging population.[29]

2.3 | Cognitive outcomes

Global cognitive function was assessed using the Chinese versions of the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA). Both instruments are valid and reliable for Chinese populations, considering cultural and linguistic differences.[30,31] The MMSE and MoCA were administered face-to-face by trained investigators, with completion times of 5–10 min and 10–15 min, respectively. Each instrument has a total score range of 0 to 30, with higher scores indicating better cognitive function.

In addition, memory function was measured using the World Health Organization/University of California Los Angeles Auditory Verbal Learning Test (AVLT),[32] which assesses both immediate recall and long-delay free recall. During the test, a list of 15 nouns is read aloud by the examiner, and the participant is asked to recall the words immediately. After three learning and recall trials, a delayed recall test is administered ≈30 min later without any reminders. Consistent with the previous study,[33,34] the memory status was determined by standard *z*-scores for the AVLT scores, which were calculated based on the respective mean and SD of test scores. The composite *z*-score for memory function was constructed by averaging the *z*-scores for each test in the AVLT.

These cognitive outcomes were assessed at the baseline period and further measured during the follow-up period to capture cognitive changes. As MoCA offers better sensitivity and broader assessment across cognitive domains,[35] the MoCA was chosen as the primary outcome for EWAS analysis, whereas MMSE and AVLT scores were also applied to confirm the robustness of results.

2.4 | Quality assurance and control

To minimize recall bias and subjective inaccuracies, we established a comprehensive quality control process throughout the study. During the questionnaire development phase, to enhance the accuracy of the survey, standardized scales (Activities of Daily Living scale and GDS-30) were used to assess ADLs and psychological factors. For dietary assessments, to avoid inaccuracies in estimating intake, we prepared food models to help participants precisely measure their food consumption. In addition, before surveys, participants were informed to bring their medical records to assist in verifying their personal medication and disease history. We aimed to ensure the accuracy of our study by using tools and records that were as objective as possible. Furthermore, standard operating procedures (SOPs) were applied for information collection and cognitive measurements. Before the beginning of the survey, all investigators received standardized investigation training. A minimum of six investigators were required to participate at each survey site, and changes of investigators midway through the process were avoided to the extent possible. The sampled participants were notified in advance and asked to make reasonable agenda arrangements to participate in the survey. During the survey, investigators were asked to question the participants meticulously and carefully,

especially in the cognitive measurement section, which was required to be conducted in strict accordance with the guidelines formulated by the project team. After the survey was completed, data verification teams were set up at local survey sites and at the provincial level to ensure the completeness and accuracy of the data. Overall, these measures were favorable for maintaining a relatively low rate of missing data (the overall rate <3%).

2.5 | Statistical analyses

To analyze the relative effects of all available exposures, variables with more than one level were included in the analysis. Data on the remaining 207 exposure factors, measured or derived at baseline, were then collected (Table S2). Unordered categorical variables (e.g., sleep duration: short, optimal, long) were dummy-coded into dichotomous variables (sleep duration/optimal, sleep duration/long), whereas ordinal variables with a logical order (e.g., household income: <1000, 1000–3999, 4000–7999, ≥8000 CNY/moth) were treated as continuous variables. Baseline characteristics of the cohort were summarized as percentages for categorical variables, and as means (SDs) and medians (interquartile ranges [IQRs]) for continuous variables. Data were assumed to be missing at random and were imputed using multiple imputation by chained equations, with 5 imputed datasets and 10 iterations. A predictive mean matching model was specified for each variable, and final missing values were imputed using the mode for categorical features and the mean for numerical features.[36,37] To deal with multiple testing issues, the false discovery rate (FDR) procedure was applied to calculate p_{FDR} value (q value), with significance set to $\alpha = 0.05$ (two-sided). Analyses were performed using R version 4.1.2.

2.6 | Exposure factor identification and domain construction

2.6.1 | ML analysis to identify the contribution of exposure factors on global cognitive function

To identify potential factors influencing baseline cognitive scores, we predicted MoCA scores using gradient-boosting decision trees based on the extreme Gradient Boosting (XGBoost) algorithm.[38] XGBoost was an ensemble ML algorithm with a gradient-boosting framework and showed superior performance compared to other ML models in terms of neurological disease.[39] The XGBoost models were optimized through fivefold cross-validation and evaluated using root mean square error (RMSE) to predict MoCA levels in the left-out group (model construction details are reported in Supplementary Methods). In In addition, we calculated the explained variance (EV) values for each exposure domain separately to estimate their relative contributions to cognitive function. To address the challenge of interpreting ML models, SHapley Additive exPlanation (SHAP) analysis was used to rank the importance of input features and overcome the "black-box" issue.[40] SHAP values were used to interpret our XGBoost and objectively evaluate the factors influencing MoCA scores in participants (more details are included in Supplementary Methods).

2.6.2 | Protective cognitive exposure score construction for each domain

Least absolute shrinkage and selection operator (LASSO) has been used widely for simultaneous estimation and consistent variable selection in field of cognitive outcomes.[41] In our study, LASSO regression was used to ensure model simplicity and minimize overfitting during training. LASSO, proposed by Tibshirani,[42] shrinks the coefficients of variables considered unimportant to zero. To select the optimal hyperparameter λ for the LASSO regression model, we employed the *glmnet* package in R. Specifically, we used the *cv.glmnet* function to perform 10-fold cross-validation. We set the regularization of LASSO regression as 1 standard error to achieve the most concise model.[43] In this process, 207 variables were included in LASSO model, and variables with coefficients reduced to zero were removed from further analysis. The variables selected were further divided into seven domains: SES, DM, diet, ADLs, medical history, physical activities, and psychosocial factors. Protective cognitive exposure scores for each domain were generated based on the β coefficients of each standardized variable in the LASSO models.[44] The standardized variables were multiplied by the β coefficients, summed, and divided by the sum of the β coefficients. Higher protective cognitive exposure domain scores indicated higher MoCA scores. We then assessed the association of each domain score with MoCA scores in the baseline population using linear regression models. To further validate the consistency and performance of the LASSO model, we recalculated the RMSE and EV of the XGBoost model using the selected variables (concise model) and compared the performance with the original XGBoost model using all available variables (full model).

2.7 | The association of exposure domains and longitudinal cognitive change

2.7.1 | Individual domain contribution

After the exposure domains were identified based on train data during the baseline period, the longitudinal cognitive change was further evaluated to validate the above findings. The longitudinal association between different domain scores and cognitive function was analyzed using linear mixed-effects models, which incorporated survey time and a random intercept at the individual level to examine the relationships between each domain and MoCA scores. As random intercept was included, effect sizes from the linear mixed model indicated each domain's impact on cognitive function, with positive values showing improvement and negative values indicating decline, adjusted for individual variability. Moreover, alternative measures of global cognition (MMSE) and memory function (composite z-score of the AVLT) were also evaluated using the same linear mixed-effects models.

2.7.2 | Phenotypic clustering for joint contribution

To identify clusters of participants with similar multidomain characteristics and further find the high-risk subpopulation, we performed an unsupervised k-means clustering analysis on the longitudinal data (n = 1083) to identify distinct exposure phenotypes. The number of clusters (k) was prespecified using the elbow method and the silhouette method, which compare measures of cluster cohesion and separation for different choices of k.[45] Once the optimal number of phenotypes was determined, patterns of exposure variables were visualized in ranked plots by the mean standardized difference between phenotype pairs,[46] and domain scores were visualized using radar charts for each phenotype. The longitudinal associations between different phenotypes and cognitive outcomes were also assessed using linear mixed-effects models adjusted to the same covariates as individual domain analyses.

2.7.3 | Monte Carlo simulation for population cognitive benefits

To better understand the implications of phenotypes for public cognitive health and policy programs, we conducted Monte Carlo simulations to generate a series of phenotype scenarios. Under each scenario, we simulated the same sample size population (1083 participants for 10,000 iterations) and ensured that the proportion of phenotypes in the baseline population remained consist with the original data. Next, for each scenario, we assumed that the public and policy interventions on phenotypes occurred by increasing the proportion of participants in specific phenotypes during the follow-up period and decreasing the proportion of participants in other phenotypes to compensate. In each simulated population, cognitive changes were estimated according to the effects calculated in the original cohort and phenotypes. Based on this assignment, the difference in cognitive scores between the original cohort and simulated intervention scenarios was analyzed using the t-test or Wilcoxon test according to data distribution. For a given trial, if the test was not significant at the 0.05 level, we concluded no benefit from the intervention. However, if the p-value was significant at the .05 level, the superiority or inferiority was determined by the direction of the cognitive scores. The differences in cognitive scores were calculated based on an average of 10,000 simulations.[46] This approach explored the heterogeneity of cognitive effects based on the frequency distributions of these phenotypes. We assessed how population cognitive health could change with alterations in the relative distribution of the identified phenotypes.

