目录

Heliyon 9 (2023) e18758

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

A dynamic machine learning model for prediction of NAFLD in a health checkup population: A longitudinal study

Yuhan Deng [a,b], Yuan Ma [c], Jingzhu Fu [d,e,f], Xiaona Wang [g], Canqing Yu [d,e,f,h], Jun Lv [d,e,f,h], Sailimai Man [b,d,e,f,***], Bo Wang [b,e,h,*], Liming Li [d,e,f,h,**]

[a] Chongqing Research Institute of Big Data, Peking University, Chongqing, China
[b] Meinian Institute of Health, Beijing, China
[c] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
[d] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China
[e] Peking University Health Science Center Meinian Public Health Institute, Beijing, China
[f] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
[g] MJ Health Screening Center, Beijing, China
[h] Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing, China

ARTICLE INFO

Keywords:
Machine learning
Dynamic prediction
Time-series data
Checkup records
NAFLD

ABSTRACT

Background: Non-alcoholic fatty liver disease (NAFLD) is one of the most common liver diseases worldwide. Currently, most NAFLD prediction models are diagnostic models based on cross-sectional data, which failed to provide early identification or clarify causal relationships. We aimed to use time-series deep learning models with longitudinal health checkup records to predict the onset of NAFLD in the future, and update the model stepwise by incorporating new checkup records to achieve dynamic prediction.

Methods: 10,493 participants with over 6 health checkup records from Beijing MJ Health Screening Center were included to conduct a retrospective cohort study, in which the constantly updated initial 5 checkup data were incorporated stepwise to predict the risk of NAFLD at and after their sixth health checkups. A total of 33 variables were considered, consisting of demographic characteristics, medical history, lifestyle, physical examinations, and laboratory tests. L1-penalized logistic regression (LR) was used for feature selection. The long short-term memory (LSTM) algorithm was introduced for model development, and five-fold cross-validation was conducted to tune and choose optimal hyperparameters. Both internal validation and external validation were conducted, using the 20% randomly divided holdout test dataset and previously unseen data from Shanghai MJ Health Screening Center, respectively, to evaluate model performance. The evaluation metrics included area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, Brier score, and decision curve. Bootstrap sampling was implemented to generate 95% confidence intervals of all the metrics. Finally, the Shapley additive explanations (SHAP) algorithm was applied in the holdout test dataset for model interpretability to obtain time-specific and sample-specific contributions of each feature.

Results: Among the 10,493 participants, 1662 (15.84%) were diagnosed with NAFLD at and after their sixth health checkups. The predictive performance of the deep learning model in the internal

* Corresponding author. Meinian Institute of Health, Beijing, China.
** Corresponding author. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China
*** Corresponding author. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China
E-mail addresses: sailimai.man@meinianresearch.com (S. Man), paul@meinianresearch.com (B. Wang), lmlee@bjmu.edu.cn (L. Li).

https://doi.org/10.1016/j.heliyon.2023.e18758
Received 13 July 2023; Received in revised form 25 July 2023; Accepted 26 July 2023
Available online 27 July 2023

validation dataset improved over the incorporation of the checkups, with AUROC increasing from 0.729 (95% CI: 0.698,0.760) at baseline to 0.818 (95% CI: 0.798,0.844) when consecutive 5 checkups were included. The external validation dataset, containing 1728 participants, was used to verify the results, in which AUROC increased from 0.700 (95% CI: 0.657,0.740) with only the first checkups to 0.792 (95% CI: 0.758,0.825) with all five. The results of feature significance showed that body fat percentage, alanine transaminase (ALT), and uric acid owned the greatest impact on the outcome, time-specific, individual-specific and dynamic feature contributions were also produced for model interpretability.

Conclusion: A dynamic prediction model was successfully established in our study, and the prediction capability kept improving with the renewal of the latest checkup records. In addition, we identified key features associated with the onset of NAFLD, making it possible to optimize the prevention and control strategies of the disease in the general population.

1. Introduction

As one of the most common chronic hepatic diseases worldwide, non-alcoholic fatty liver disease (NAFLD) has affected 29.6% of the Asian population and the prevalence rate is increasing significantly over time [1]. NAFLD can transit to hepatic inflammation and fibrosis, and is highly associated with other non-liver-specific diseases, especially cardiovascular and metabolic disorders [2], presenting severe healthcare burdens globally [3]. However, given the multifactorial and intricate etiology of the disease, it's still difficult to determine a specific prevention strategy and achieve early identification of high-risk groups to reduce the prevalence of the disease. Therefore, improved prediction of the risk of NAFLD may be of great value in the prevention and control of the disease in the general population.

Artificial intelligence methods, together with massive data collected in the medical field, make it possible to precisely predict the risk of NAFLD. As expected, more and more research has been conducted to solve such issues. However, previous related studies were mostly cross-sectional and mainly focused on the development of diagnostic prediction models, which failed to determine causal relationships or provide early risk probabilities sometime before the confirmed diagnosis of NAFLD [4–6]. Besides, most existing studies were based on conventional machine learning models with a single measurement of variables [5,7,8], ignoring the variation tendency contained in multiple measurements of data.

To date, the valuable information contained in health checkup data, characterized by annually repeat-measured items, is underutilized, while the commonly used machine learning models, such as random forest and XGBoost, may be unable to handle these kinds of time-series data well [9]. Presently, deep learning-based models [10], including the recurrent neural network (RNN) and its derived models, can make full use of the constantly updated records and have been proven to show good performance, but almost in ICU settings [11]. However, despite the promising prediction performance, the complex structures and large amounts of parameters in deep learning models restrict their acceptance and practical application [12,13], while the introduction of model interpretability algorithms largely solved this problem and facilitate the understanding of the black boxes. Among all the algorithms, Shapley additive explanations (SHAP) algorithm [14], a widely used model interpretability method for any machine learning models, seems both effective and prospective in quantifying the impact of each variable on the outcomes [15,16].

To the best of our knowledge, no studies have ever used constantly updated time-series health checkup data to develop a dynamic prediction model based on deep learning algorithms to predict NAFLD. In this study, we use the long short-term memory (LSTM) model with SHAP interpretability algorithms to predict the risk of NAFLD one year after health checkups and update the prediction model

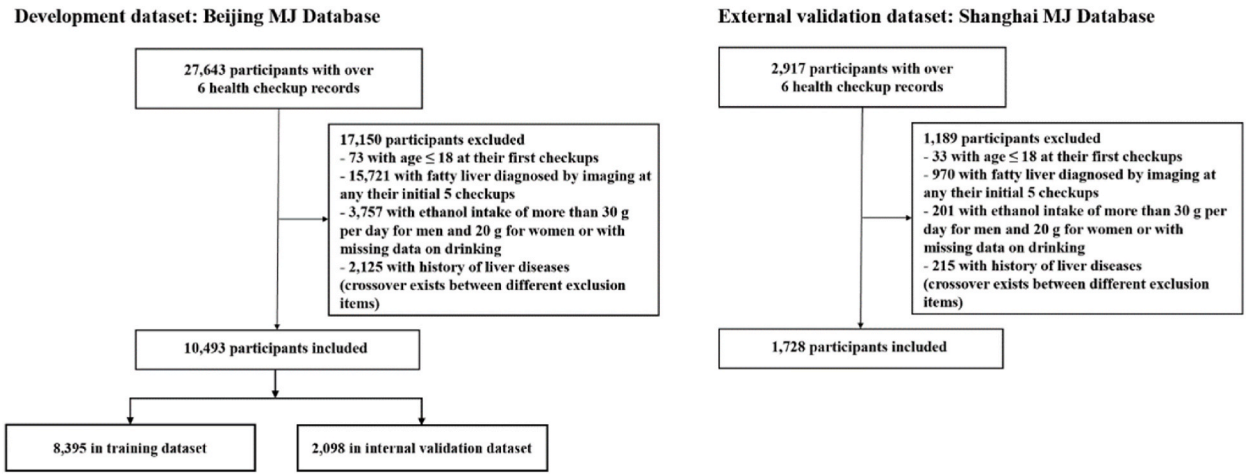


Fig. 1. Inclusion flow chart of study participants.

over checkups, and further verify the results on an external validation dataset.

2. Method

2.1. Data source and study participants

A retrospective cohort study was conducted based on data from Beijing MJ Health Screening Center. Beijing MJ Health Checkup Database contains routinely annual checkup data from 2003 to 2021, with more than 50,000 checkup records collected every year. During each checkup, a series of standardized protocols including physical examinations, laboratory tests, imaging diagnoses and a questionnaire survey about demographic characteristics, dietary habits, physical activity and other comprehensive lifestyle and health conditions were applied and inquired to all the participants. The external validation dataset was obtained from Shanghai MJ Health Screening Center, a health checkup center independent of that in Beijing, containing checkup records from 2003 to 2021.

In our study, participants who attended more than 6 checkups were included. The exclusion criteria were as follows: (1) age under 18 years at the first checkup; (2) having been diagnosed with fatty liver on liver ultrasound at any of the initial five checkups; (3) having ethanol intake of more than 30 g per day for men and 20 g for women or with missing data on drinking at any of the initial five checkups; (4) having a self-reported history of hepatic diseases at any of the initial five checkups. After excluding those who meet the exclusion criteria, a total number of 10,493 participants were included for model development and 1728 for model validation in our study (Fig. 1). The median time intervals between two consecutive checkups of the participants were presented in Supplemental F. 1.

2.2. Predictors and outcomes

The following variables at each of the initial 5 visits were extracted: (i) demographic characteristics: sex, age, education, income; (ii) physical examinations: body mass index (BMI), waist circumference (WC), body fat percentage (BFP), systolic blood pressure (SBP), diastolic blood pressure (DBP); (iii) laboratory tests: fasting blood glucose (FBG), total cholesterol (TC), triglyceride (TG), high-density lipoprotein cholesterol (HDLC), low-density lipoprotein cholesterol (LDLC), alanine transaminase (ALT), aspartate aminotransferase (AST), direct bilirubin (DB), total bilirubin (TB), total protein (TP), albumin (ALB), globulin (GLB), glutamyl transferase (GGT), alkaline phosphatase (ALP), lactate dehydrogenase (LD), uric acid (UA); (iv) smoking status; (v) dietary habits: dietary regularity, fruit intake, dairy intake, fried food intake; (vi) medical history: hypertension, diabetes and hyperlipidemia.

The consumption of food was classified into three tiers of low, moderate, and high, delineated by the following specifics: the low fruit intake is defined as either no consumption or less than 150 g per day, moderate intake is defined as consuming 150 g–300 g fruit per day, and high intake is defined as consuming more than 300 g fruit per day; the low intake of dairy products is defined as no consumption or less than one cup of milk (240 ml) or one serving of dairy products (30 g of cheese or 1 slice of cheese) per week, moderate intake is defined as consuming 1–3 cups of milk or 1–3 servings of dairy products per week, and high intake is defined as consuming more than four cups of milk or more than four servings of dairy products per week; a low level of intake of fried food indicates either no consumption or consumption of less than one serving per week, a moderate level of intake indicates consumption of 1–3 servings per week, and a high level of intake indicates consumption of four or more servings per week, with one serving defined as half a bowl.

The outcome was defined as whether the participant was diagnosed with NAFLD on ultrasonography from their sixth health checkups until the end of their follow-ups.

2.3. Data preprocessing and statistical analysis

Participants in the development dataset from Beijing MJ database were randomly divided into a training set (80%) and an internal validation dataset (20%), and imputations of missing values were conducted in the two datasets, together with the external validation dataset, respectively. Except for certain variables (sex for example) that were constant at each checkup, most of the variables were time-series forms, which means that they were measured repetitively at each checkup and not independent of each other, so the last observation carried forward (LOCF) method was conducted to impute missing values for repeated measurements. After the first imputation, variables with missing rates over 30% were excluded from our analysis, including TP and ALB. For continuous variables, their means were used to impute the rest missingness. Outliers were identified as values distributed more or less than three standard deviations from the mean and were handled the same way as missing values. All continuous variables were maintained as their original forms in case of information loss, and categorical variables with multiple levels were re-classified, then the unordered were one-hot encoded, while the ordered were kept as their reclassification forms. Continuous variables were presented as means with standard deviations (SDs) and utilized Student's t-test or Wilcoxon rank-sum test for statistical analysis. Categorical variables were presented as counts and percentages, and the Chi-square test was employed for comparison.

After the preprocessing step, L1-penalized logistic regression (LR) was conducted based on the newest records in the training set to select a subset of the total features generated from the preprocessing procedure, then the LSTM algorithm was introduced to develop the prediction model, and the model was updated as the number of visit times increased and new measurements of predictors entered. The hyperparameters were selected through a five-fold cross-validation process, in which all combinations of hyperparameters were tested and the optimal combination was chosen based on the mean area under the receiver operating curve (AUROC) across five validation sets derived from the training dataset. Besides, considering the imbalanced issue of the dataset [17], in which participants with NAFLD were less than those without, we kept all the samples in the minority class in each epoch and randomly selected the same

number of samples from the majority class to train the model, and both balanced datasets and imbalanced datasets were tested through five-fold cross-validation. In addition, to establish a benchmark for the prediction of NAFLD, LR was conducted using data based on each health checkup.

In the internal validation dataset and external validation dataset, discrimination and calibration of the prediction models were assessed, in which AUROC, sensitivity, and specificity were used as metrics of discrimination ability, while Brier score was used to evaluate calibration capability. Besides, Platt scaling method [18] was used to transform outputs into risk probabilities by fitting LR to improve calibration, and the Brier score was calculated after recalibration. Bootstrap sampling was performed to generate 95% confidence intervals (CIs) of the metrics above. Specifically, the same size of samples as the validation dataset was randomly selected with replacement 1000 times, and the distribution of the results was used to obtain 95% CIs. Meanwhile, decision-curve characteristic on the internal validation dataset was evaluated. The Delong test was utilized to compare the consecutive AUCs resulting from model updates. Furthermore, as predictive performance with model updates was not independent, we employed repeated measures analysis of variance (ANOVA) to compare the overall performance between models based on LSTM algorithms and LR across all five time points. After the comparison of the updated models, the best-performed model was selected to show feature significance. The SHAP algorithm was applied in the internal validation dataset to enhance models' interpretability by obtaining patient-specific contributions of each feature at a specific time point.