2.8 | Sensitivity analyses

We conducted several additional sensitivity analyses to evaluate the robustness of our findings. First, to further investigate the interaction effects between multidomain factors on longitudinal cognitive change, interaction terms between each pair of domains were added individually to the original linear mixed-effects models to assess their statistical significance. Significant interaction terms were then included together in a model. The final model incorporated all statistically significant interaction terms to evaluate the interaction effects of multiple domains on longitudinal cognitive change. Second, another method for handling missing data was applied. Participants at baseline with missing data for key variables (36 variables used to calculate exposure domain scores) were excluded, and a new dataset was constructed. This updated dataset was used to reassess the association between exposure domains and longitudinal cognitive changes. The simulation trials were then reanalyzed to confirm the robustness of the findings.

3 | RESULTS

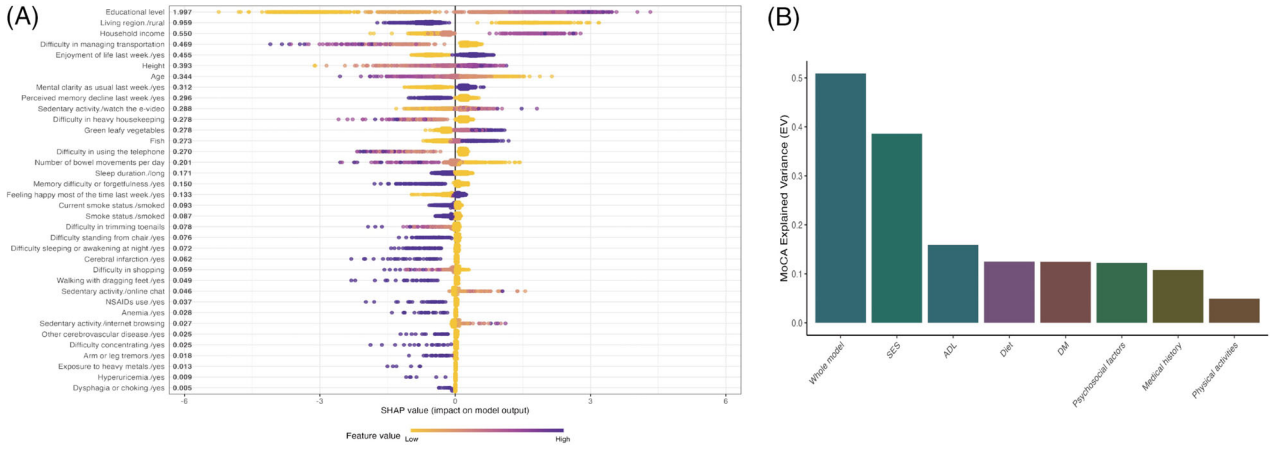
3.1 | Participants characteristics

The mean age of the participants was 67.3 years (SD 7.27 years), and 85.46% completed high school or a lower level of education; 72.76% of the participants lived in a rural region, and over half of them (59.11%) had household incomes between 1000 and 3999 Chinese yuan per month. Table S3 shows the detailed characteristics of the exposure variables (36 variables selected in LASSO).

3.2 | Identification of individual variables affecting global cognitive function

To better understand and identify which variables (Table S2) affect global cognitive function the most, we used XGBoost models (full model) and assessed the influence of how individual variables drive MoCA-predicting models using SHAP analysis. The variables were ranked by importance, and SHAP values were shown for the top 40 features. The SHAP summary dot plot (Figure S1) highlights the contribution of various factors to MoCA scores based on mean SHAP values. Variables related to SES—including educational level, living region, and household income—were found to have the greatest effect on MoCA scores. Next, ADL-related variables such as difficulty in managing transportation, heavy housekeeping, and using the telephone, also had a strong impact. Meanwhile, other domains also affected this model outcome, including DMs (e.g., age), dietary factors (e.g., green leafy vegetables), medical history (e.g., number of bowel movements per day), physical activities (e.g., sedentary activity/watch the e-video), and psychosocial factors (e.g., enjoyment of life last week/yes). In addition, the SHAP summary dot plot also visually represented the direction of each variable's influence on MoCA scores, such as higher educational attainment, living in an urban area, no difficulty in managing transport, and higher household income appeared to increase MoCA scores.

To reduce model complexity and dependency on multiple variables, we also constructed LASSO models to predict MoCA scores using all available variables. Of 207 risk factors, 36 were significantly associated with global cognitive function (Table S4 and Figure S2). Of these, 25 factors were detrimental to cognitive scores, whereas 11 were protective. Among the top factors, five were beneficial to cognitive scores:



(C)

Domain component	MoCA scores	β (95% CI)	P value	q value	MMSE scores	β (95% CI)	P value	q value	AVLT scores	β (95% CI)	P value	q value
SES	2.91 (2.61, 3.21)		<0.0001	<0.0001	2.03 (1.78, 2.29)		<0.0001	<0.0001	0.09 (0.04, 0.14)		0.0005	0.0010
ADL	1.88 (1.61, 2.15)		<0.0001	<0.0001	1.61 (1.38, 1.84)		<0.0001	<0.0001	0.11 (0.07, 0.16)		<0.0001	<0.0001
Diet	1.23 (0.94, 1.51)		<0.0001	<0.0001	0.82 (0.58, 1.06)		<0.0001	<0.0001	0.28 (0.23, 0.33)		<0.0001	<0.0001
DM	0.87 (0.47, 1.27)		<0.0001	<0.0001	0.83 (0.48, 1.17)		<0.0001	<0.0001	0.11 (0.04, 0.18)		0.0013	0.0013
Psychosocial factors	1.61 (1.27, 1.95)		<0.0001	<0.0001	1.67 (1.38, 1.96)		<0.0001	<0.0001	0.14 (0.08, 0.20)		<0.0001	<0.0001
Medical history	2.20 (1.60, 2.79)		<0.0001	<0.0001	1.45 (0.93, 1.96)		<0.0001	<0.0001	0.18 (0.07, 0.28)		0.0009	0.0013
Physical activities	0.27 (0.00, 0.54)		0.0497	0.0497	0.46 (0.23, 0.69)		0.0001	0.0001	-0.08 (-0.13, -0.03)		0.0011	0.0013

FIGURE 2 Contribution rank of variables and domain components. (A) SHAP summary dot plot illustrates the impact on model output (SHAP values) for the whole population during the baseline period ($N = 1142$). The MoCA scores increase with the SHAP values of the features. Each dot represents a participant's SHAP value for a specific variable, with colors indicating the actual variable values: higher values are shown in purple, and lower values in orange. Dots are stacked vertically to reflect density. The figure displays SHAP values for concise model. (B) Explained variance of predicted MoCA scores was determined by XGBoost algorithms after fivefold cross-validation in the left-out group. These were trained exclusively on variables from each domain component (see Table S3 for variables included in each group) or the concise model (all selected 36 variables). (C) The longitudinal association of protective cognitive exposure scores with global cognitive function and memory function was analyzed using a linear mixed-effects model. ADLs, activities of daily living; DM, demographic factors; MoCA, Montreal Cognitive Assessment; SES, socioeconomic status; SHAP, SHapley Additive exPlanation; XGBoost, extreme Gradient Boosting.

educational level, household income, enjoyment of life last week/yes, green leafy vegetables, and mental clarity as usual last week/yes, whereas three were harmful to cognitive scores: difficulty in managing transportation, living region/rural, and difficulty in using the telephone. Detailed information on all selected factors is provided in Table S3. To validate the effectiveness of the LASSO model in variable selection, we assessed the goodness of fit for the XGBoost model (concise model) by including only the 36 selected variables. Similar performances were observed between the full model (all 207 variables) and concise model (Table S5). SHAP analyses also supported the variable selection results, as the selected variables consistently made significant contributions to model predictions in both the full model (Figure S1) and the concise model (Figure 2A).

3.3 | Relative individual contributions of domains to cognitive function

After constructing protective cognitive exposure scores, we evaluated the contribution of each domain component to cognitive function. In baseline participants, all seven favorable domain scores significantly improved cognitive function: SES ($\beta = 3.58$, 95% confidence interval [CI] = 3.19 to 3.97), medical history ($\beta = 2.79$, 95% CI = 2.10 to 3.48), psychosocial factors ($\beta = 1.68$, 95% CI = 1.24 to 2.12), ADLs ($\beta = 1.66$, 95% CI = 1.30 to 2.02), DM ($\beta = 1.05$, 95% CI = 0.55 to 1.55), diet ($\beta = 0.70$, 95% CI = 0.32 to 1.08), and physical activities ($\beta = 0.54$, 95% CI = 0.16 to 0.92) (Table S6). We used fivefold cross-validation to predict the EV of each domain group or the whole model on MoCA scores in the left-out baseline participants (Figure 2B). SES was the best predictor (EV = 38.59%), followed by ADLs (explaining 15.90% of global cognition scores). Diet, DM, and psychosocial factors each contributed significantly, explaining ≈12% of the variance. Physical activities (EV 4.93%) and medical history (EV 10.78%) exhibited the lowest predictive power (Table S5 for specific model parameters).

3.4 | The longitudinal association of domains with cognitive function

We used linear mixed-effects models to analyze the impact of domain components on the longitudinal changes in cognitive function consistent with findings during baseline period. The results showed that

higher SES had the strongest positive effect on MoCA scores ($\beta = 2.91$, 95% CI = 2.61 to 3.21, $p < 0.0001$, $q < 0.0001$), followed by better medical health status ($\beta = 2.2$, 95% CI = 1.60 to 2.79, $p < 0.0001$, $q < 0.0001$), good ADL ability ($\beta = 1.88$, 95% CI = 1.61 to 2.15, $p < 0.0001$, $q < 0.0001$), positive psychosocial status ($\beta = 1.61$, 95% CI = 1.27 to 1.95, $p < 0.0001$, $q < 0.0001$), suitable diet ($\beta = 1.23$, 95% CI = 0.94 to 1.51, $p < 0.0001$, $q < 0.0001$), favorable demographic characteristics ($\beta = 0.87$, 95% CI = 0.47 to 1.27, $p < 0.0001$, q < 0.0001), and active physical exercise ($\beta = 0.27$, 95% CI = 0 to 0.54, $p = 0.0497$, $q = 0.0497$) (Figure 2C).

Similar findings were observed when analyzing MMSE scores, with participants exhibiting favorable domain components scoring higher than those without (all q's ≤ 0.0001). Furthermore, we analyzed the association between domain components and memory performance. Our results also demonstrated the positive impact of domain components on memory function. Notably, diet played an important role for cognitive function ($\beta = 0.28$, 95% CI = 0.23 to 0.33, $p < 0.0001$, $q < 0.0001$), whereas physical activity was the only domain that had a negative effect (Figure 2C).

3.5 | Specific phenotypes of identified domains and factors among populations

In the cluster analysis, four phenotypes were identified as the most optimal to represent the cohort data (Figure S3). These phenotypes, with distinctive patterns of domain features and the summary statistics, are presented in Figure 3. Phenotypes 1 and 2 were considered favorable due to their relatively high protective cognitive exposure domain scores, whereas Phenotypes 3 and 4 were categorized as middle and unfavorable, respectively, based on their relatively lower domain scores. Post hoc analysis of the phenotypes revealed that Phenotype 1, comprising 87 participants (8.03%), was characterized by high SES, active physical activities, and good ADL ability, along with intermediate-high levels of demographic, dietary, and psychosocial status. Phenotype 2 was marked by a high level of diet, high SES, and good ADL ability but low physical activity levels, with intermediate-high levels of demographic and psychosocial status. Nearly half of the participants in Phenotype 3 had good ADL ability but low SES, with moderate levels of demographic, dietary, and psychosocial status. Phenotype 4, which included about 10% of the participants, was characterized by impaired ADL ability and moderate dietary status, along with low levels of demographic, dietary, psychosocial status, and physical activity (Table 1).

When the variables were ranked by the standardized mean difference between phenotypes, participants in Phenotype 1 had higher physical activities, especially in leisure and social contact time, including online chatting, watching e-videos, and internet browsing. Participants in Phenotype 2 had a healthier diet and were more likely to consume green leafy vegetables and fish. In addition, participants in Phenotypes 1 and 2 had higher SES (including higher educational levels, higher household income, and urban residence) compared to those in Phenotypes 3 and 4. Participants in Phenotype 4 exhibited the poorest ADL ability, particularly in managing transportation and using the telephone (Figure 3B). Therefore, four phenotypes were characterized by having high/low levels of distinctive exposure pattern compared with other phenotypes. There were "active leisure/social activities" (Phenotype 1) and "superior diet" (Phenotype 2), whereas people in "low SES" (Phenotype 3) and "unfavorable lifestyle" (Phenotype 4) needed attention and further intervention.

3.6 | Association of phenotypes with cognitive outcomes

The cognitive outcomes for all identified phenotypes were analyzed and are presented in Table 1 and Table 2. In this cohort, "active leisure/social activities" and "superior diet" dominated phenotypes exhibited the best cognitive outcomes with the highest MoCA scores (24.5 to 25), MMSE scores (28), and AVLT scores (0.02 to 0.03). In longitudinal analyses of global cognition scores, compared to the "superior diet" dominated phenotype, participants in the "active leisure/social activities" dominated phenotype did not show differences in MoCA and MMSE scores. However, those with "active leisure/social activities" experienced significantly worse memory function outcomes (ALVT scores, $\beta = -0.18$, 95% CI = −0.34 to −0.03, $p = 0.0175$, $q = 0.0175$). The participants in "low SES" (vs "superior diet") were observed to have significantly lower scores of all cognitive outcome including MoCA scores ($\beta = -4.95$, 95% CI = −5.51 to −4.40, $p < 0.0001$, $q < 0.0001$), MMSE scores ($\beta = -3.33$, 95% CI = −3.81 to −2.86, $p < 0.0001$, $q < 0.0001$), and AVLT scores ($\beta = -0.49$, 95% CI = −0.58 to −0.40, $p < 0.0001$, $q < 0.0001$). The "unfavorable lifestyle" dominated phenotype was associated with the poorest cognitive outcomes, showing the largest deficits in MoCA ($\beta = -10.70$, 95% CI = −11.58 to −9.83, $p < 0.0001$, $q < 0.0001$), MMSE ($\beta = -8.22$, 95% CI = −8.96 to −7.47, $p < 0.0001$, $q < 0.0001$), and AVLT ($\beta = -0.87$, 95% CI = −1.01 to −0.73, $p < 0.0001$, $q < 0.0001$), compared to the "superior diet" dominated phenotype.

3.7 | Monte Carlo simulation for cognitive function improvement

The "active leisure/social activities" and "superior diet" dominated phenotypes had relatively higher domain scores and better cognitive outcomes, offering more cognitive benefits compared to "low SES" and "unfavorable lifestyle." A Monte Carlo simulation was performed to analyze the impact on cognitive function in the total population, assuming that some interventions took effect by varying the proportion of phenotypes during the follow-up period. The results are presented in Figure 4 and Figure S4. In the simulated phenotype distribution, the chance of finding benefit increased as the proportion of participants in "active leisure/social activities" and "superior diet" dominated phenotypes rose. For global cognition, when the proportion of Phenotype 1 increased by 10% (scenario 3), there was a 67.52% chance of observing a benefit, with the average MoCA score of the total population