All the data preprocessing was conducted using SAS 9.4 and statistical analyses using Python 3.7.

2.4. LSTM for model development

LSTM [19] is introduced on the basis of the RNN, which can tackle time-series problems and selectively retain important information through its gating systems: input gate (I_t), forget gate (F_t) and output gate (O_t). Specifically, the input gate determines how much information to input at the present step, the forget gate controls how much information to throw away from the previous steps, and the output gate decided how much current information to output. After the disposal conducted by the gate systems, redundant information is filtered out, largely releasing the memory of the hidden layer. The equations and diagram of an LSTM neuron (Fig. 2) are as below:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$$

$$F_t = \sigma\left(X_t W_{xf} + H_{t-1} W_{hf} + b_f\right)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

$$\widetilde{C}_t = \tanh\ (X_t W_{xc} + H_{t-1} W_{hc} + b_c)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \widetilde{C}_t$$

$$H_t = O_t \odot tanh(C_t)$$

In our study, the NAFLD risk prediction was updated timely by incorporating new checkup records to construct dynamic models, and all these longitudinal data integrated into the models were used for the prediction of NAFLD in future checkups.

2.5. SHAP for model interpretability

In the SHAP algorithm [14], feature importance is presented according to Shapley values. For a single sample, a probability is a certain output from a prediction model, indicating the outcome risk of this sample, and in the same way, we can get a mean probability of all samples. Through the SHAP algorithm, each of the current features of this sample corresponds to a Shapley value, which indicates the contribution of this feature (considering its interactions with other features) to the outcome. The sum of all Shapley values is the difference between the actual probability of a single sample and the mean probability of whole samples, which can be represented as:

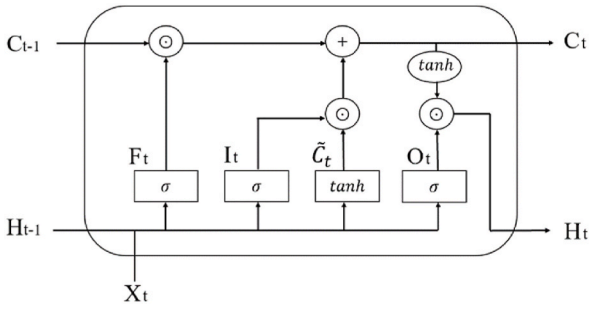


Fig. 2. Diagram of an LSTM neuron.

Table 1
Characteristics of study participants at baseline stratified by outcome.

Characteristic, n (%)	Total (N = 10,493)	NAFLD		*P* value
		Yes (N = 1662)	No (N = 8831)	
Demographic characteristics				
Sex				<0.001
Male	3162 (30.13)	714 (42.96)	2448 (27.72)	.
Female	7331 (69.87)	948 (57.04)	6383 (72.28)	.
Age/year	35.40 ± 8.73	37.13 ± 9.02	35.08 ± 8.63	<0.001
Education				<0.001
University degree or below	1781 (16.97)	390 (23.47)	1391 (15.75)	
Undergraduate	6018 (57.35)	961 (57.82)	5057 (57.26)	.
University degree and above	2694 (25.67)	311 (18.71)	2383 (26.98)	.
Income				<0.001
Low	7867 (74.97)	1326 (79.78)	6541 (74.07)	
Moderate	1456 (13.88)	191 (11.49)	1265 (14.32)	.
High	1170 (11.15)	145 (8.72)	1025 (11.61)	.
Body measurement				
BMI/(kg/m^2)	21.30 ± 2.27	22.45 ± 2.15	21.08 ± 2.23	<0.001
Waist circumference/cm	72.35 ± 7.25	75.93 ± 7.11	71.68 ± 7.08	<0.001
Body fat percentage/%	24.66 ± 5.54	26.33 ± 5.80	24.35 ± 5.44	<0.001
SBP/mmHg	105.86 ± 11.65	108.00 ± 12.34	105.46 ± 11.47	<0.001
DBP/mmHg	64.88 ± 8.72	65.91 ± 9.32	64.69 ± 8.59	<0.001
Laboratory tests				
FBG/(mmol/L)	5.25 ± 0.36	5.33 ± 0.35	5.24 ± 0.36	<0.001
TC/(mmol/L)	4.37 ± 0.70	4.47 ± 0.73	4.35 ± 0.70	<0.001
TG/(mmol/L)	0.84 ± 0.34	0.95 ± 0.37	0.82 ± 0.33	<0.001
HDLC/(mmol/L)	1.57 ± 0.29	1.50 ± 0.28	1.59 ± 0.29	<0.001
LDLC/(mmol/L)	2.58 ± 0.59	2.67 ± 0.59	2.56 ± 0.59	<0.001
ALT/(U/L)	15.93 ± 7.36	17.70 ± 7.84	15.59 ± 7.22	<0.001
AST/(U/L)	17.94 ± 4.51	18.44 ± 4.61	17.84 ± 4.48	<0.001
Direct bilirubin/(μmol/L)	3.00 ± 2.15	2.42 ± 2.15	3.11 ± 2.13	<0.001
Total bilirubin/(μmol/L)	8.63 ± 6.25	7.29 ± 6.41	8.88 ± 6.18	<0.001
Total protein/(g/L)	72.80 ± 2.44	72.81 ± 2.08	72.80 ± 2.50	0.838
Albumin/(g/L)	46.43 ± 1.53	46.48 ± 1.34	46.42 ± 1.56	0.087
Globulin/(g/L)	26.38 ± 2.15	26.34 ± 1.83	26.38 ± 2.21	0.351
GGT/(U/L)	15.24 ± 5.95	16.36 ± 5.84	15.03 ± 5.95	<0.001
ALP/(U/L)	57.65 ± 10.59	59.28 ± 9.59	57.34 ± 10.74	<0.001
LD/(U/L)	148.94 ± 15.65	150.34 ± 14.61	148.68 ± 15.83	<0.001
Uric acid/(μmol/L)	277.90 ± 58.55	292.31 ± 58.32	275.19 ± 58.19	<0.001
Living habits				
Smoking status				<0.001
Current	971 (9.25)	237 (14.26)	734 (8.31)	.
Former	200 (1.91)	25 (1.50)	175 (1.98)	.
Never	9322 (88.84)	1400 (84.24)	7922 (89.71)	.
Dietary regularity				0.161
Regular	8563 (81.61)	1336 (80.39)	7227 (81.84)	
Irregular	1930 (18.39)	326 (19.61)	1604 (18.16)	.
Fruit intake				0.061
Low	2117 (20.18)	357 (21.48)	1760 (19.93)	.
Moderate	7468 (71.17)	1183 (71.18)	6285 (71.17)	.
High	908 (8.65)	122 (7.34)	786 (8.90)	.
Dairy intake				0.003
Low	4905 (46.75)	840 (50.54)	4065 (46.03)	.
Moderate	2707 (25.80)	401 (24.13)	2306 (26.11)	.
High	2881 (27.46)	421 (25.33)	2460 (27.86)	.
Fried food intake				0.132
Low	6177 (58.87)	951 (57.22)	5226 (59.18)	.
Moderate	3734 (35.59)	604 (36.34)	3130 (35.44)	.
High	582 (5.55)	107 (6.44)	475 (5.38)	.
Personal history				
Hypertension				<0.001
Yes	404 (3.85)	106 (6.38)	298 (3.37)	
No	10,089 (96.15)	1556 (93.62)	8533 (96.63)	.
Diabetes				0.483
Yes	133 (1.27)	24 (1.44)	109 (1.23)	
No	10,360 (98.73)	1638 (98.56)	8722 (98.77)	.
Hyperlipidemic				<0.001
Yes	2625 (25.02)	516 (31.05)	2109 (23.88)	
No	7868 (74.98)	1146 (68.95)	6722 (76.12)	.

BMI = body mass index. SBP = systolic blood pressure. DBP = diastolic blood pressure. FBG = fasting blood glucose. TC = total cholesterol. TG = triglyceride. HDL-C = high-density lipoprotein cholesterol. LDL-C = low-density lipoprotein cholesterol. ALT = alanine transaminase. AST = aspartate aminotransferase. GGT = glutamyl transferase. ALP = alkaline phosphatase. LD = lactate dehydrogenase.

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \ldots + f(x_{ij})$$

In our study, the global contribution, together with the time-specific and sample-specific contributions of each feature can be generated by this algorithm.

3. Results

After excluding individuals who did not meet the inclusion criteria, a total of 10,493 participants were included for model development in our study. Among them, 1662 (15.84%) were diagnosed with NAFLD at and after their sixth checkup. Characteristics of study participants at baseline stratified by the outcome are presented in Table 1. People with the onset of NAFLD were older ($P < 0.001$), comprised more males ($P < 0.001$), with lower educational degrees ($P < 0.001$) and lower income ($P < 0.001$). Moreover, they had higher values of anthropometric measures, including BMI ($P < 0.001$), WC ($P < 0.001$), BFP ($P < 0.001$), SBP ($P < 0.001$), and DBP ($P < 0.001$). The differences in most of the laboratory tests between the two groups were also statistically significant ($P < 0.05$). Besides, the NAFLD group comprised more current smokers ($P < 0.001$), and had lower dairy intake ($P = 0.003$), and the prevalence of hypertension ($P < 0.001$) and hyperlipidemia ($P < 0.001$) was also significantly higher in the NAFLD group. To comprehensively

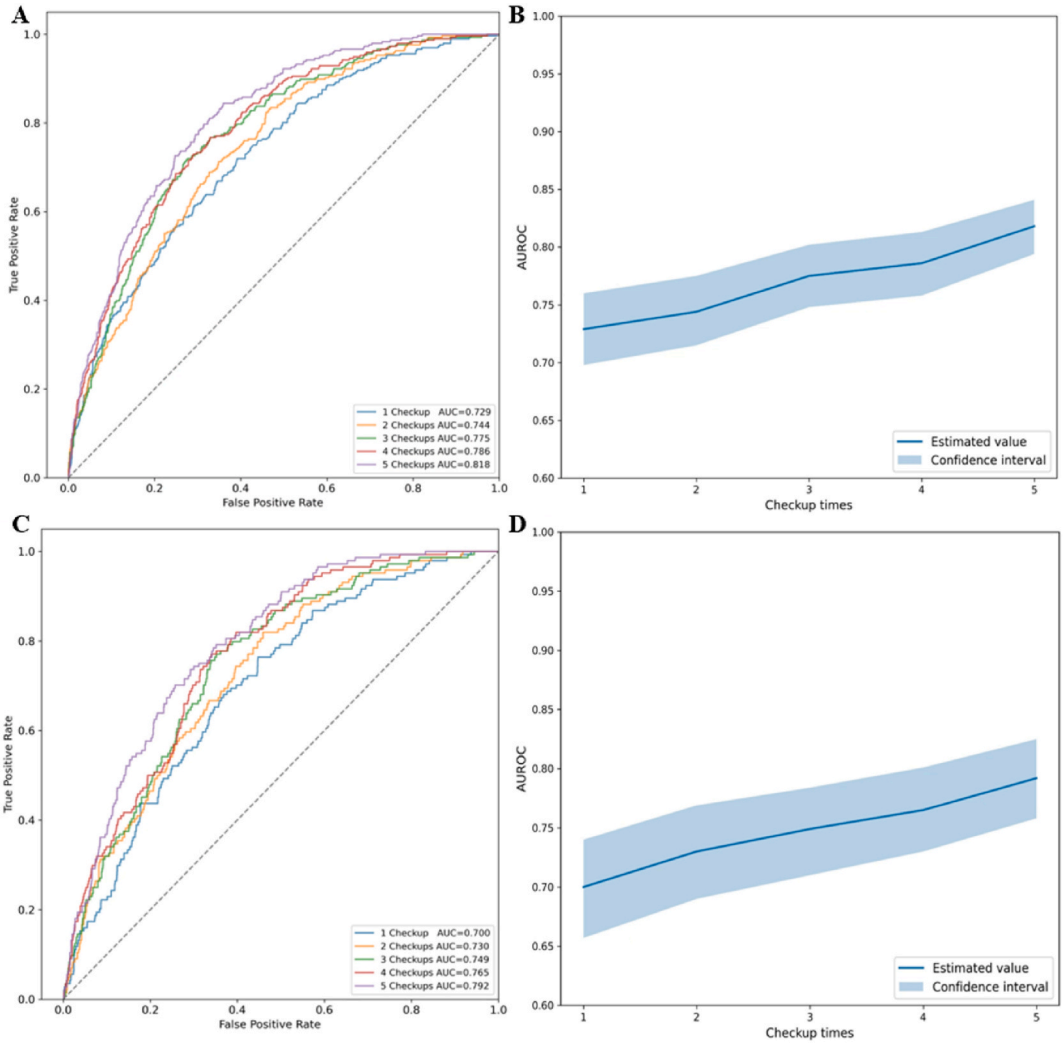


Fig. 3. Comparison of AUCs as the number of checkups increased. ROC curves of internal validation (A) and external validation (C); AUC as a function of checkup times in the internal validation dataset (B) and external validation dataset (D).

demonstrate the longitudinal characteristics of the study participants, we also presented the characteristics of their fifth health checkup and the checkup prior to their last follow-up in Supplemental Tab. 1 and Supplemental Tab. 2, respectively. Following the random partitioning of the dataset, we investigated differences in the baseline characteristics of the study participants between the training and internal validation sets. Supplemental Tab. 3 displays the results of this assessment, which demonstrated no statistically significant differences in any of the analyzed features.