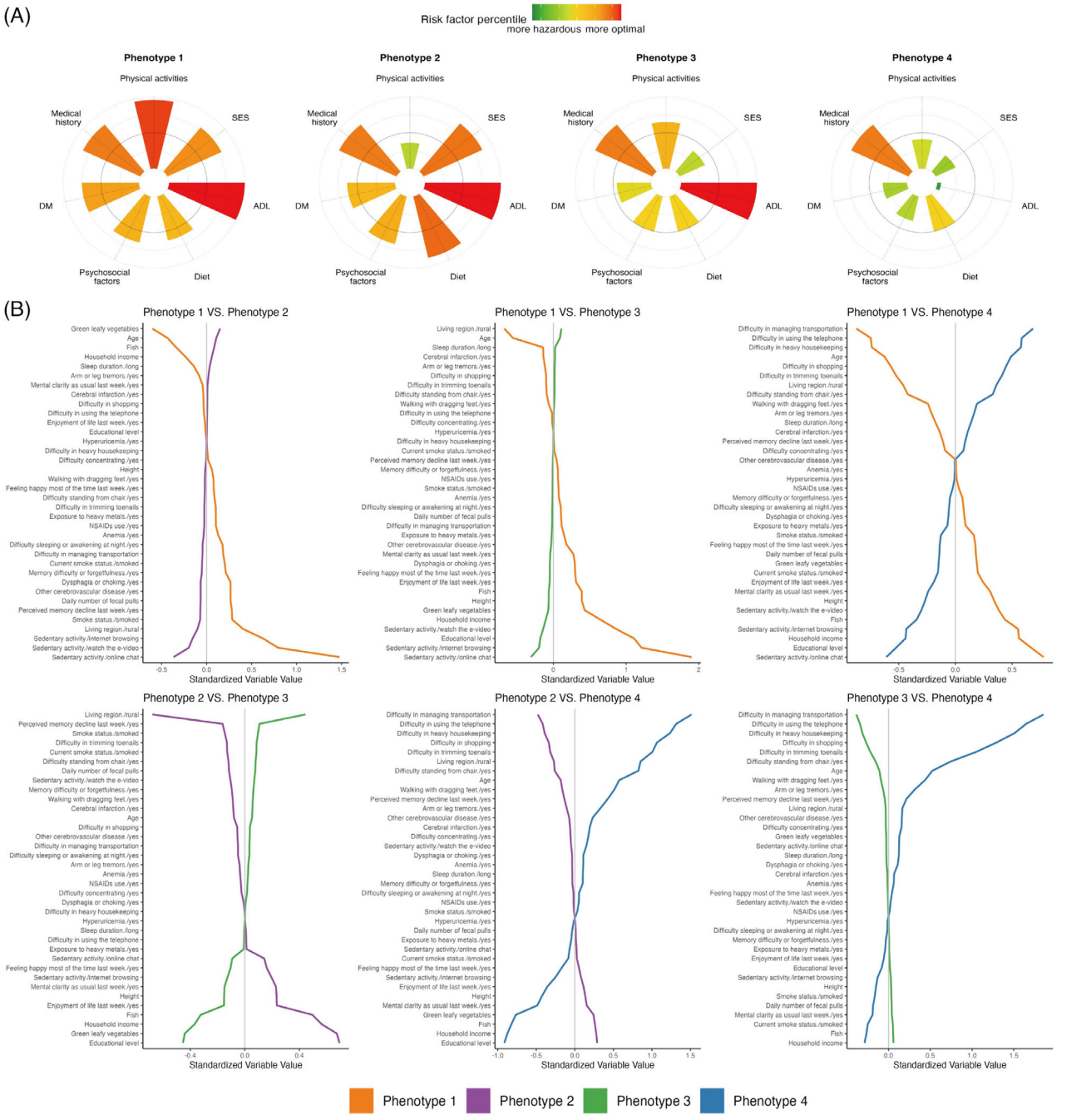


FIGURE 3 The characteristic distribution of domain components by phenotype for the longitudinal cohort. (A) The radar chart shows domain components by phenotype. Each panel corresponds to a phenotype, with each bar showing the median value of a domain factor for all participants in that phenotype. The concentric circles represent the minimum, 25th, 50th, and 75th percentiles, and maximum in the whole sample, with the median shown in a darker color. The height and color of the bars represent the median level of each risk factor, positioned relative to the distribution in the whole population, ensuring a common scale across all phenotypes. (B) The comparison of variables between phenotypes. In all panels, the variables are standardized so that all means are scaled to 0 and SDs to 1. A value of 1 for the standardized variable (x-axis) signifies that the mean value for the phenotype was 1 SD higher than the mean value for both phenotypes shown in the graph as a whole. SD, standard deviation.

rising by 0.97 points compared to the original distribution. When Phenotype 1 increased by one-fourth of the total population (Scenario 4: +25%, proportion = 33%), the probability of benefit reached 100% and the average MoCA score would increase by 1.82. When Phenotype 1 constituted the majority of the population (scenario 5: +50%, proportion = 58%), the average cognitive scores improved by 2.28 points. Similar trends were observed in MMSE and AVLT outcomes, where average scores increased by 1.09 and 0.30 points, respectively, as Phenotype 1 increased to the same extent (+25%). When Phenotype 2 increased by the same proportion (scenario 5: +25%), the chance of finding cognitive benefit also increased to 100%. The average cognitive scores improved by 2.08 points in MoCA, 1.06 points in MMSE,

TABLE 1 Phenotypes characteristics of the studied population.

	Phenotype 1	Phenotype 2	Phenotype 3	Phenotype 4	*p*-value	*q* value[a]
Survey period (%)					<0.0001	<0.0001
Baseline	87 (8.03)	272 (25.12)	611 (56.42)	113 (10.43)		
Follow-up	87 (8.03)	428 (39.52)	459 (42.38)	109 (10.06)		
Cognitive outcome						
MoCA scores	24.5 (20, 28)	25 (20, 29)	19 (14, 23)	12 (8, 16)	<0.0001	<0.0001
MMSE scores	28 (24, 30)	28 (24, 30)	24 (20, 27)	18 (14, 22)	<0.0001	<0.0001
AVLT scores	0.02 (−0.46, 0.71)	−0.03 (−0.53, 0.99)	−0.31 (−0.71, 0.2)	−0.71 (−0.96, −0.29)	<0.0001	<0.0001
Domain component						
SES	0.30 (−0.05, 0.79)	0.78 (0.30, 1.14)	−0.42 (−0.79, −0.06)	−0.78 (−1.14, −0.41)	<0.0001	<0.0001
ADL	0.34 (0.34, 0.34)	0.34 (0.34, 0.34)	0.34 (0.34, 0.34)	−1.94 (−2.76, −1.39)	<0.0001	<0.0001
Diet	0.14 (−0.39, 0.55)	0.97 (0.55, 1.5)	−0.39 (−0.81, 0.14)	−0.39 (−0.81, 0.14)	<0.0001	<0.0001
DM	0.30 (−0.07, 0.74)	0.21 (−0.18, 0.53)	−0.05 (−0.42, 0.28)	−0.29 (−0.72, 0.09)	<0.0001	<0.0001
Psychosocial factors	0.16 (−0.21, 0.74)	0.46 (−0.21, 0.74)	0.07 (−0.78, 0.74)	−0.41 (−0.78, 0.16)	<0.0001	<0.0001
Medical history	0.11 (0.08, 0.11)	0.11 (0.11, 0.11)	0.11 (−0.01, 0.11)	0.11 (−0.36, 0.11)	<0.0001	<0.0001
Physical activities	1.84 (1.49, 2.42)	−0.31 (−0.49, −0.04)	−0.08 (−0.40, 0.01)	−0.09 (−0.40, 0.01)	<0.0001	<0.0001

Abbreviations: ADLs, activities of daily living; AVLT, Auditory Verbal Learning Test; DM, demographic factors; MoCA, Montreal Cognitive Assessment; MMSE, Mini-Mental State Examination; SES, socioeconomic status.
[a]q values were calculated using the false discovery rate (FDR) procedure.

TABLE 2 The longitudinal association of phenotypes with global cognitive function and memory function using linear mixed-effects model.

	MoCA scores			MMSE scores			AVLT scores		
	β (95% CI)	*p*-value	q value[a]	*β* (95% CI)	*p*-value	q value[a]	*β* (95% CI)	*p*-value	*q* value[a]
Phenotype 1	−0.45 (−1.37, 0.48)	0.3458	0.3458	0.1 (−0.69, 0.89)	0.8003	0.8003	−0.18 (−0.34, −0.03)	0.0175	0.0175
Phenotype 2	–	–	–	–	–	–	–	–	–
Phenotype 3	−4.95 (−5.51, −4.40)	<0.0001	<0.0001	−3.33 (−3.81, −2.86)	<0.0001	<0.0001	−0.49 (−0.58, −0.40)	<0.0001	<0.0001
Phenotype 4	−10.70 (−11.58, −9.83)	<0.0001	<0.0001	−8.22 (−8.96, −7.47)	<0.0001	<0.0001	−0.87 (−1.01, −0.73)	<0.0001	<0.0001

Abbreviations: AVLT, Auditory Verbal Learning Test; CI, confidence interval; FDR, false discovery rate; MoCA, Montreal Cognitive Assessment; MMSE, Mini-Mental State Examination.
[a]q values were calculated using the false discovery rate (FDR) procedure.

and 0.48 points in AVLT. Conversely, increasing the proportion of participants in Phenotypes 3 and 4 was significantly detrimental to all cognitive outcomes.

3.8 | Sensitivity analyses

The results of the sensitivity analyses are presented in the Supplementary Material. The first sensitivity analysis was conducted to assess the interaction effects of multidomain factors on cognitive function (Table S7). All domains were found to be positively associated with cognitive scores, other than physical activity for memory. The analysis further revealed several significant negative interactions between domains. For MoCA scores, negative interaction effects were identified between SES and diet ($\beta = -0.51$, 95% CI = −0.83 to −0.18, $p = 0.0021$, $q = 0.0023$), ADLs and physical activities ($\beta = -0.56$, 95% CI = −0.99 to −0.13, $p = 0.0113$, $q = 0.0113$), and diet and physical activities ($\beta = -0.53$, 95% CI = −0.87 to −0.20, $p = 0.0016$, $q = 0.0020$). Similar trends were observed for MMSE scores, with significant negative interactions between SES and diet, ADLs and physical activities, and SES and DM (all p's < 0.05, $q < 0.05$). For AVLT scores, additional interactions between diet and psychosocial factors ($\beta = 0.27$, 95% CI = 0.20 to 0.34, $p < 0.0001$, $q < 0.0001$) and DM and physical activities ($\beta = 0.14$, 95% CI = 0.05 to 0.23, $p = 0.0016$, $q = 0.0021$) were identified. Furthermore, a different strategy for handling missing data was implemented to test the robustness of assumptions regarding the missing data mechanism. The results from the analysis of longitudinal effects between exposure domains, phenotypes, and cognitive function were consistent with findings from datasets imputed using multiple imputation (see Figure S5 and Table S8). Although statistically significant associations between psychosocial status and Phenotype 1 with AVLT scores were not observed ($p > 0.05$), the direction of the associations remained consistent. Simulation trials were reanalyzed, further confirming the robustness of these findings (see Figures S6 and S7).

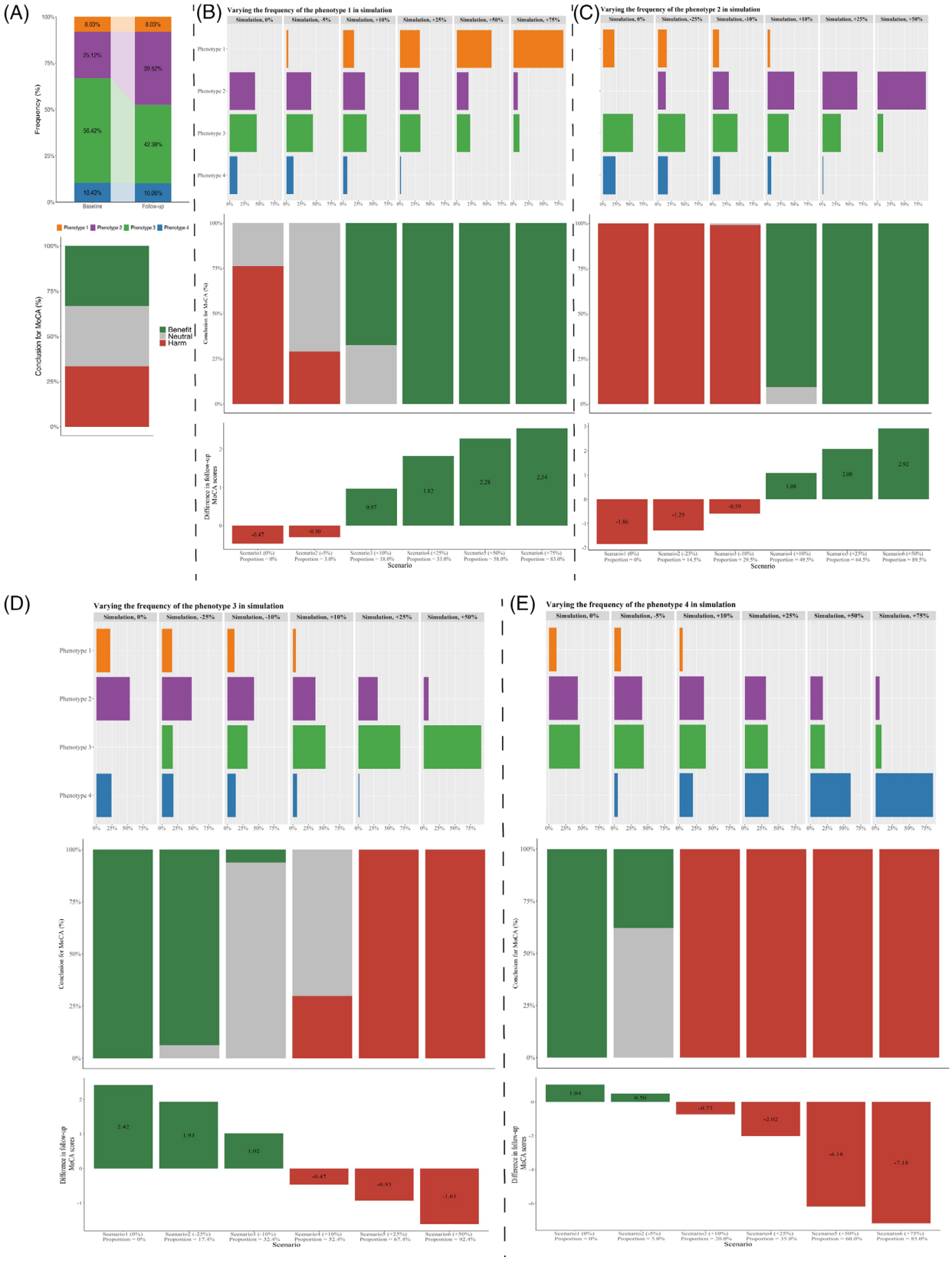


FIGURE 4 Monte Carlo simulation of MoCA score changes with varying phenotypic frequencies in the total population. In this cohort, the actual distribution of the data in the given phenotypes and the legend for harms, benefits, and neutral effects using Monte Carlo simulation analysis are presented in A. Each simulation was conducted with 10,000 iterations using sampling with replacement. The MoCA scores analyzed by changing the phenotype distributions of the data are presented in B (Phenotype 1), C (Phenotype 2), D (Phenotype 3), and E (Phenotype 4). MoCA, Montreal Cognitive Assessment.

4 | DISCUSSION

As cognitive function was influenced by multifactorial characteristics, we conducted an EWAS study design that innovatively incorporated exposure factors that were as comprehensive as possible, covering seven domains: SES, DM, diet, ADLs, medical history, physical activities, and psychosocial factors. The 36 of 207 exposure factors were selected and their relative contributions were further assessed by ML models. The rationality of 36 variables affecting cognitive function has been reported in previous studies (Table S9). Improved profiles in these seven domains independently enhanced global cognitive scores. Moreover, we conducted a cluster analysis to explore the joint phenotypes across domains. By promoting more favorable exposure patterns to dominate the majority of the population, the overall cognitive function of the population will be improved.

Previous studies have concentrated primarily on the individual effects of specific factors, such as diet[20] and sleep duration.[11] Few studies have collected multiple lifestyle factors on cognitive function simultaneously,[47] making it challenging to compare the relative contributions of different factors to cognitive decline. Our cluster analysis indicated that participants who maintained healthier exposure patterns ("active leisure/social activities" and "superior diet" dominated phenotypes) experienced significant cognitive benefits compared to those with less healthy patterns. This information could not only help people develop favorable lifestyles but also provide suggestions for the intervention policy from the perspective of government and society.