The characteristics of participants at baseline and at fifth checkup in the external validation dataset after feature selection were presented in Supplemental Tab. 4 and Supplemental Tab. 5.

After feature selection, a total of 16 features, including sex, education, income, BMI, WC, BFP, ALT, UA, DB, LD, TG, AST, HDLC, ALP, GGT and GLB were included in our model. Model performance in predicting NAFLD improved as checkup records were renewed. In the internal validation dataset, the AUC increased significantly from 0.729 (95% CI: 0.698,0.760) at baseline to 0.818 (95% CI: 0.798,0.844) when consecutive 5 checkups were included (Fig. 3A–B). The recalibrated Brier scores decreased when more checkup records were included, indicating improved calibration. The decision curve analysis also showed that the net benefits of the model incorporated all five checkups significantly outperformed those of the other throughout all threshold probabilities (Fig. 4). The same growth trend in AUC can be seen in the external validation set, from 0.700 (95% CI: 0.657,0.740) at baseline to 0.792 (95% CI: 0.758,0.825) with all 5 checkups (Fig. 3C–D), although less accurate compared with those in the internal validation dataset. Brier scores also showed an improved calibration with the update of the checkups. The results of Delong test indicated that the improvement of AUC was statistically significant both in internal and external validation datasets (Supplemental Tab. 6). All metrics are presented in Table 2.

For comparison, LR was also used to develop prediction models at each health checkup, the model performance was presented in Supplemental Tab. 7. Although AUC increased with updated health checkup records, from 0.724 (95% CI: 0.695,0.753) to 0.784 (95% CI: 0.758,0.810) in internal validation dataset, and from 0.713 (0.673,0.753) to 0.787 (0.750,0.823) in external validation dataset, the overall performance of the model was inferior to that of LSTM models ($P < 0.0001$) based on the result of repeated measures ANOVA, and the improvement trend was not as strong as the latter ($P < 0.0001$), as evidenced by the results presented in Supplemental Tab. 8 and Supplemental Tab. 9. Specifically, significant differences ($P < 0.001$) were observed in the prediction performance of the models in the internal validation set as the number of health checkups varied. Additionally, the LSTM models outperformed the LR models in terms of overall performance over five time points, with mean AUC values of 0.770 and 0.752, respectively. Moreover, the trend in AUC improvement with the update of checkup records differed significantly between the two models, with the LSTM model showing a stronger improvement trend than the LR model (difference in AUC between final and initial visits: 0.089 vs. 0.060, respectively). These results were further supported by external validation. Besides, to more clearly compare the predictive performance differences between LR and LSTM models, we also used Wilcoxon signed rank tests to compare the AUC differences between the two models at or before each time point. The results are presented in Supplemental Tab. 10.

To address the concern of potential improvement in model performance due to the decreasing time interval between the measurement time of the variables and the occurrence of the outcome with the incorporation of new health checkup records, we further developed models relative to the time of the fifth checkup, incorporating the fourth, the third, and until the first gradually, and evaluating its performance. The results are presented in Fig. 5A and B and Table 3, which also showed an increased trend in model performance.

The contributions of each feature in the model related to the risk of NAFLD are shown in Fig. 6A and B. Among all these features, body fat percentage, ALT, and uric acid own the biggest impact on the outcome, with higher values above indicating greater risks of the development of NAFLD. Besides, some demographic characteristics, like income and education, are also at the priority level.

The dynamic prediction model can also be interpreted at a specific checkup time, the summary plots of which can be seen in Supplemental F. 2–6. The results of the mean importance rank of the features over time are presented in Fig. 7, which illustrates the changing contributions of a certain feature. We note that body fat percentage and ATL occupied the top ranks in almost all checkup times, HDLC gained importance with the extension of time steps, while the ranks of DB decreased, meaning losing importance over

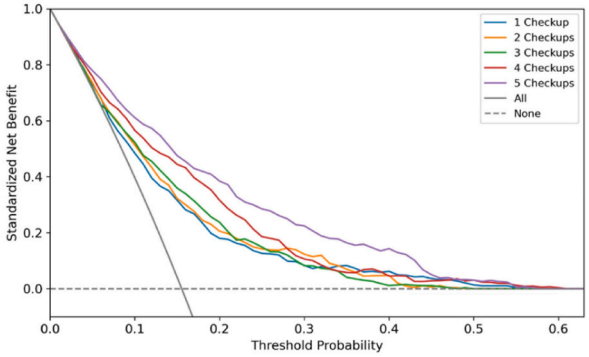


Fig. 4. Decision curve analysis with the update of checkups in the internal validation dataset. The gray solid line and the gray dashed line represent the net benefit generated by intervention, and non-intervention to all participants, respectively. Other colored lines present intervention to all potential patients identified by each model throughout threshold probabilities.

Table 2
Model performance with the update of health checkups based on LSTM.

Model	AUC	Sensitivity	Specificity	Brier score
Internal validation dataset				
1 Checkup	0.729 (0.698,0.760)	0.700 (0.548,0.870)	0.641 (0.459,0.780)	0.131 (0.119,0.142)
2 Checkups	0.744 (0.715,0.775)	0.774 (0.649,0.884)	0.605 (0.481,0.721)	0.112 (0.101,0.123)
3 Checkups	0.775 (0.748,0.802)	0.735 (0.653,0.808)	0.715 (0.652,0.784)	0.087 (0.076,0.094)
4 Checkups	0.786 (0.758,0.813)	0.748 (0.660,0.872)	0.700 (0.558,0.775)	0.067 (0.058,0.077)
5 Checkups	0.818 (0.798,0.844)	0.804 (0.713,0.873)	0.690 (0.627,0.772)	0.052 (0.041,0.066)
External validation dataset				
1 Checkup	0.700 (0.657,0.740)	0.741 (0.532,0.897)	0.590 (0.422,0.788)	0.076 (0.066,0.087)
2 Checkups	0.730 (0.690,0.769)	0.765 (0.646,0.925)	0.579 (0.446,0.745)	0.077 (0.066,0.088)
3 Checkups	0.749 (0.710,0.784)	0.786 (0.702,0.877)	0.642 (0.524,0.686)	0.076 (0.066,0.088)
4 Checkups	0.765 (0.730,0.801)	0.792 (0.702,0.904)	0.646 (0.516,0.710)	0.067 (0.057,0.078)
5 Checkups	0.792 (0.758,0.825)	0.791 (0.611,0.911)	0.695 (0.503,0.790)	0.066 (0.055,0.077)

LSTM = long short-term memory. AUC = area under the curve.

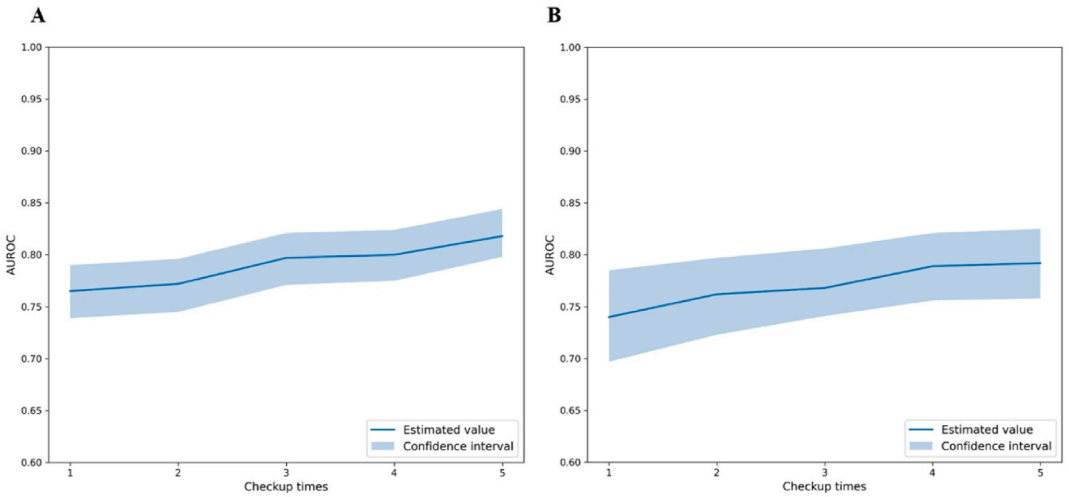


Fig. 5. AUC as a function of checkup times incorporated from the fifth checkup till the first in the internal validation dataset (A) and external validation dataset (B).

Table 3
Model performance incorporated checkup data from the fifth till the first.

Model	AUC	Sensitivity	Specificity	Brier score
Internal validation dataset				
5th Checkup	0.765 (0.739,0.790)	0.770 (0.679,0.858)	0.655 (0.564,0.729)	0.118 (0.110,0.127)
4th-5th Checkups	0.772 (0.745,0.796)	0.774 (0.677,0.855)	0.666 (0.566,0.750)	0.115 (0.106,0.123)
3rd-5th Checkups	0.797 (0.771,0.821)	0.789 (0.699,0.865)	0.684 (0.608,0.763)	0.092 (0.083,0.101)
2nd-5th Checkups	0.800 (0.775,0.824)	0.773 (0.686,0.844)	0.712 (0.649,0.807)	0.071 (0.060,0.084)
1st-5th Checkups	0.818 (0.798,0.844)	0.804 (0.713,0.873)	0.690 (0.627,0.772)	0.052 (0.041,0.066)
External validation dataset				
5th Checkup	0.740 (0.697,0.785)	0.629 (0.494,0.748)	0.770 (0.684,0.879)	0.088 (0.082,0.095)
4th-5th Checkups	0.762 (0.723,0.797)	0.717 (0.494,0.888)	0.682 (0.449,0.888)	0.091 (0.084,0.098)
3rd-5th Checkups	0.768 (0.741,0.806)	0.784 (0.696,0.892)	0.647 (0.533,0.723)	0.081 (0.073,0.088)
2nd-5th Checkups	0.789 (0.756,0.821)	0.804 (0.607,0.950)	0.640 (0.476,0.832)	0.079 (0.073,0.086)
1st-5th Checkups	0.792 (0.758,0.825)	0.791 (0.611,0.911)	0.695 (0.503,0.790)	0.066 (0.055,0.077)

AUC = area under the curve.

time.

We presented the fourth and fifth checkups of a representative case from the internal validation dataset in Fig. 8A and B, respectively. The participant was a 45-year-old male, with an undergraduate degree and a moderate level of income. At his fourth checkup, his BMI was 26.93 kg/m^2, waist circumference 82 cm, and body fat percentage 39.0%, with none of the three chronic diseases. Laboratory tests showed his direct bilirubin 2.67 μmol/L, ALT 16.0U/L, TG 1.95 mmol/L, and HDLC 1.01 mmol/L, which

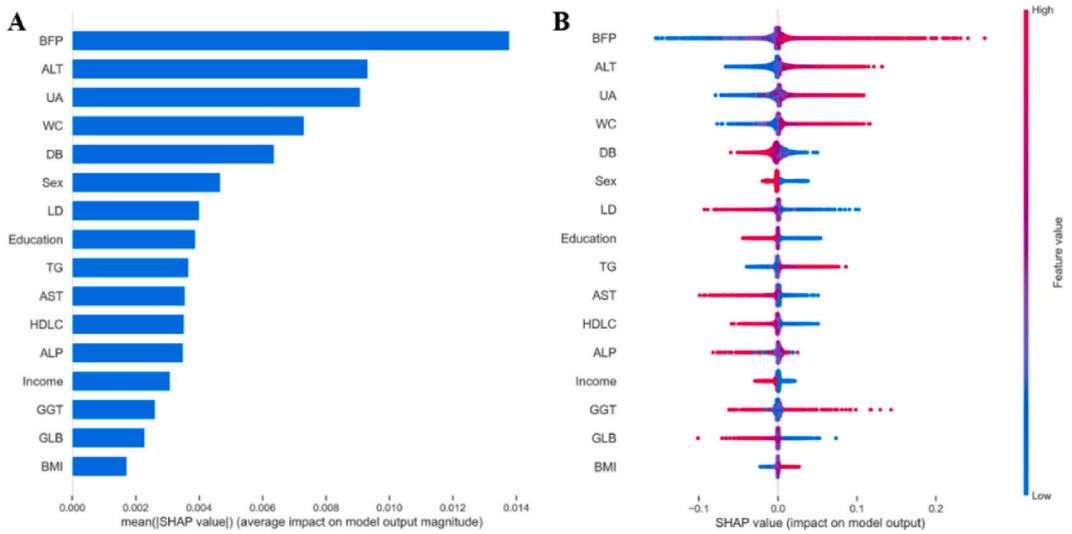


Fig. 6. Summary plot of the impact of features on the risk of NAFLD. (A) The blue bars represent the sum of the absolute SHAP value on features. The higher SHAP values indicate more contributions to prediction. (B) Feature significance decreased from top to bottom. Each point represents the impact of a feature on NAFLD prediction for one participant at a given checkup time. Red points with positive SHAP values, which are the same as blue points with negative SHAP values, represent positive associations with the outcome, while red points with negative SHAP values, similarly to blue points with positive SHAP values, mean negative relationships. BFP = body fat percentage. ALT = alanine transaminase. UA = uric acid. WC = waist circumference. DB = direct bilirubin. LD = lactate dehydrogenase. TG = triglyceride. AST = aspartate aminotransferase. HDLC = high-density lipoprotein cholesterol. ALP = alkaline phosphatase. GGT = glutamyl transferase. GLB = globulin. BMI = body mass index.