SES emerged as the best predictor for cognitive scores, whereas education level was a more critical indicator than other factors within the SES domain (Figure 2A). A range of complex and multifaceted pathways were reported between SES and cognition function. Research indicates that SES is associated with brain structures, which mediate cognitive skills such as executive processing and language skills.[48] In addition, individuals from lower socioeconomic backgrounds are more frequently exposed to social and environmental stressors, including overcrowding, crime, noise pollution, and discrimination. These stressors significantly contribute to negative psychological outcomes.[49] SES also impacts cognitive abilities indirectly through factors such as educational attainment, household spending, and social engagement.[50] Given that SES is more likely to be modified at the community level, it is imperative for governments and public institutions to focus on improving socioeconomic conditions. Especially in lower middle-income areas, the socioeconomic inequalities were more pronounced than in high-income areas. Some socioeconomic interventions are recommended including improving the level of education and implementing resource redistribution measures toward less-advantaged populations (e.g., affordable housing, taxes, and social security).[51]

Exposure factors associated with the ADL domain also played a significant role in cognitive function. Previous studies have found that cognitive decline is closely linked to reductions in daily tasks essential for independent living.[52,53] In line with previous studies,[54] our analysis revealed that the MoCA score had a stronger association with IADL items, particularly in managing transportation, housekeeping, and using the telephone. These performance declines may stem from a reduction in cognitive capacity, limiting both the range of tasks individuals can complete and the skills they can apply to these tasks.

Diet is widely recognized as an important lifestyle intervention for public health. In our analysis, diet significantly contributed to cognitive function; similar to findings by Jia et al.,[33] diet domain acted the strongest association with longitudinal memory change. Consistent with previous studies, our EWAS study indicated that consuming the fish and green leafy vegetables components of the MIND Diet proved more effective in enhancing cognitive function, which is potentially due to the high content of essential nutrients (e.g., n-3 long-chain polyunsaturated fatty acids and low-molecular-weight peptides) and their anti-inflammatory properties.[55–57]

Our results indicated that physical activities had multiple effects on cognitive function. Although higher physical activity scores were associated with better global cognitive function (as measured by MoCA and MMSE), they were also linked to worse memory function. The identified exposure factors (online chat, watch the e-video, and internet browsing) in physical activities are generally considered leisure and social activities, which have been associated with a lower risk of cognitive impairment,[47,58] potentially due to increased interpersonal communication, which enhances social support[59] and self-perceived emotional support.[60] However, the negative association with memory function warrants attention. Activities with an energy expenditure of 1.5 metabolic equivalents while sitting or reclining are classified as sedentary behaviors according to previous studies.[61] The observed memory decline may be attributed to reduced cerebral blood flow[62] and increased levels of inflammatory cytokines[63] caused by prolonged sedentary behaviors.

Psychological factors are increasingly recognized for their impact on cognitive processes.[64] A cohort study in Korea reported that psychological resilience was positively related to global cognitive function.[65] Research has shown that timely implementation of depression intervention strategies can reduce the risk of dementia.[66] Therefore, appropriate psychological interventions may not only improve mental health but also enhance cognitive function.

As the exposure factors and domains were identified, it is necessary to pinpoint that evaluating a single factor's contributions, without considering other exposure factors, may overlook potential joint effects, especially when multiple exposure factors are causally linked. Some potential interaction effects had been identified in multidomain factors, which highlighted the complexity of multidomain influences on cognition. For example, the interaction of SES with other domains was identified, whereas previous studies have reported the joint association between SES and lifestyle factors.[67] According to the nutrition transition hypothesis, increasing wealth is frequently linked to shifts in dietary and physical activity patterns.[68] However, the complex mechanisms underlying the relationships between lifestyle factors, SES, and cognitive function remain insufficiently understood. Moreover, consistent with previous findings,[69] the negative interaction effects were commonly observed in all cognitive outcomes and suggested that stronger cognitive improvements occurred when domain scores were relatively low. This indicated that individuals with lower scores across multiple domains were likely to benefit more from targeted

interventions. These findings underscored the importance of prioritizing interventions for individuals with worse exposure phenotypes to maximize cognitive benefits.

Here, we conducted a cluster analysis to define the joint distribution of exposure domains and the simulation results indicated that participants with favorable combinations of domains experienced greater cognitive benefits. When Phenotypes 1 and 2 increased by 25%, compared to the original scenario, the average population scores increased by 1.82–2.08 points on the MoCA and 1.09–1.06 points on the MMSE. These improvements in cognitive scores are considered clinically significant.[70] These findings provide further evidence supporting the implementation of multidomain interventions for cognitive health by transferring more people with unfavorable phenotypes into healthy exposure patterns.

By examining the differences in characteristics between phenotype pairs (Figure 3B), significant discrepancies were observed between the unfavorable phenotype (Phenotype 4) and the favorable phenotypes (Phenotypes 1 and 2) across domains such as ADLs, SES, diet, and physical activity. Because ADL ability and financial status are challenging to modify, greater emphasis should be placed on factors that are more easily influenced. Practical, multidomain interventions could focus on improving education levels, increasing the consumption of green leafy vegetables and fish, and promoting active leisure and social engagement. In addition, due to the interconnected nature of these factors, interventions targeting modifiable behaviors may also indirectly improve less modifiable ones. For example, cognitive decline in individuals with lower incomes can be attenuated if they have a healthy lifestyle.[71] Similarly, enhancing education levels may build cognitive reserve, making individuals more resilient to the effects of brain aging and related pathologies.[72] Therefore, some precise and targeted interventions were recommended for individuals with unfavorable phenotypes.

This study has several strengths. First, this is the first EWAS study combining ML technology to integrate exposure information on such a large scale and systematically and comprehensively identify the exposure factors related to cognitive function. Second, the diversity of variables also allowed us to identify relatively complete exposure patterns and better explore the individual and joint contributions of the multiplexed exposure factors to cognitive function. Finally, this cohort study captures the longitudinal fluctuation of cognitive function and exposure domains that further infer the cognitive benefits for the population through simulation trials.

However, there are still some limitations. First, the assessments of factors in this study were based on self-reports, and genetic information was not available. Nonetheless, a previous cohort[36] and cross-sectional[73] study found that the association of lifestyle profiles with dementia and cognitive function remained largely unchanged regardless of whether genetic risk factors were included. This may be because the variance attributed to genetic factors was largely accounted for by other exposure domains.[2] Second, the relatively short follow-up period limited our ability to explore cognitive-related clinical endpoints, thus longitudinal studies with long follow-ups are recommended. However, multiple cognitive tests, including MoCA, MMSE, and AVLT, are widely recognized as effective tools for assessing cognition. We utilized these tests to represent cognitive conditions to the greatest extent possible. Third, due to the data availability, the scope of the exposure domain may not have been fully represented by existing exposure factors. Further EWAS studies collecting more comprehensive exposure information should be conducted to facilitate a more extensive investigation. Fourth, although the exposure variables were derived from objective and quantifiable components of the questionnaire, the biases arising from the biases of self-reports and questionnaires can lead to exposure misclassification and distort the association between exposure and outcomes in the study. More objective assessment methods, such as physical activity measured by accelerometers,[74] should be applied in future studies. Fifth, simulation trials are conducted using existing data, and their predictive accuracy is limited by potential unmeasured confounding factors and the complex dynamics of phenotype transitions. More comprehensive data collection and developing advanced models should be considered to address these limitations.

In conclusion, cognitive function was associated with multidomain exposure factors. The SES and diet domains were found to have the strongest association with longitudinal cognitive function and memory change. The phenotype analysis and simulation trials highlight that the priority of interventions for subpopulations with the unfavorable exposure phenotype could bring greater cognitive benefits, especially when health costs were considered.

ACKNOWLEDGMENTS

We sincerely appreciate the participants of the Community Cohort Study of Nervous System Diseases. This project was supported by the National Key R&D Program of China, Precision Medicine Project–Cohort Study on Nervous System Diseases (No. 2017YFC0907700); Community-based Cohort Study on Nervous System Diseases (No. 2017YFC0907701); Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0513200).

CONFLICT OF INTEREST STATEMENT

The authors declare they have no actual or potential competing financial interests. Author disclosures are available in the Supporting Information.

CONSENT STATEMENT

All subjects provided written informed consent.