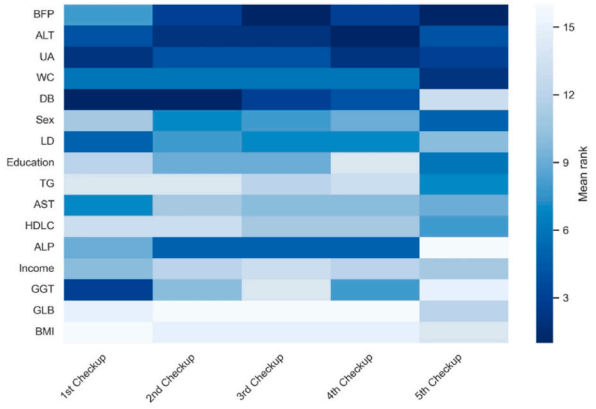


Fig. 7. The dynamic impact of features on the risk of NAFLD. BFP = body fat percentage. ALT = alanine transaminase. UA = uric acid. WC = waist circumference. DB = direct bilirubin. LD = lactate dehydrogenase. TG = triglyceride. AST = aspartate aminotransferase. HDLC = high-density lipoprotein cholesterol. ALP = alkaline phosphatase. GGT = glutamyl transferase. GLB = globulin. BMI = body mass index.

were all risk factors identified by the model. One year later, he tested a 10U/L increase in GGT, which was out of the normal range and pulled the prediction towards the onset of NAFLD. The values of waist circumstance, TG and direct bilirubin decreased but were still identified as risk factors. Meanwhile, the measurements of BMI, body fat percentage, HDLC, ALT and LD, increased a little compared with the fourth checkup, although most of which, especially the laboratory test results, were still within their normal limits, identified as strong risk factors to the prediction. This participant was diagnosed with NAFLD at his sixth checkup. More information about this individual can be seen in Supplemental Tab. 11.

BFP = body fat percentage. GGT = glutamyl transferase. DB = direct bilirubin. ALT = alanine transaminase. TG = triglyceride. HDLC = high-density lipoprotein cholesterol. UA = uric acid. WC = waist circumference. LD = lactate dehydrogenase. BMI = body mass index. GLB = globulin.

We further explored the relationships between features and the outcome, and detailed how some features interact (Fig. 9). A non-linear relationship exists between UA measured at the third checkup, ALP measured at the fourth checkup and the outcome (Fig. 9A and B). The normal ranges of UA and ALP are 180–420 mmol/L and 40–150U/L, respectively, meaning that the results of these two laboratory tests in most participants are within the normal ranges. Besides, the relationship between the measurement value and the

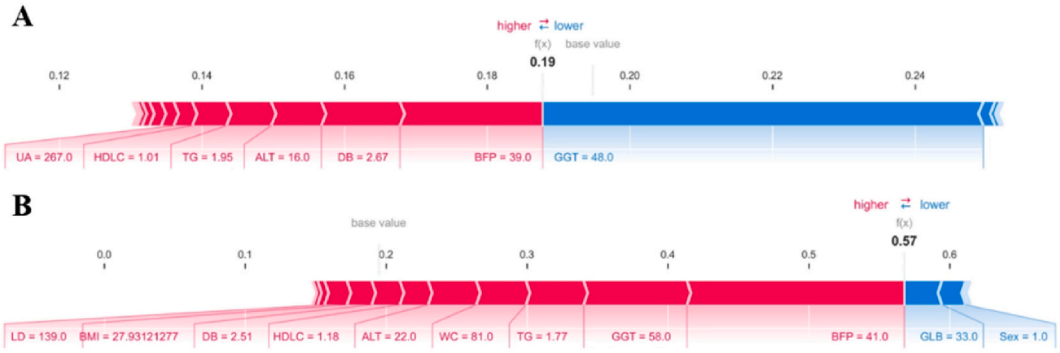


Fig. 8. Time-specific impact of features for a single participant (A) the 4th checkup (B) the 5th checkup. The features in red color drive the outcome towards the side of the development of the NAFLD, while the features in blue color pull down the risk probability.

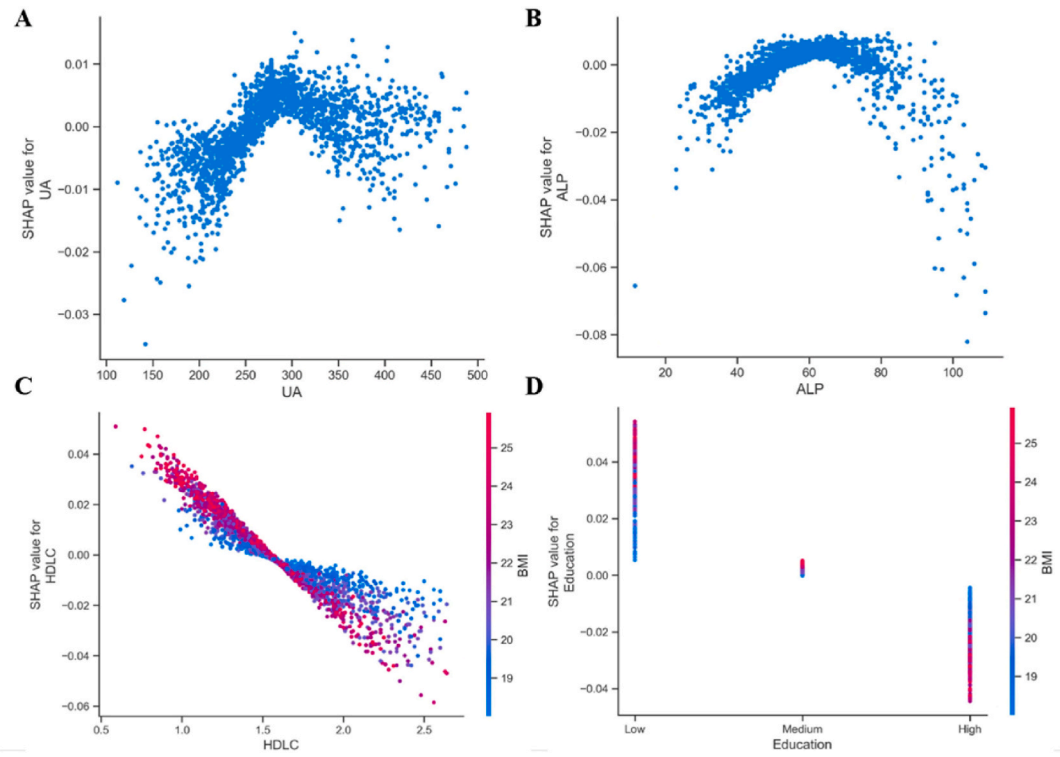


Fig. 9. Impact of feature interactions on NAFLD prediction. (A) the relationship between UA and the outcome; (B) the relationship between ALP and the outcome; (C) the interaction of HDLC and BMI; (D) the interaction of education and BMI. The SHAP value corresponding to each dot represents the impact of a feature and its interaction with other features for one participant. The positive SHAP value pushes the prediction towards the onset of NAFLD, while the negative SHAP value reduces the risk of the disease. UA = uric acid. ALP = alkaline phosphatase. HDLC = high-density lipoprotein cholesterol. BMI = body mass index.

risk of NAFLD was found to be non-linear, contrary to the expected linear relationship. Specifically, the risk of NAFLD increased initially with the measurement values, followed by a decrease, indicating a complex relationship between the predictors and the outcome. Furthermore, it is obvious that a higher value of HDLC is associated with a lower risk of NAFLD, but when considering BMI, we can see that most of the dots in red color, indicating high BMI, are distributed far away from the position of 0 SHAP value, which means that the interaction of high BMI and HDLC has a stronger impact on the prediction (Fig. 9C). Similarly, a lower educational degree forebodes a higher risk of NAFLD, and a higher BMI also strengthens the impact (Fig. 9D).

4. Discussion

In this study, a time-series deep learning model was developed to predict the risk of NAFLD within the health checkup population,

and the model was updated when new checkup records were incorporated to achieve dynamic prediction. SHAP algorithm was used to enhance model interpretability and identify features that had a great impact on outcomes.

As expected, with the extension of the time step, the latest records were incorporated and more information about temporal trends was contained, so the model performance improved significantly. However, when incorporating records from the first checkups to the fifth checkups stepwise, the time of predictors measurement will approach the time of prediction, so the model performance may improve on this account. To deal with this problem, we further developed a model relative to the time of the fifth checkup, incorporating the fourth, the third, and until the first gradually, and evaluating its performance. The results also showed an increased trend in model performance, although not as significant as the forward incorporation. This finding indicates that the features measured years before the development of NAFLD are still of great value and need to be taken into account to better predict the outcome.

While cross-sectional data-based diagnostic prediction models have shown promising results, they fall short in predicting the future incidence of NAFLD and identifying high-risk populations before the onset of the disease [5,7,8,20]. Prognosis prediction models, on the other hand, excel in early risk estimation for diseases. Compared to existing studies on prognosis prediction models, the predictive performance of the model developed in our study demonstrates certain advantages. In a study by Zhu et al., multivariate LR model was constructed to predict the risk of NAFLD among lean pre-diabetics with normal blood lipid levels, achieving an AUC of 0.796 in the longitudinal internal validation with a 5-year follow-up period [21]. Abeysekera et al. utilized variables measured at the participants' adolescence to predict the incidence risk of NAFLD at the age of 24 using LR, yielding the highest AUROC of 0.79 [22]. Wang et al. conducted a four-year cohort study to predict the incidence risk of NAFLD among 13,240 baseline NAFLD-free individuals using cox proportional hazards regression analyses. In the internal validation cohort, the AUC for 1-year, 2-year, 3-year, and 4-year risk predictions were 0.817, 0.820, 0.814, and 0.813, respectively [23]. In contrast, our study focused on developing a dynamic prediction model with constantly updated health checkup records using advanced deep learning algorithms that estimated the incidence risk of NAFLD from the sixth health checkup onwards until the end of the follow-up period. The internal validation set yielded an impressive highest AUC of 0.818 (0.798, 0.844), showcasing the innovative and competitive performance of our model.

Our research reveals that the prediction model we established surpasses the LR model used as the benchmark. We found that our LSTM-based model's performance further enhanced as more recent health checkup records become available compared to the LR model. This implies that the LSTM model could offer more accurate predictions over a longer time span of data. Additionally, as both models employ the same predictive variables, the cost of using either model remains identical. Our validation set results demonstrate that while the LR model's discriminative ability improves with updated health checkup records, its calibration ability decreases. Conversely, our LSTM-based prediction model exhibits greater robustness, stability, and generalizability. Our findings suggest that the LSTM model could serve as a more dependable and precise tool for time-series data prediction, offering superior performance to the LR model.

Although in many studies, BMI was regarded as an important predictor of NAFLD [24–26], we identified the more valid body measurement indicator, body fat percentage. This indicator could help recognize those with normal weight obesity, whose BMI cannot tell [27]. Besides, this feature had a roughly increased importance rank over time, showing more value than other body measurement indicators [28]. As a practical, easily measured, and non-invasive indicator, we recommend using body fat percentage to supplement other body measurement indicators. However, further research needs to be done to test whether there exists a causal relationship between BFP and NAFLD. ALT, a standard and commonly used indicator of liver function, was found to have a great impact on the prediction in our study. As an enzyme in the cytosol of hepatocytes, only a low level of ALT can be detected in the serum in a healthy population, elevated ALT level usually indicates the injury or apoptosis of hepatocytes [29]. Previous studies have also concluded that the increased ALT value was associated with a higher risk of NAFLD [30–32]. In a cross-sectional study on elderly Chinese man and woman, ALT was observed to have a significant joint association with serum uric acid to NAFLD prevalence [33], although we haven't considered the joint contribution of the two features in our study, serum uric acid was also found to have a strong impact on the outcome. The global contribution of UA in all five checkups showed that a high level of the feature driving the outcome towards the side of NAFLD. Meanwhile, we also found a non-linear relationship between UA and the outcome that existed in a single checkup. Some cross-sectional studies have demonstrated that serum uric acid was positively associated with the prevalence of NAFLD in Chinses population [34]. Besides, a case-control study also indicated that uric acid was an independent predictor of NAFLD [35]. As the end product of purine metabolism, high levels of serum uric acid could induce fat accumulation and then lead to the development of fatty liver [36]. Waist circumference was also a significant predictor of NAFLD, consistent finding has been revealed in previous studies [7, 37]. Furthermore, our results also showed that direct bilirubin levels from five or more years ago can have a significant and continuous effect on the development of NAFLD, which may hint at its long-term effect on the onset of the disease. Nevertheless, lifestyle habits collected from the questionnaire survey, including dietary condition and smoking status, although proved to be associated with the development of NAFLD by other studies [38–40], were not regarded as important factors in our study. It makes sense when considering the long-time interval between the survey time of these habits and the prediction, along with the possible changeability of these features. From the SHAP algorithms, we also illustrated the interactions between some features. We found that a high BMI could strengthen the impact of certain features on the outcome, including HDLC and education. Specifically, the interactions between high BMI and the two features tended to drive the outcome towards two extremes. Thus, minor changes in laboratory test results for the high BMI population may have a strong impact on the development of NAFLD, which deserves more attention.

We selected a representative sample from the internal validation dataset to show the quantitative and individual-specific impact of features and provide the changing process over time. This sample was predicted to develop NAFLD in the future based on his consecutive 5 checkups. But when regardless of his initial 4 checkups and only his last checkup record was included, he would be predicted as exempted from NAFLD, which appears to be a wrong prediction and is contrary to reality. Specifically, almost all features in his fifth checkup seemed not very abnormal and most were within normal limits, but when compared with his fourth checkup, we

can find that there exists a slight increase in BMI, body fat percentage, HDLC, ALT, and many other features. Thus, we can conclude that it was the variation trend contained in the time-series checkup data that prompted the correct prediction. Although in some studies, time-series data were transformed into summary statistics, including mean, minimum and maximum values, to fully harness the information contained in the repeated measurement data [41,42], the temporal trends in detail cannot be captured adequately [43, 44]. Therefore, instead of focusing on snapshot measurements or the transformed statistics of multiple measurements, the original measurement history and long-term trends of some features also need to be taken into account.

It should be noted that our study was subject to the presence of right censoring as well as potential bias due to loss to follow-up, which are inherent limitations associated with utilizing health checkup data. As presented in Supplementary Tab. 12, we report the number and proportion of participants lost to follow-up at each subsequent health checkup after the sixth one. However, we performed a thorough analysis of participant characteristics between those who were lost to follow-up and those who were not at their sixth health checkup, as presented in Supplementary Tab. 13. Our comparison revealed that there were negligible differences between the two groups, as most indicators did not exhibit statistically significant differences ($P > 0.05$), and the standardized mean difference (SMD) results indicated that almost all indicators had an SMD of less than 0.1 [45]. In addition, we also conducted a separate comparison of baseline characteristics between participants lost to follow-up and those who were not, taking into account the outcome (Supplementary Tab. 14). This allowed for a better comparison with Table 1. The analysis revealed that the distribution of most characteristic differences between different outcomes was similar across both the lost to follow-up and not lost to follow-up groups. These findings partly mitigate the potential impact of loss to follow-up on our results.

While our health check-up database contained data spanning 2003–2012, we chose to analyze only those individuals with six or more health checkups. This decision aimed to balance the trade-off between selecting too few or too many years of data, as the former would not fully reflect changes in the time-series data while the latter would result in an unacceptably small sample size. Supplementary Tab. 15 illustrates the relationship between the number of health checkups included and corresponding sample sizes and NAFLD incidence rates. Although selecting patients with seven or more health checkup records would provide richer data, it would reduce the sample size to below 10,000, compromising the statistical power of our deep learning model. Additionally, if the selected time span is too long, the corresponding population may be relatively healthy, with a low incidence rate of NAFLD, and thus the practical value of the model would be limited. Conversely, choosing five or fewer health check-up records would produce a relatively short time-series that might not capture the full spectrum of changes in the underlying data. Therefore, we opted to focus on patients with at least six recorded health checkups to maintain an optimal balance between sample size and data richness.

We own some strengths in this study. Firstly, we made full use of the routine health checkup data and kept their time-series forms to extract the valuable information contained, and used the SHAP algorithm to visualize changing contributions of top predictors and provided the individual-level model explanation. Secondly, we explored the dynamic prediction by incorporating newly updated checkups to adapt to the evolving health checkup picture, which is more useful than a static one and can provide timely guidance to participants. Thirdly, we focused on developing a prognostic prediction model with a time interval of over one year to help identify high-risk groups at early stages, which may own more value compared with most of the current diagnostic prediction models. Fourthly, we verified the predictive performance of our model through external validation on a completely independent dataset, so the generalization and the practical value of the prediction model are in prospect.

There are still several limitations in our study. Firstly, as the high prevalence rate of NAFLD, we filtered out those diagnosed with NAFLD in their initial five checkups and kept a relatively healthier population, which may limit the application of the model. Secondly, we trained the LSTM model with consecutively updated 5 checkups of each participant. Even though our longitudinal data spanned over 5 years, the checkup records hadn't been renewed at a very high frequency. While the strength of the LSTM algorithm lies in its learning about long-term dependencies, which may partially affect our results [46]. Thirdly, our study did not comprehensively investigate the potential differences in missingness patterns between outcome groups and over time during the imputation of missing values, and the use of the LOCF method may not be optimal for our checkup data. Therefore, a more in-depth exploration of missingness imputation techniques is warranted to enhance the robustness of our results. Fourthly, although the MJ Health Screening Center performs calibration of its examination instruments and equipment regularly, the lengthy time frame of our study may still result in natural year-to-year differences in biochemical values. Fifthly, our feature selection process did not consider the longitudinal nature of the data, but instead relied solely on variables obtained during a single checkup, which may partly bias our results. Lastly, despite the relatively high sensitivity exhibited by our prediction model, there are still individuals who are incorrectly predicted as negative, resulting in false negatives. This may be due to the right censoring of data, which could lead to some study participants being diagnosed with NAFLD long after the fifth check-up. Nevertheless, it is essential to recognize that various factors may contribute to the manifestation of false negatives, and their quantification poses a significant challenge.

In conclusion, we developed a dynamic and continuously updated deep learning model based on time-series health checkup data for the prediction of NAFLD in the future checkup and achieved good performance both in calibration and discrimination. Predictors of top significance were identified and their impact on the outcome was visualized both in global settings and individual settings.

Ethics statement

Ethical approval from the Institutional Review Board of Peking University Health Science Center (ID of the approval: IRB00001052-19077) was received by this study. And as all data were de-identified in this study, individual informed consent was waived.

Author contribution statement

Yuhan Deng, Yuan Ma and Jingzhu Fu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jianchun Yin, Canqing Yu, Jun Lv: Conceived and designed the experiments; Analyzed and interpreted the data.

Sailimai Man, Bo Wang, Liming Li: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was funded by the Ministry of Science and Technology of the People's Republic of China (grant number 2020YFC2004703); the National Natural Science Foundation of China (grant number 82192901, 82192904, 82192900); and the Ministry of Science and Technology of the People's Republic of China (grant number 2020YFC2003405).

Data availability statement

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18758.

References

[1] J. Li, B. Zou, Y.H. Yeo, Y. Feng, X. Xie, D.H. Lee, et al., Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999-2019: a systematic review and meta-analysis, Lancet Gastroenterol Hepatol 4 (5) (2019) 389–398, https://doi.org/10.1016/S2468-1253(19)30039-1.

[2] T. Marjot, A. Moolla, J.F. Cobbold, L. Hodson, J.W. Tomlinson, Nonalcoholic fatty liver disease in adults: current concepts in etiology, outcomes, and management, Endocr. Rev. 41 (1) (2020) bnz009, https://doi.org/10.1210/endrev/bnz009.

[3] J.M. Paik, K. Kabbara, K.E. Eberly, Y. Younossi, L. Henry, Z.M. Younossi, Global burden of NAFLD and chronic liver disease among adolescents and young adults, Hepatology 75 (5) (2022) 1204–1217, https://doi.org/10.1002/hep.32228.

[4] M. Noureddin, F. Ntanios, D. Malhotra, K. Hoover, B. Emir, E. McLeod, et al., Predicting NAFLD prevalence in the United States using National Health and Nutrition Examination Survey 2017-2018 transient elastography data and application of machine learning, Hepatol Commun 6 (7) (2022) 1537–1548, https://doi.org/10.1002/hep4.1935.

[5] X. Ma, C. Yang, K. Liang, B. Sun, W. Jin, L. Chen, et al., A predictive model for the diagnosis of non-alcoholic fatty liver disease based on an integrated machine learning method, Am J Transl Res 13 (11) (2021) 12704–12713.

[6] A. Atsawarungruangkit, P. Laoveeravat, K. Promrat, Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database, World J. Hepatol. 13 (10) (2021) 1417–1427, https://doi.org/10.4254/wjh.v13.i10.1417.

[7] W. Ji, M. Xue, Y. Zhang, H. Yao, Y. Wang, A machine learning based framework to identify and classify non-alcoholic fatty liver disease in a large-scale population, Front. Public Health 10 (2022), 846118, https://doi.org/10.3389/fpubh.2022.846118.

[8] Y.X. Liu, X. Liu, C. Cen, X. Li, J.M. Liu, Z.Y. Ming, et al., Comparison and development of advanced machine learning tools to predict nonalcoholic fatty liver disease: an extended study, Hepatobiliary Pancreat. Dis. Int. 20 (5) (2021) 409–415, https://doi.org/10.1016/j.hbpd.2021.08.004.

[9] Æ.Ö. Kristinsson, Y. Gu, S.M. Rasmussen, J. Mølgaard, C. Haahr-Raunkjær, C.S. Meyhoff, E.K. Aasvang, H.B.D. Sørensen, Prediction of serious outcomes based on continuous vital sign monitoring of high-risk patients, Comput. Biol. Med. 147 (2022), 105559, https://doi.org/10.1016/j.compbiomed.2022.105559.

[10] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539.

[11] S. Nemati, A. Holder, F. Razmi, M.D. Stanley, G.D. Clifford, T.G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the ICU, Crit. Care Med. 46 (4) (2018) 547–553, https://doi.org/10.1097/CCM.0000000000002936.

[12] A.J. London, Artificial intelligence and black-box medical decisions: accuracy versus explainability, Hastings Cent. Rep. 49 (1) (2019) 15–21, https://doi.org/10.1002/hast.973.

[13] W.X. Lim, Z. Chen, A. Ahmed, The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review, Med. Biol. Eng. Comput. 60 (3) (2022) 633–642, https://doi.org/10.1007/s11517-021-02487-8.

[14] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Proceedings of the 31st International Conference on Neural Information Processing Systems (2017) 4768–4777.

[15] M.D. Nemesure, M.V. Heinz, R. Huang, N.C. Jacobson, Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence, Sci. Rep. 11 (1) (2021) 1980.

[16] Y. Zoabi, S. Deri-Rozov, N. Shomron, Machine learning-based prediction of COVID-19 diagnosis based on symptoms, NPJ Digit Med 4 (1) (2021) 3.

[17] E. Tasci, Y. Zhuge, K. Camphausen, A.V. Krauze, Bias and class imbalance in oncologic data-towards inclusive and transferrable AI in large scale oncology data sets, Cancers 14 (12) (2022) 2897.

[18] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Advances in Large Margin Classifiers 10 (1999) 61–74.

[19] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[20] S. Qin, X. Hou, Y. Wen, C. Wang, X. Tan, H. Tian, Q. Ao, J. Li, S. Chu, Machine learning classifiers for screening nonalcoholic fatty liver disease in general adults, Sci. Rep. 13 (1) (2023) 3638.

[21] W. Zhu, P. Shi, J. Fu, A. Liang, T. Zheng, X. Wu, S. Yuan, Development and application of a novel model to predict the risk of non-alcoholic fatty liver disease among lean pre-diabetics with normal blood lipid levels, Lipids Health Dis. 21 (1) (2022) 149.

[22] K.W.M. Abeysekera, J.G. Orr, F.H. Gordon, L.D. Howe, J. Hamilton-Shield, J. Heron, M. Hickman, Evaluating future risk of NAFLD in adolescents: a prediction and decision curve analysis, BMC Gastroenterol. 22 (1) (2022) 323.
[23] J. Wang, Y. Tang, K. Peng, H. Liu, J. Xu, Development and validation of a nomogram for predicting nonalcoholic fatty liver disease in the non-obese Chinese population, Am. J. Tourism Res. 12 (10) (2020) 6149–6159.
[24] S. Saokaew, S. Kanchanasuwan, P. Apisarnthanarak, A. Charoensak, P. Charatcharoenwitthaya, P. Phisalprapa, et al., Clinical risk scoring for predicting non-alcoholic fatty liver disease in metabolic syndrome patients (NAFLD-MS score), Liver Int. 37 (10) (2017) 1535–1543.
[25] A.E. Rigamonti, A. Bondesan, E. Rondinelli, S.G. Cella, A. Sartorio, The role of aspartate transaminase to platelet ratio index (APRI) for the prediction of non-alcoholic fatty liver disease (NAFLD) in severely obese children and adolescents, Metabolites 12 (2) (2022) 155.
[26] K.W.M. Abeysekera, J.G. Orr, F.H. Gordon, L.D. Howe, J. Hamilton-Shield, J. Heron, et al., Evaluating future risk of NAFLD in adolescents: a prediction and decision curve analysis, BMC Gastroenterol. 22 (1) (2022) 323.
[27] E. Oliveros, V.K. Somers, O. Sochor, K. Goel, F. Lopez-Jimenez, The concept of normal weight obesity, Prog. Cardiovasc. Dis. 56 (4) (2014) 426–433.
[28] A.G. Mainous 3rd, B.J. Rooks, J.F. Medley, S.B. Dickmann, Body composition among adults at a healthy body mass index and association with undetected non-alcoholic fatty liver, Int. J. Obes. 46 (7) (2022) 1403–1405.
[29] W.R. Kim, S.L. Flamm, A.M. Di Bisceglie, et al., Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease, Hepatology 47 (2008) 1363–1370.
[30] X. Ma, S. Liu, J. Zhang, M. Dong, Y. Wang, M. Wang, et al., Proportion of NAFLD patients with normal ALT value in overall NAFLD patients: a systematic review and meta-analysis, BMC Gastroenterol. 20 (2020) 10.
[31] D.N. Amarapurka, A.D. Amarapurkar, N.D. Patel, et al., Nonalcoholic steatohepatitis (NASH) with diabetes: predictors of liver fibrosis, Ann. Hepatol. 5 (2006) 30–33.
[32] M.T. Long, A. Pedley, L.D. Colantonio, J.M. Massaro, U. Hoffmann, P. Muntner, et al., Development and validation of the framingham steatosis index to identify persons with hepatic steatosis, Clin. Gastroenterol. Hepatol. 14 (2016) 1172–1180.
[33] H. Yang, D. Li, X. Song, F. Liu, X. Wang, Q. Ma, et al., Joint associations of serum uric acid and ALT with NAFLD in elderly men and women: a Chinese cross-sectional study, J. Transl. Med. 16 (1) (2018) 285.
[34] Y. Li, C. Xu, C. Yu, L. Xu, M. Miao, Association of serum uric acid level with non-alcoholic fatty liver disease: a cross-sectional study, J. Hepatol. 50 (5) (2009) 1029–1034.
[35] A. Lonardo, P. Loria, F. Leonardi, A. Borsatti, P. Neri, M. Pulvirenti, et al., Fasting insulin and uric acid levels but not indices of iron metabolism are independent predictors of non-alcoholic fatty liver disease. A case-control study, Dig. Liver Dis. 34 (3) (2002) 204–211.
[36] Y.J. Choi, H.S. Shin, H.S. Choi, J.W. Park, I. Jo, et al., Uric acid induces fat accumulation via generation of endoplasmic reticulum stress and SREBP-1c activation in hepatocytes, Lab. Invest. 94 (10) (2014) 1114–1125.
[37] A. Atsawarungruangkit, P. Laoveeravat, K. Promrat, Machine learning models for predicting non-alcoholic fatty liver disease in the general United States population: NHANES database, World J. Hepatol. 13 (10) (2021) 1417–1427.
[38] E. Molina-Molina, G.E. Furtado, J.G. Jones, P. Portincasa, A. Vieira-Pedrosa, A.M. Teixeira, et al., The advantages of physical exercise as a preventive strategy against NAFLD in postmenopausal women, Eur. J. Clin. Invest. 52 (3) (2022), e13731.
[39] S. Ebrahimi Mousavi, N. Dehghanseresht, F. Dashti, Y. Khazaei, S. Salamat, et al., The association between Dietary Diversity Score and odds of nonalcoholic fatty liver disease: a case-control study, Eur. J. Gastroenterol. Hepatol. 34 (6) (2022) 678–685.
[40] J.H. Lee, H.S. Lee, S.B. Ahn, Y.J. Kwon, Dairy protein intake is inversely related to development of non-alcoholic fatty liver disease, Clin Nutr 40 (10) (2021) 5252–5260.
[41] Z. Zeng, S. Yao, J. Zheng, X. Gong, Development and validation of a novel blending machine learning model for hospital mortality prediction in ICU patients with Sepsis, BioData Min. 14 (1) (2021 Aug 16) 40.
[42] Y. Zhu, J. Zhang, G. Wang, R. Yao, C. Ren, G. Chen, et al., Machine learning prediction models for mechanically ventilated patients: analyses of the MIMIC-III database, Front. Med. 8 (2021), 662340.
[43] J. Lopez Bernal, S. Soumerai, A. Gasparrini, A methodological framework for model selection in interrupted time series studies, J. Clin. Epidemiol. 103 (2018) 82–91.
[44] Y. Xue, D. Klabjan, Y. Luo, Predicting ICU readmission using grouped physiological and medication trends, Artif. Intell. Med. 95 (2019) 27–37.
[45] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, second ed., Lawrence Erlbaum Associates, Hillsdale, *NJ*, 1988, pp. 1–17.
[46] I. Gandin, A. Scagnetto, S. Romani, G. Barbati, Interpretability of time-series deep learning models: a study in cardiovascular patients admitted to Intensive care unit, J. Biomed. Inf. 121 (2021), 103876.