ORCID

Jingtao Wu https://orcid.org/0000-0002-0634-3894

REFERENCES

1. Mitchell AJ, Kemp S, Benito-León J, Reuber M. The influence of cognitive impairment on health-related quality of life in neurological disease. *Acta Neuropsychiatr*. 2010;22:2-13. doi:10.1111/j.1601-5215.2009.00439.x
2. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396:413-446. doi:10.1016/S0140-6736(20)30367-6

3. Shen X, Wang C, Zhou X, et al. Nonlinear dynamics of multi-omics profiles during human aging. *Nat Aging*. 2024;4:1619-1634. doi:10.1038/s43587-024-00692-2
4. Liu L, Saba A, Pascual JR, et al. Lecanemab and vascular-amyloid deposition in brains of people with Down syndrome. *JAMA Neurol*. 2024;81:1066-1072.doi:10.1001/jamaneurol.2024.2579
5. Mahase E. Lecanemab: European drug agency rejects Alzheimer's drug amid debate over efficacy and safety. *BMJ*. 2024;386:q1692. doi:10.1136/bmj.q1692
6. Reardon S. FDA approves Alzheimer's drug lecanemab amid safety concerns. *Nature*. 2023;613:227-228. doi:10.1038/d41586-023-00030-3
7. Sha F, Zhao Z, Wei C, Li B. Modifiable factors associated with reversion from mild cognitive impairment to cognitively normal status: a prospective cohort study. *J Alzheimers Dis*. 2022;86:1897-1906. doi:10.3233/JAD-215677
8. Sachdev PS, Lipnicki DM, Crawford J, et al. Factors predicting reversion from mild cognitive impairment to normal cognitive functioning: a population-based study. *PLoS One*. 2013;8:e59649. doi:10.1371/journal.pone.0059649
9. Ding E, Deng F, Fang J, et al. Exposome-wide ranking to uncover environmental chemicals associated with dyslipidemia: a panel study in healthy older Chinese adults from the BAPE study. *Environ Health Perspect*. 2024;132:097005. doi:10.1289/EHP13864
10. Wang Y, Tang S, Hou T, Cong L, Du Y, Qiu C. Multimodal interventions to delay dementia and disability in rural China (MIND-China): an update 2018-2021. *Alzheimers Dementia*. 2022;18:e061546. doi:10.1002/alz.061546
11. Chen W-C, Wang X-Y. Longitudinal associations between sleep duration and cognitive impairment in Chinese elderly. *Front Aging Neurosci*. 2022;14:1037650. doi:10.3389/fnagi.2022.1037650
12. Andrikopoulos P. Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine N-oxide. *Nat Commun*. 2023;14(1):5843.
13. Graham SA, Lee EE, Jeste DV, et al. Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: a conceptual review. *Psychiatry Res*. 2020;284:112732. doi:10.1016/j.psychres.2019.112732
14. Choi KW, Stein MB, Nishimi KM, et al. An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *AJP*. 2020;177:944-954. doi:10.1176/appi.ajp.2020.19111158
15. Bloomberg M, Brocklebank L, Hamer M, Steptoe A. Joint associations of physical activity and sleep duration with cognitive ageing: longitudinal analysis of an English cohort study. *Lancet Healthy Longev*. 2023;4:e345-53. doi:10.1016/S2666-7568(23)00083-1
16. Pan XF, Wang L, Pan A. Epidemiology and determinants of obesity in China. *Lancet Diabetes Endocrinol*. 2021;9:e2. doi:10.1016/S2213-8587(21)00045-0
17. Zhang S, Wang L, Jia X, et al. A comparison between dietary consumption status and healthy dietary pattern among adults aged 55 and older in China. *Nutrients*. 2022;14:2778. doi:10.3390/nu14132778
18. Barnes LL, Dhana K, Liu X, et al. Trial of the MIND diet for prevention of cognitive decline in older persons. *N Engl J Med*. 2023;389:602-611. doi:10.1056/NEJMoa2302368
19. Barnes JL, Tian M, Edens NK, Morris MC. Consideration of nutrient levels in studies of cognitive decline. *Nutr Rev*. 2014;72:707-719. doi:10.1111/nure.12144
20. Morris MC, Tangney CC, Wang Y, et al. MIND diet slows cognitive decline with aging. *Alzheimers Dement*. 2015;11:1015-1022. doi:10.1016/j.jalz.2015.04.011
21. Tangney CC, Li H, Wang Y, et al. Relation of DASH- and Mediterranean-like dietary patterns to cognitive decline in older persons. *Neurology*. 2014;83:1410-1416. doi:10.1212/WNL.0000000000000884
22. Zhang M, Yu E, He Y. Tools for dementia epidemiological investigations and their applications. Shanghai Arch. *Psychiatry*. 1995;7:1-62.
23. Su X, Shang L, Xu Q, et al. Prevalence and predictors of mild cognitive impairment in Xi'an: a community-based study among the elders. *PLoS One*. 2014;9:e83217. doi:10.1371/journal.pone.0083217
24. Chen X, Pan X. *Neurological examination and common scales Quick Reference(Chinese Edition)*. Chemical Industries Press; 2013.
25. Jia J-P, Wang Y-H, Wei C-B, et al. Chinese guidelines for diagnosis and management of cognitive impairment and dementia (V): dementia therapy. *Chin Med J*. 2011;91:940-945. doi:10.3760/cma.j.issn.0376-2491.2011.14.002
26. Zhou J, Britigan DH, Rajaram SS, Wang H, Su D. Association between leisure time physical activity preference and behavior: evidence from the China Health & Nutrition Survey, 2004-2011. *BMC Public Health*. 2017;17:451. doi:10.1186/s12889-017-4386-5
27. Wang J, Li R, Zhang L, et al. Associations between sedentary behaviour patterns and depression among people aged 60 and older in Hebei Province of China. *BMC Public Health*. 2022;22:283. doi:10.1186/s12889-022-12727-7
28. Liang Y, Li X, Yang T, et al. Patterns of physical activity and their relationship with depression among community-dwelling older adults in Shanghai, China: a latent class approach. *BMC Geriatr*. 2021;21:587. doi:10.1186/s12877-021-02537-8
29. Huang F, Wang H, Wang Z, et al. Is geriatric depression scale a valid instrument to screen depression in Chinese community-dwelling elderly?. *BMC Geriatr*. 2021;21:310. doi:10.1186/s12877-021-02266-y
30. Lu J, Li D, Li F, et al. Montreal cognitive assessment in detecting cognitive impairment in Chinese elderly individuals: a population-based study. *J Geriatr Psychiatry Neurol*. 2011;24:184-190. doi:10.1177/0891988711422528
31. Katzman R, Zhang MY, Ouang-Ya-Qu N, et al. A Chinese version of the Mini-Mental State Examination; impact of illiteracy in a Shanghai dementia survey. *J Clin Epidemiol*. 1988;41:971-978. doi:10.1016/0895-4356(88)90034-0
32. Maj M, Satz P, Janssen R, et al. WHO Neuropsychiatric AIDS study, cross-sectional phase II. Neuropsychological and neurological findings. *Arch Gen Psychiatry*. 1994;51:51-61. doi:10.1001/archpsyc.1994.03950010051007
33. Jia J, Zhao T, Liu Z, et al. Association between healthy lifestyle and memory decline in older adults: 10 year, population based, prospective cohort study. *BMJ*. 2023;380:e072691. doi:10.1136/bmj-2022-072691
34. Koetsier J, Cavill R, Reijnders R, et al. Blood-based multivariate methylation risk score for cognitive impairment and dementia. *Alzheimers Dementia*. 2024;20:6682-6698. doi:10.1002/alz.14061
35. Pinto TCC, Machado L, Bulgacov TM, et al. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's disease (AD) in the elderly? *Int Psychogeriatr*. 2019;31:491-504. doi:10.1017/S1041610218001370
36. Zhang Y, Chen S-D, Deng Y-T, et al. Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank. *Nat Hum Behav*. 2023;7:1185-1195. doi:10.1038/s41562-023-01585-x
37. Asgari S, Khalili D, Zayeri F, Azizi F, Hadaegh F. Dynamic prediction models improved the risk classification of type 2 diabetes compared with classical static models. *J Clin Epidemiol*. 2021;140:33-43. doi:10.1016/j.jclinepi.2021.08.026
38. Chen T, Guestrin C, Xgboost: a scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794. Association for Computing Machinery (ACM); 2016.