Received: 18 December 2022 | Revised: 19 March 2023 | Accepted: 15 April 2023

DOI: 10.1111/liv.15591

Liver INTERNATIONAL WILEY

ORIGINAL ARTICLE

Metabolic-associated fatty liver disease in relation to site-specific and multiple-site subclinical atherosclerosis

Xinyu Wang[1] | Ruosu Zhang[1] | Sailimai Man[1,2,3] | Jun Lv[1,4,5] | Canqing Yu[1,4,5] | Jianchun Yin[6] | Xiaona Wang[7] | Yuhan Deng[2,8] | Bo Wang[2,3,4] | Liming Li[1,4,5] | Yuanjie Pang[1,5]

[1]Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing, China

[2]Meinian Institute of Health, Beijing, China

[3]Peking University Health Science Center, Meinian Public Health Institute, Beijing, China

[4]Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing, China

[5]Key Laboratory of Epidemiology of Major Diseases, Ministry of Education, Peking University, Beijing, China

[6]MJ Health Care Group, Shanghai, China

[7]Beijing MJ Health Check-up Center, Beijing, China

[8]Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

Correspondence
Dr. Yuanjie Pang, Department of Epidemiology and Biostatistics, School of Public Health, Peking University, 38 Xueyuan Road, Beijing 100191, China.
Email: yuanjie_p@163.com

Dr. Bo Wang, Meinian Public Health Institute, Health Science Center, Peking University, 38 Xueyuan Road, Beijing 100191, China.
Email: paul@meinianresearch.com

Funding information
China Postdoctoral Science Foundation, Grant/Award Number: 2019TQ0008 and 2020M670071; Fundamental Research Funds for the Central Universities, Peking University Start-up Grant, Grant/Award

Abstract

Background & Aims: Non-alcoholic fatty liver disease (NAFLD) and the newly proposed metabolic-associated fatty liver disease (MAFLD) were each associated with subclinical atherosclerosis. However, there is limited evidence on risk of atherosclerosis in individuals who meet the criteria for one but not the other. We aimed to investigate the associations of MAFLD or NAFLD status with site-specific and multiple-site atherosclerosis.

Methods: This is a prospective cohort study involving 4524 adults within the MJ health check-up cohort. Logistic regression model was used to estimate odds ratios (ORs) and confidence intervals (CIs) for subclinical atherosclerosis (elevated carotid intima-media thickness [CIMT], carotid plaque [CP], coronary artery calcification [CAC] and retinal atherosclerosis [RA]) associated with MAFLD or NAFLD status, MAFLD subtypes and fibrosis status.

Results: MAFLD was associated with higher risks of elevated CIMT, CP, CAC and RA (OR: 1.41 [95% CI 1.18–1.68], 1.23 [1.02–1.48], 1.60 [1.24–2.08], and 1.79 [1.28–2.52], respectively), whereas NAFLD per se did not increase risk of atherosclerosis except for elevated CIMT. Individuals who met both definitions or the definition for MAFLD but not NAFLD had higher risk of subclinical atherosclerosis. Among MAFLD subtypes, MAFLD with diabetes had the highest risk of subclinical atherosclerosis, but the associations did not differ by fibrosis status. Stronger positive associations were observed of MAFLD with multiple-site than single-site atherosclerosis.

Conclusions: In Chinese adults, MAFLD was associated with subclinical atherosclerosis, with stronger associations for multiple-site atherosclerosis. More attention should be paid to MAFLD with diabetes, and MAFLD might be a better predictor for atherosclerotic disease than NAFLD.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; AVR, arteriole-to-venule ratio; BMI, body mass index; CAC, coronary artery calcification; CI, confidence interval; CIMT, carotid intima-media thickness; CP, carotid plaque; CRP, C-reactive protein; CT, computed tomography; CVD, cardiovascular disease; DM, diabetes mellitus; FBG, fasting blood glucose; FIB-4, fibrosis-4; GGT, gamma-glutamyl transferase.; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homeostasis model assessment of insulin resistance; HU, Hounsfield unit; IQR, interquartile range; LDL-C, low-density lipoprotein cholesterol; MAFLD, metabolic-associated fatty liver disease; NAFLD, non-alcoholic fatty liver disease; NHANES, the National Health and Nutrition Examination Survey; OR, odds ratio; OW, overweight; PLT, platelet; RA, retinal atherosclerosis; RRR, relative risk ratio; SD, standard deviation; WC, waist circumference.

Number: BMU2022PY014; Ministry of Science and Technology of the People's Republic of China, Grant/Award Number: 2020YFC2004703; Peking University Medicine Fund of Fostering Young Scholars' Scientific & Technological Innovation, Grant/Award Number: BMU2022RCZX022

Handling Editor: Luca Valenti

KEYWORDS
advanced fibrosis, metabolic-associated fatty liver disease, non-alcoholic fatty liver disease, subclinical atherosclerosis

1 | INTRODUCTION

Metabolic-associated fatty liver disease (MAFLD), which emphasizes the presence of metabolic dysfunction in definition, is an updated nomenclature of non-alcoholic fatty liver disease (NAFLD).[1] While most patients with hepatic steatosis fit the definitions for both MAFLD and NAFLD, a considerable proportion meets the criteria for one but not the other.[2] Individuals who meet the definition of NAFLD but not MAFLD have little metabolic dysfunction, whereas those with MAFLD but not NAFLD have metabolic dysfunction and other chronic liver conditions or excessive alcohol use. Comparing the prognosis of the two discordant groups may provide insights into the necessity of re-definition.

Cardiovascular disease (CVD) is one of the leading causes of death in patients with NAFLD.[3] Since subclinical atherosclerosis is predictive for future cardiovascular events depending on the location and the number of sites affected, examining the associations of MAFLD or NAFLD status with site-specific and multiple-site subclinical atherosclerosis may inform the early detection of cardiovascular risk in patients with MAFLD or NAFLD.[4–7] In addition, a recent prospective cohort study conducted in the United States reported that MAFLD, but not NAFLD, was associated with all-cause mortality independently of known metabolic risk factors.[8] Investigating the associations with atherosclerosis might help to identify the potential mechanism under the differences in risks of death between MAFLD and NAFLD.

Cross-sectional and prospective cohort studies have shown consistently positive associations of MAFLD or NAFLD, respectively, with subclinical atherosclerosis (including carotid intima-media thickness [CIMT], carotid plaque [CP], coronary artery calcification [CAC], etc.).[9–15] However, it is not clear whether the discordant groups that meet the criteria for one but not the other definition have the same magnitude of associations with subclinical atherosclerosis, particularly for CP and CAC. Moreover, there is little evidence on the prospective associations of MAFLD or NAFLD with multiple-site atherosclerosis or whether the associations differ by advanced fibrosis or subtypes of MAFLD.

Therefore, we aimed to (1) examine the associations between MAFLD or NAFLD status and site-specific subclinical atherosclerosis; (2) compare the associations of MAFLD or NAFLD status with multiple-site atherosclerosis and that with single-site atherosclerosis; and (3) examine whether the associations of MAFLD with subclinical atherosclerosis differed by subtypes of MAFLD and advanced fibrosis.

Key points

In Chinese adults, metabolic-associated fatty liver disease (MAFLD) showed stronger associations with subclinical atherosclerosis than non-alcoholic fatty liver disease (NAFLD), suggesting that it is a better indicator for atherosclerotic disease. Patients with diabetes-accompanied MAFLD had the highest risks of subclinical atherosclerosis among all MAFLD subtypes. Our study highlighted the necessity of early detection for MAFLD and supported the redefinition for NAFLD.

2 | MATERIALS AND METHODS

2.1 | Study population

This is a prospective cohort study in the MJ Health Check-up study, which is an open cohort of participants who underwent a comprehensive annual health check-up at the clinic of the MJ Health Check-up Center in Beijing, China. Initially, we included adults aged ≥18 years with baseline health data collected from January 1, 2009, to December 31, 2018 ($n = 161\,681$). Participants who did not perform carotid artery ultrasound, chest computed tomography (CT) scan or fundus examination ($n = 139\,563$), had missing data on MAFLD or NAFLD status ($n = 3559$) or had elevated CIMT, CP, CAC, retinal atherosclerosis (RA) or history of CVD ($n = 5660$) at baseline were excluded (Figure S1). We further excluded participants who lost to follow up ($n = 7282$) or had missing data on CIMT, CP, CAC or RA ($n = 1093$) during the follow-up period. Finally, 4524 participants were included in the analyses.

2.2 | Diagnosis of NAFLD and MAFLD

The diagnosis of fatty liver was based on abdominal B-type ultrasound examination. Participants were required to fast before the examination. An Aplio 300 colour Doppler ultrasound scanner (Canon, Japan) with a probe frequency of 3.5 MHz was used in the ultrasound examination. The presence of fatty liver was determined if the participants met the ultrasound criteria for fatty liver.[16] In the questionnaire, each individual was asked about the frequency of

drinking alcohol, the type of alcohol (a reflection of alcohol content) and the number of cups per drink. The daily total alcohol consumption was calculated by multiplying the frequency by the alcohol content and the number of cups per drink. The definition of NAFLD was the presence of fatty liver in the absence of excessive alcohol consumption (<20 g/day for females and <30 g/day for males) or other identifiable causes.[17,18]

MAFLD was diagnosed as the presence of fatty liver with ≥1 of the following: (1) overweight (OW) or obese (body mass index [BMI] ≥23 kg/m^2), (2) diabetes mellitus (DM, fasting blood glucose [FBG] ≥7.0 mmol/L or self-reported diabetes mellitus or using anti-diabetic drugs) and (3) at least 2 metabolic risk abnormalities. Metabolic risk abnormalities consisted of (1) waist circumference (WC) ≥90 cm for men and ≥80 cm for women, (2) blood pressure ≥130/85 mmHg or specific drug treatment, (3) fasting plasma triglycerides ≥1.70 mmol/L or specific drug treatment, (4) plasma high-density lipoprotein (HDL) cholesterol <1.0 mmol/L for men and <1.3 mmol/L for women or specific drug treatment, (5) prediabetes (FBG 5.6–6.9 mmol/L), (6) homeostasis model assessment of insulin resistance (HOMA-IR) score ≥2.5 and (7) plasma high-sensitivity C-reactive protein (CRP) level >2 mg/L.[1] Individuals with MAFLD were further categorized into three mutually exclusive subtypes: (1) MAFLD with DM (DM-MAFLD), (2) overweight or obese MAFLD without diabetes (OW-MAFLD) and (3) lean or normal-weight MAFLD with ≥2 metabolic risk abnormalities but without DM (lean-MAFLD).[19]

The non-invasive fibrosis marker fibrosis-4 (FIB-4) was used to assess the probability of advanced fibrosis in all NAFLD or MAFLD events. The FIB-4 was calculated using the following equation: FIB-4 = age×aspartate aminotransferase (AST, U/L)/platelet (PLT) count (×10^9/L)×√alanine aminotransferase (ALT, U/L).[20] Fatty liver with intermediate or high probability of advanced fibrosis was identified if FIB-4 ≥ 1.30.[20]

2.3 | Diagnosis of subclinical atherosclerosis

Details for diagnoses of subclinical atherosclerosis, including elevated CIMT, CP, CAC and RA, can be found in Supplementary methods. Multiple-site atherosclerosis was defined by the presence of atherosclerotic lesions in more than one of the examined territories among the carotid artery, the coronary artery and the retinal arteriole.

2.4 | Statistical analysis

Continuous variables with normal or skewed distribution were described as the mean ± standard deviation (SD) or median (interquartile range [IQR]), and categorical variables were presented as frequencies and percentages. Follow-up was calculated from baseline (the date of the first health check-up) until the date of the outcome or the last follow-up (before December 31, 2018), whichever came first. Logistic regression model was used to calculate the odds ratios (ORs) and 95% confidence intervals (CIs) for incident subclinical atherosclerosis. Multinomial logistic regression model was used to assess the associations between MAFLD or NAFLD status and the number of affected sites, with no atherosclerosis as the reference category. All models were adjusted for age, gender, education level, income level, smoking status, alcohol consumption and physical activity according to previously reported risk factors for both NAFLD and subclinical atherosclerosis.[21–23]

All statistical analyses were performed using R (version 3.6.2; R Foundation for Statistical Computing, Vienna, Austria). All p values were two-sided, and statistical significance was defined as $p < 0.05$.

3 | RESULTS

3.1 | Baseline characteristics of study participants

The baseline characteristics of study participants by MAFLD or NAFLD status and incident elevated CIMT status are presented in Table 1 and Table S1, respectively. Among the 4524 study participants, the mean (SD) age at baseline was 45.3 (7.0) years, and 2716 (60.0%) were male. The prevalence of NAFLD, MAFLD, and concordant MAFLD and NAFLD was 33.0%, 43.4% and 32.4%, respectively (Figure 1). In terms of discordant groups, the prevalence of MAFLD only and NAFLD only was 11.0% and 0.6%, while that of simple hepatic steatosis (those who had hepatic steatosis but did not meet the definitions for MAFLD or NAFLD) was 0.2%.

Compared with those with normal liver, participants with MAFLD or NAFLD at baseline were older and more likely to be male, current smokers and current drinkers. Moreover, participants with MAFLD had a less healthy metabolic profile and were more likely to have prevalent diabetes and hypertension.

3.2 | Associations between MAFLD or NAFLD status and subclinical atherosclerosis

During a median follow-up of 2.0 years, there were 709, 578, 287 and 177 cases of elevated CIMT, CP, CAC and RA, respectively. After adjustments for demographic and lifestyle risk factors, MAFLD demonstrated positive associations with all four markers of subclinical atherosclerosis (OR: 1.41 [95% CI, 1.18–1.68] for elevated CIMT, 1.23 [1.02–1.48] for CP, 1.60 [1.24–2.08] for CAC and 1.79 [1.28–2.52] for RA, Table 2). The corresponding ORs of NAFLD were 1.36 (1.14–1.63), 1.08 (0.89–1.31), 1.39 (1.07–1.79) and 1.22 (0.88–1.70), respectively. When additionally adjusted for diabetes and hypertension, NAFLD was no associated with CAC (Table S2).

When cohorts were classified according to the absence and/or presence of MAFLD and NAFLD, MAFLD+NAFLD was associated with higher risks of subclinical atherosclerosis (OR: 1.44 [1.20–1.74] for elevated CIMT, 1.16 [0.95–1.42] for CP, 1.59 [1.20–2.10] for CAC and 1.56 [1.09–2.25] for RA). In particular, there were positive associations between MAFLD only and each marker of subclinical

TABLE 1 Baseline characteristics of study participants.

	Normal liver (*n* = 2526)	Simple hepatic steatosis (*n* = 9)	NAFLD only (*n* = 25)	MAFLD only (*n* = 497)	MAFLD + NAFLD (*n* = 1467)	Overall (*n* = 4524)
Age, year[a]	44.7 (6.9)	44.0 (6.8)	46.0 (7.5)	45.3 (6.2)	46.4 (7.1)	45.3 (7.0)
Female (*n*, %)	1364 (54.0)	2 (22.2)	12 (48.0)	25 (5.0)	405 (27.6)	1808 (40.0)
College education (*n*, %)	759 (30.0)	1 (11.1)	7 (28.0)	141 (28.4)	421 (28.7)	1329 (29.4)
Current smokers (*n*, %)	481 (19.0)	3 (33.3)	8 (32.0)	247 (49.7)	381 (26.0)	1120 (24.8)
Current drinkers (*n*, %)	615 (24.3)	6 (66.7)	8 (32.0)	412 (82.9)	370 (25.2)	1411 (31.2)
Inactive (*n*, %)	508 (20.1)	4 (44.4)	7 (28.0)	105 (21.1)	305 (20.8)	929 (20.5)
BMI, kg/m^2[a]	22.9 (2.6)	22.4 (0.5)	21.7 (1.4)	27.1 (2.7)	26.6 (2.9)	24.5 (3.3)
WC, cm[a]	78.1 (8.3)	79.4 (7.1)	76.1 (5.0)	92.3 (7.6)	89.7 (8.2)	83.4 (10.2)
SBP, mmHg[a]	109.8 (13.4)	107.6 (10.0)	107.2 (14.5)	119.1 (13.6)	118.0 (13.9)	113.5 (14.2)
DBP, mmHg[a]	69.6 (10.3)	67.9 (7.9)	67.6 (8.1)	77.3 (10.8)	75.3 (10.5)	72.3 (10.9)
FBG, mmol/L[a]	5.5 (0.6)	5.6 (0.5)	5.4 (0.3)	6.1 (1.2)	5.9 (1.0)	5.7 (0.9)
TC, mmol/L[a]	4.7 (0.8)	4.9 (0.7)	5.1 (0.9)	5.0 (0.9)	5.0 (0.9)	4.8 (0.9)
TG, mmol/L[a]	1.2 (0.7)	1.1 (0.4)	1.2 (0.5)	2.2 (1.6)	1.9 (1.2)	1.5 (1.1)
HDL-C, mmol/L[a]	1.5 (0.4)	1.5 (0.3)	1.5 (0.3)	1.2 (0.3)	1.2 (0.3)	1.4 (0.4)
LDL-C, mmol/L[a]	2.9 (0.8)	3.1 (0.6)	3.2 (1.0)	3.1 (0.8)	3.2 (0.8)	3.0 (0.8)
AST, U/L[b]	18.0 (6.0)	20.0 (2.0)	21.0 (6.0)	22.0 (9.0)	20.0 (7.0)	19.0 (7.0)
ALT, U/L[b]	16.0 (9.0)	25.0 (19.0)	22.0 (15.0)	29.0 (19.0)	25.0 (17.0)	19.0 (14.0)
GGT, U/L[b]	16.0 (13.0)	21.0 (12.0)	21.0 (16.0)	43.0 (45.0)	28.0 (23.0)	21.0 (22.0)
PLT count, (*10^9/L)[a]	228.8 (55.0)	233.4 (48.0)	240.6 (64.7)	220.7 (49.4)	233.2 (53.6)	229.4 (54.1)
Diabetes (*n*, %)	53 (2.1)	0 (0)	0 (0)	61 (12.3)	148 (10.1)	262 (5.8)
Hypertension (*n*, %)	368 (14.6)	0 (0)	3 (12.0)	193 (38.8)	499 (34.0)	1063 (23.5)

Note: Categorical variables are presented as frequencies and percentages.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CIMT, carotid intima-media thickness; DBP, diastolic blood pressure; FBG, fasting blood glucose; GGT, gamma-glutamyl transferase; HDL-C, high-density lipoprotein cholesterol; IQR, interquartile range; LDL-C, low-density lipoprotein cholesterol; PLT, platelet; SBP, systolic blood pressure; SD, standard deviation; TC, total cholesterol; TG, triglycerides; WC, waist circumference.

[a]Data are presented as the mean (SD).

[b]Data are presented as median (IQR).

atherosclerosis (OR: 1.42 [1.06–1.89] for elevated CIMT, 1.43 [1.05–1.92] for CP, 1.78 [1.19–2.64] for CAC and 2.84 [1.71–4.70] for RA). However, these associations were not observed in those with NAFLD only (OR: 0.99 [0.23–2.97] for CP, 1.94 [0.30–6.90] for CAC and 1.19 [0.06–6.64] for RA) except for elevated CIMT (OR: 3.10 [1.15, 7.55]).

3.3 | Associations between MAFLD or NAFLD status and the number of sites affected by subclinical atherosclerosis

Compared with those without MAFLD, the relative risk ratios (RRRs) for 1-site, 2-site and 3-site atherosclerosis were 1.28 (1.07–1.53), 1.73 (1.19–2.51) and 3.96 (1.40–11.12) among participants with MAFLD, respectively (Table 3). NAFLD was only associated with 1-site atherosclerosis (RRR: 1.21 [1.02–1.45]). The corresponding RRRs were 1.28 (1.06–1.54), 1.55 (1.03–2.32) and 3.27 (0.96–11.15) among those with MAFLD+NAFLD. MAFLD only was associated with increased risks of multiple-site atherosclerosis (RRR: 2.44 [1.42–4.19] for 2-site and 20.73 [4.71–91.43] for 3-site), while NAFLD only was only associated with 3-site atherosclerosis (RRR: 23.05 [1.38–384.61]).

3.4 | Associations between MAFLD subtypes and subclinical atherosclerosis

Among those with MAFLD, 209 (10.6%) participants were classified as DM-MAFLD, 1672 (85.1%) as OW-MAFLD and 83 (4.2%) as lean-MAFLD at baseline (Table 4). Compared with the non-MAFLD group, the ORs of the DM-MAFLD group were 1.99 (1.40–2.78) for elevated CIMT, 1.83 (1.27–2.60) for CP, 3.15 (2.04–4.77) for CAC, and 3.10 (1.79–5.23) for RA, respectively. The corresponding ORs were 1.40 (1.17–1.68), 1.18 (0.97–1.44), 1.40 (1.07–1.85) and 1.68 (1.18–2.40) in the OW-MAFLD group. No significant associations between lean-MAFLD and subclinical atherosclerosis were observed compared to non-MAFLD except for elevated CIMT.

Compared to participants with DM-MAFLD, those with OW-MAFLD or lean-MAFLD had lower risks of subclinical atherosclerosis (Table S3). For example, for OW-MAFLD, the ORs were 0.69

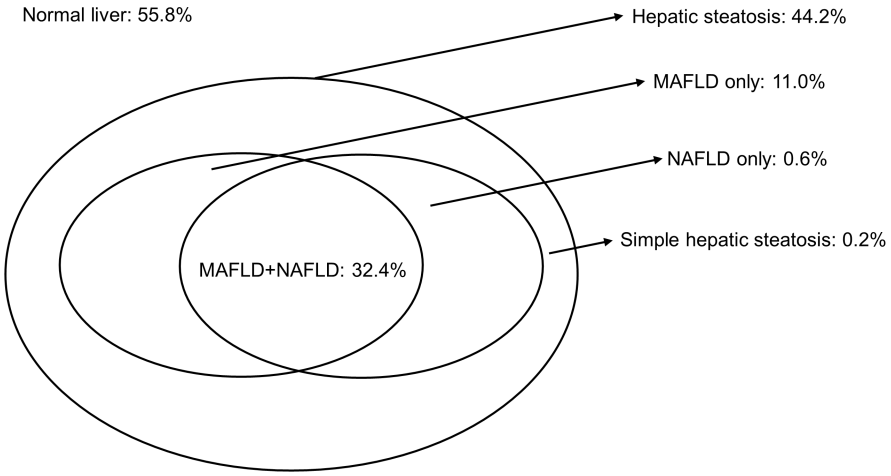


FIGURE 1 Relative proportions of MAFLD and NAFLD in the study participants. MAFLD+NAFLD denotes individuals who met both MAFLD and NAFLD definitions. In terms of discordant groups, MAFLD only denotes individuals who met the definition of MAFLD but not NAFLD, whereas NAFLD only denotes those with NAFLD but not MAFLD. Among individuals with hepatic steatosis diagnosed by abdominal ultrasound, we defined simple hepatic steatosis as individuals who met neither the definition for MAFLD nor NAFLD (i.e. individuals with steatosis but had neither metabolic risk factors to be diagnosed with MAFLD nor absence of other fatty liver causes to be diagnosed with NAFLD).

TABLE 2 Associations between MAFLD or NAFLD status and subclinical atherosclerosis.

	Elevated CIMT		CP		CAC		RA	
	No. cases	OR (95% CI)	No. cases	OR (95% CI)	No. cases	OR (95% CI)	No. cases	OR (95% CI)
NAFLD								
Non-NAFLD	396	1.00	357	1.00	159	1.00	101	1.00
NAFLD	313	1.36 (1.14, 1.63)	221	1.08 (0.89, 1.31)	128	1.39 (1.07, 1.79)	76	1.22 (0.88, 1.70)
MAFLD								
Non-MAFLD	304	1.00	269	1.00	111	1.00	66	1.00
MAFLD	405	1.41 (1.18, 1.68)	309	1.23 (1.02, 1.48)	176	1.60 (1.24, 2.08)	111	1.79 (1.28, 2.52)
MAFLD & NAFLD[a]								
Normal liver	295	1.00	266	1.00	108	1.00	65	1.00
NAFLD only	7	3.10 (1.15, 7.55)	3	0.99 (0.23, 2.97)	2	1.94 (0.30, 6.90)	1	1.19 (0.06, 6.64)
MAFLD only	99	1.42 (1.06, 1.89)	91	1.43 (1.05, 1.92)	50	1.78 (1.19, 2.64)	36	2.84 (1.71, 4.70)
MAFLD+NAFLD	306	1.44 (1.20, 1.74)	218	1.16 (0.95, 1.42)	126	1.59 (1.20, 2.10)	75	1.56 (1.09, 2.25)

Note: Logistic regression model was used in this analysis, adjusted for age, gender, education, income, smoking, drinking and physical activity.
Abbreviations: CAC, coronary artery calcification; CI, confidence interval; CIMT, carotid intima-media thickness; CP, carotid plaque; MAFLD, metabolic-associated fatty liver disease; NAFLD, non-alcoholic fatty liver disease; OR, odds ratio; RA, retinal atherosclerosis.
[a]The group of simple hepatic steatosis was analysed as a single group but not reported because of the wide CIs.