39. Ryu S-E, Shin D-H, Chung K. Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization. *IEEE Access*. 2020;8:177708-177720. doi:10.1109/ACCESS.2020.3025553
40. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:1-10.
41. Christensen GM, Li Z, Pearce J, et al. The complex relationship of air pollution and neighborhood socioeconomic status and their association with cognitive decline. *Environ Int*. 2022;167:107416. doi:10.1016/j.envint.2022.107416
42. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Stat Methodol*. 1996;58:267-288.
43. Zhang Y, Guo L, Dai Q, et al. A signature for pan-cancer prognosis based on neutrophil extracellular traps. *J Immunother Cancer*. 2022;10:e004210. doi:10.1136/jitc-2021-004210
44. Lourida I, Hannon E, Littlejohns TJ, et al. Association of lifestyle and genetic risk with incidence of dementia. *JAMA*. 2019;322:430-437. doi:10.1001/jama.2019.9879
45. Lhoste VPF, Zhou B, Mishra A, et al. Cardiometabolic and renal phenotypes and transitions in the United States population. *Nat Cardiovasc Res*. 2023;3:46-59. doi:10.1038/s44161-023-00391-y
46. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA*. 2019;321:2003-2017. doi:10.1001/jama.2019.5791
47. Ran W, Yu Q. Data-driven clustering approach to identify novel clusters of high cognitive impairment risk among Chinese community-dwelling elderly people with normal cognition: a national cohort study. *J Glob Health*. 2024;14:04088. doi:10.7189/jogh.14.04088
48. Thomas MSC, Coecke S. Associations between socioeconomic status, cognition, and brain structure: evaluating potential causal pathways through mechanistic models of development. *Cogn Sci*. 2023;47:e13217. doi:10.1111/cogs.13217
49. Baum A, Garofalo JP, Yali AM. Socioeconomic status and chronic stress: does stress account for SES effects on health?. *Ann NY Acad Sci*. 1999;896:131-144. doi:10.1111/j.1749-6632.1999.tb08111.x
50. Ye X, Zhu D, He P. Direct and indirect associations between childhood socioeconomic status and cognitive function in the middle-aged and older adults in China. *Aging Ment Health*. 2022;26:1730-1737. doi:10.1080/13607863.2021.1935459
51. Ni Y, Zhou Y, Kivimäki M, et al. Socioeconomic inequalities in physical, psychological, and cognitive multimorbidity in middle-aged and older adults in 33 countries: a cross-sectional study. *Lancet Healthy Longev*. 2023;4:e618-28. doi:10.1016/S2666-7568(23)00195-2
52. Tucker-Drob EM. Neurocognitive functions and everyday functions change together in old age. *Neuropsychology*. 2011;25:368-377. doi:10.1037/a0022348
53. Allaire JC, Marsiske M. Well- and ill-defined measures of everyday cognition: relationship to older adults' intellectual ability and functional status. *Psychol Aging*. 2002;17:101-115. doi:10.1037/0882-7974.17.1.101
54. Armstrong M, Barr E, Gruber-Baldini A, et al. Association of cognitive screening tests with disability in Parkinson's disease (P5.259). *Neurology*. 2014;82:P5259. doi:10.1212/WNL.82.10_supplement.P5.259
55. Chataigner M, Mortessagne P, Lucas C, et al. Dietary fish hydrolysate supplementation containing n-3 LC-PUFAs and peptides prevents short-term memory and stress response deficits in aged mice. *Brain Behav Immun*. 2021;91:716-730. doi:10.1016/j.bbi.2020.09.022
56. Yang YJ, Yu A, Kim J, et al. Fish consumption and cognitive function in older adults. *FASEB J*. 2013;27:lb389-lb389. doi:10.1096/fasebj.27.1_supplement.lb389
57. Morris MC, Wang Y, Barnes LL, Bennett DA, Dawson-Hughes B, Booth SL. Nutrients and bioactives in green leafy vegetables and cognitive decline: prospective study. *Neurology*. 2018;90:e214-22. doi:10.1212/WNL.0000000000004815
58. Fangfang H, Xiao H, Shuai Z, et al. Living environment, built environment and cognitive function among older chinese adults: results from a cross-sectional study. *J Prev Alzheimers Dis*. 2022;9:126-135. doi:10.14283/jpad.2021.59
59. Haslam DM, Tee A, Baker S. The use of social media as a mechanism of social support in parents. *J Child Fam Stud*. 2017;26:2026-2037. doi:10.1007/s10826-017-0716-6
60. Shensa A, Sidani JE, Lin LY, Bowman ND, Primack BA. Social media use and perceived emotional support among US young adults. *J Community Health*. 2016;41:541-549. doi:10.1007/s10900-015-0128-8
61. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary". *Exerc Sport Sci Rev*. 2008;36:173-178. doi:10.1097/JES.0b013e3181877d1a
62. Carter SE, Draijer R, Holder SM, Brown L, Thijssen DHJ, Hopkins ND. Regular walking breaks prevent the decline in cerebral blood flow associated with prolonged sitting. *J Appl Physiol (1985)*. 2018;125:790-798. doi:10.1152/japplphysiol.00310.2018
63. Wirth K, Klenk J, Brefka S, et al. Biomarkers associated with sedentary behaviour in older adults: a systematic review. *Ageing Res Rev*. 2017;35:87-111. doi:10.1016/j.arr.2016.12.002
64. Farhanginasab F, Lajevardi L, Taghizadeh G. The relationship between psychological factors and cognitive function in patients with Parkinson's disease who have chronic fatigue. *Func Disabil J*. 2021;4:E39. doi:10.32598/fdj.4.39
65. Jung SJ, Lee GB, Nishimi K, Chibnik L, Koenen KC, Kim HC. Association between psychological resilience and cognitive function in older adults: effect modification by inflammatory status. *GeroScience*. 2021;43:2749-2760. doi:10.1007/s11357-021-00406-1
66. Yang L, Deng Y-T, Leng Y, et al. Depression, depression treatments, and risk of incident dementia: a prospective cohort study of 354,313 participants. *Biol Psychiatry*. 2023;93:802-809. doi:10.1016/j.biopsych.2022.08.026
67. Li R, Li R, Xie J, et al. Associations of socioeconomic status and healthy lifestyle with incident early-onset and late-onset dementia: a prospective cohort study. *Lancet Healthy Longev*. 2023;4:e693-702. doi:10.1016/S2666-7568(23)00211-8
68. Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. *Int J Obes*. 2004;28:S2-S9. doi:10.1038/sj.ijo.0802804
69. Parrott MD, Shatenstein B, Ferland G, et al. Relationship between diet quality and cognition depends on socioeconomic position in healthy older adults. *J Nutr*. 2013;143:1767-1773. doi:10.3945/jn.113.181115
70. Muir RT, Hill MD, Black SE, Smith EE. Minimal clinically important difference in Alzheimer's disease: rapid review. *Alzheimers Dement*. 2024;20:3352-3363. doi:10.1002/alz.13770
71. Weng P-H, Chen J-H, Chiou J-M, et al. The effect of lifestyle on late-life cognitive change under different socioeconomic status. *PLoS One*. 2018;13:e0197676. doi:10.1371/journal.pone.0197676
72. Lövdén M, Fratiglioni L, Glymour MM, Lindenberger U, Tucker-Drob EM. Education and cognitive functioning across the life span. *Psychol Sci Public Interest*. 2020;21:6-41. doi:10.1177/1529100620920576
73. Jin X, He W, Zhang Y, et al. Association of APOE ε4 genotype and lifestyle with cognitive function among Chinese adults aged 80 years and older: a cross-sectional study. *PLoS Med*. 2021;18:e1003597. doi:10.1371/journal.pmed.1003597
74. Cao Z, Min J, Chen H, et al. Accelerometer-derived physical activity and mortality in individuals with type 2 diabetes. *Nat Commun*. 2024;15:5164. doi:10.1038/s41467-024-49542-0

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wu J, Yin B, Wen R, et al. A systematic exposure-wide framework leveraging machine learning to identify multidomain exposure factors and their joint influence on cognitive function: Evidence from a neurological cohort. *Alzheimer's Dement*. 2025;21:e14624.
https://doi.org/10.1002/alz.14624