(0.49–0.97) for elevated CIMT, 0.63 (0.44–0.91) for CP, 0.44 (0.29–0.68) for CAC and 0.54 (0.33–0.92) for RA.

3.5 | Associations between advanced fibrosis and subclinical atherosclerosis in participants with MAFLD

When participants with MAFLD were stratified by risk of advanced fibrosis, no significant associations were observed between intermediate or high probability of advanced fibrosis and subclinical atherosclerosis (Table S4). For example, for elevated FIB-4, the ORs were 1.13 (0.78–1.62) for elevated CIMT, 1.08 (0.73–1.57) for CP, 0.82 (0.49–1.35) for CAC and 0.87 (0.49–1.49) for RA.

4 | DISCUSSION

In the current study, individuals with MAFLD had higher risk of developing subclinical atherosclerosis, with stronger associations for

TABLE 3 Associations between MAFLD or NAFLD status and number of sites affected by subclinical atherosclerosis.

	1-site vs no atherosclerosis		2-site vs no atherosclerosis		3-site vs no atherosclerosis	
	No. cases	RRR (95% CI)	No. cases	RRR (95% CI)	No. cases	RRR (95% CI)
NAFLD						
Non-NAFLD	422	1.00	78	1.00	13	1.00
NAFLD	281	1.21 (1.02, 1.45)	57	1.24 (0.86, 1.79)	10	1.33 (0.55, 3.19)
MAFLD						
Non-MAFLD	331	1.00	50	1.00	5	1.00
MAFLD	372	1.28 (1.07, 1.53)	85	1.73 (1.19, 2.51)	18	3.96 (1.40, 11.12)
MAFLD & NAFLD[a]						
Normal liver	329	1.00	49	1.00	4	1.00
NAFLD only	1	0.26 (0.03, 1.94)	1	1.81 (0.23, 14.22)	1	23.05 (1.38, 384.61)
MAFLD only	92	1.22 (0.91, 1.63)	29	2.44 (1.42, 4.19)	9	20.73 (4.71, 91.43)
MAFLD+NAFLD	280	1.28 (1.06, 1.54)	56	1.55 (1.03, 2.32)	9	3.27 (0.96, 11.15)

Note: Multi-nominal logistic regression model was used in this analysis, adjusted for age, gender, education, income, smoking, drinking and physical activity.

Abbreviations: CI, confidence interval; MAFLD, metabolic-associated fatty liver disease; NAFLD, non-alcoholic fatty liver disease; RRR, relative risk ratio.

[a]The group of simple hepatic steatosis was analysed as a single group but not reported because of the wide CIs.

TABLE 4 Associations between MAFLD subtypes and subclinical atherosclerosis.

		Elevated CIMT		CP		CAC		RA	
	No. participants	No. cases	OR (95% CI)	No. cases	OR (95% CI)	No. cases	OR (95% CI)	No. cases	OR (95% CI)
Non-MAFLD	2560	304	1.00	269	1.00	111	1.00	66	1.00
DM-MAFLD	209	62	1.99 (1.40, 2.78)	51	1.83 (1.27, 2.60)	37	3.15 (2.04, 4.77)	24	3.10 (1.79, 5.23)
OW-MAFLD	1672	336	1.40 (1.17, 1.68)	249	1.18 (0.97, 1.44)	131	1.40 (1.07, 1.85)	85	1.68 (1.88, 2.40)
Lean-MAFLD	83	7	0.47 (0.19, 0.99)	9	0.79 (0.36, 1.55)	8	1.81 (0.77, 3,71)	2	0.68 (0.11, 2.32)

Note: Logistic regression model was used in this analysis, adjusted for age, gender, education, income, smoking, drinking and physical activity.

Abbreviations: CAC, coronary artery calcification; CI, confidence interval; CIMT, carotid intima-media thickness; CP, carotid plaque; MAFLD, metabolic-associated fatty liver disease; OR, odds ratio; RA, retinal atherosclerosis.

multiple-site atherosclerosis. MAFLD only, but not NAFLD only, was positively associated with subclinical atherosclerosis (except for the positive association between NAFLD only and elevated CIMT). Moreover, in individuals with MAFLD, those with DM-MAFLD had higher risk of subclinical atherosclerosis compared with those with OW-MAFLD or lean-MAFLD. Our study highlighted the importance of early detection and interventions of MAFLD to prevent atherosclerosis.

Our findings for MAFLD were generally consistent with two prospective studies in China. One showed that compared with participants without MAFLD throughout the follow-up period, those with stable MAFLD had a higher risk of elevated CIMT (OR: 1.36 [1.13–1.62]), while the other reported that MAFLD was associated with a higher risk of CP (OR: 2.36 [1.12, 4.96]).[13,24] Our study extended results from previous studies by showing stronger associations for CAC and RA than for elevated CIMT and CP. Since CAC is considered a better predictor of subsequent CVD events than carotid atherosclerosis, our results emphasized the significance of examining coronary artery in patients with MAFLD.[25,26] In terms of NAFLD, a recent systematic review and meta-analysis involving 26 studies and 85 395 participants reported that the presence of NAFLD was associated with higher risks of increased CIMT or CP (OR: 1.74 [1.47–2.06]; between-study heterogeneity $I^2 = 86\%$) and CAC (OR: 1.56 [1.24–1.96]; $I^2 = 59\%$), while a cross-sectional study in China reported that NAFLD was associated with a higher risk of retinal athersclerosis.[27,28] In contrast to previous studies, our study found no associations of NAFLD with CP and RA. Since the meta-analysis mainly consisted of cross-sectional studies, more evidence is needed on the prospective associations of NAFLD with CP and RA. In addition, two retrospective cohort studies in Korea and China (sample size: 14288 and 8020 participants; follow-up duration: 5.0 and 3.3 years, respectively) reported positive associations between advanced fibrosis and subclinical atherosclerosis in NAFLD.

However, we observed null associations between fibrosis and atherosclerosis, possibly due to the relatively short follow-up period and small sample size.[10,29]

A prospective cohort study in the National Health and Nutrition Examination Survey (NHANES) III reported that MAFLD was associated with a higher risk of all-cause mortality, but NAFLD per se (additionally adjusted for metabolic risk factors) was not. In support of this study, the positive association of NAFLD with CAC in our study disappeared with additional adjustment for diabetes and hypertension, while results for MAFLD persisted. Furthermore, MAFLD only was positively associated with four markers of subclinical atherosclerosis, whereas NAFLD only was only associated with elevated CIMT. This suggested that atherosclerosis might be the underneath pathway through which MAFLD and NAFLD were linked to higher risk of all-cause mortality.[8] The positive association between NAFLD only and elevated CIMT might be because participants in this group developed metabolic dysfunction during the follow-up, and elevated CIMT tended to appear earlier in the course of atherosclerosis than other markers.[30] Future studies using a longitudinal approach are warranted to explore the relationships between transitions of liver status and subclinical atherosclerosis.

Another finding of this study is that associations of MAFLD with subclinical atherosclerosis were stronger for multiple-site atherosclerosis, which indicated that atherosclerotic disease in patients with MAFLD might have higher probability of concurrence. In a cross-sectional study in France involving 2554 adults, the presence of steatosis was associated with at least 2-site atherosclerosis (OR: 1.21 [1.01–1.45]).[9] Our study confirmed the association using a prospective cohort design and emphasized the necessity of a comprehensive evaluation of atherosclerosis in individuals with MAFLD.

Among the MAFLD subtypes, DM-MAFLD conferred the highest risk of subclinical atherosclerosis, which was generally consistent with a Korean study examining the associations of MAFLD subtypes and long-term cardiovascular events.[19] Our study suggested that DM-MAFLD subtype could serve as a strong predictor not only for future cardiovascular events but also for subclinical atherosclerosis. However, our results did not show higher risk of subclinical atherosclerosis in lean-MAFLD group. Since the high CVD risk of lean-MAFLD in the Korean study was mainly driven by the stronger association among those with advanced fibrosis, our null findings may be due to the relatively low fibrosis burden in participants with MAFLD in the current study.[19]

The strengths of the current study included a prospective study design, detailed categorization of MAFLD or NAFLD status and coverage of carotid arteries, coronary arteries and retinal arterioles for the evaluation of subclinical atherosclerosis. However, several limitations must be stated. First, the diagnoses of hepatic steatosis and advanced fibrosis were based on abdominal ultrasound and non-invasive biomarkers, respectively, rather than liver biopsy which is considered the gold standard.[31] However, ultrasonography is widely used in epidemiological studies for its convenience, cost-effectiveness and non-invasiveness, which has been proven to have high reliability and accuracy for hepatic steatosis diagnosis.[32] In addition, FIB-4 has been externally validated and accepted by clinical practice guidelines.[33] Second, NAFLD only accounted for a small proportion (0.6%) of the study participants. Therefore, the results should be interpreted with caution. Third, the study population involved participants attending health check-ups who had relatively higher levels of income and education, limiting the generalizability of the results. Finally, although demographic and lifestyle risk factors were adjusted for, residual confounding due to unmeasured or unknown variables cannot be ruled out.

In conclusion, MAFLD was positively associated with site-specific and multiple-site subclinical atherosclerosis, and particular attention should be paid to those with DM-MAFLD. The definition of MAFLD outperformed NAFLD to predict risk of subclinical atherosclerosis. The current study reinforced the necessity of early detection of MAFLD in preventing atherosclerosis and improving cardiovascular health.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation and data analysis were performed by Xinyu Wang, Ruosu Zhang and Yuanjie Pang. The first draft of the manuscript was written by Xinyu Wang and Yuanjie Pang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

FUNDING INFORMATION

This work was supported by the Peking University Medicine Fund of Fostering Young Scholars' Scientific & Technological Innovation (BMU2022RCZX022), the Fundamental Research Funds for the Central Universities, the Peking University Start-up Grant (BMU2022PY014), the China Postdoctoral Science Foundation (2019TQ0008, 2020M670071), and the Ministry of Science and Technology of China (2020YFC2004703).

CONFLICT OF INTEREST STATEMENT

The authors have no relevant financial or non-financial interests to disclose.

ETHICS STATEMENT

The study was approved by the Institutional Review Board of Peking University Health Science Center (ID of the approval: IRB00001052-19077). The requirement for informed consent was waived due to the use of deidentified data obtained as part of routine check-ups.

ORCID

Xinyu Wang https://orcid.org/0000-0003-1794-510X
Sailimai Man https://orcid.org/0000-0003-3032-4113
Yuanjie Pang https://orcid.org/0000-0002-4826-8861

REFERENCES

Author names in bold designate shared co-first authorship.

1. Eslam M, Sanyal AJ, George J, International Consensus Panel. MAFLD: a consensus-driven proposed nomenclature for

metabolic associated fatty liver disease. *Gastroenterology*. 2020;158(7):1999-2014.
2. Wai-Sun Wong V, Kanwal F. On the proposed definition of metabolic-associated fatty liver disease. *Clin Gastroenterol and Hepatol*. 2021;19(5):865-870.
3. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73-84.
4. Willeit P, Tschiderer L, Allara E, et al. Carotid intima-media thickness progression as surrogate marker for cardiovascular risk. *Circulation*. 2020;142(7):621-642.
5. Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson KE. Coronary calcium score and cardiovascular risk. *J Am Coll Cardiol*. 2018;72(4):434-447.
6. Polak JF, Szklo M, Kronmal RA, et al. The value of carotid artery plaque and intima-media thickness for incident cardiovascular disease: the multi-ethnic study of atherosclerosis. *J Am Heart Assoc*. 2013;2(2). doi:10.1161/JAHA.113.000087
7. Laclaustra M, Casasnovas JA, Fernández-Ortiz A, et al. Femoral and carotid subclinical atherosclerosis association with risk factors and coronary calcium: the AWHS study. *J Am Coll Cardiol*. 2016;67(11):1263-1274.
8. Kim D, Konyn P, Sandhu KK, Dennis BB, Cheung AC, Ahmed A. Metabolic dysfunction-associated fatty liver disease is associated with increased all-cause mortality in the United States. *J Hepatol*. 2021;75(6):1284-1291.
9. Pais R, Redheuil A, Cluzel P, Ratziu V, Giral P. Relationship among fatty liver, specific and multiple-site atherosclerosis, and 10-year Framingham score. *Hepatology*. 2019;69(4):1453-1463.
10. Sinn DH, Cho SJ, Gu S, et al. Persistent nonalcoholic fatty liver disease increases risk for carotid atherosclerosis. *Gastroenterology*. 2016;151(3):481-488.
11. Targher G. Non-alcoholic fatty liver disease as driving force in coronary heart disease? *Gut*. 2017;66(2):213-214.
12. Park HE, Kwak MS, Kim D, Kim MK, Cha MJ, Choi SY. Nonalcoholic fatty liver disease is associated with coronary artery calcification development: a longitudinal study. *J Clin Endocrinol Metab*. 2016;101(8):3134-3143.
13. **Liu S**, **Wang J**, Wu S, et al. The progression and regression of metabolic dysfunction-associated fatty liver disease are associated with the development of subclinical atherosclerosis: a prospective analysis. *Metabolism*. 2021;120:154779.
14. **Wang J**, **Liu S**, Cao Q, et al. New definition of metabolic dysfunction-associated fatty liver disease with elevated brachial-ankle pulse wave velocity and albuminuria: a prospective cohort study. *Front Med*. 2022;16:714-722. doi:10.1007/s11684-021-0888-8
15. Lin TY, Chen YJ, Chen WL, Peng TC. The relationship between non-alcoholic fatty liver disease and retinopathy in NHANES III. *PLoS One*. 2016;11(11). doi:10.1371/journal.pone.0165970
16. Saadeh S, Younossi ZM, Remer EM, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology*. 2002;123(3):745-750.
17. Man S, Lv J, Yu C, et al. Association between metabolically healthy obesity and non-alcoholic fatty liver disease. *Hepatol Int*. 2022;16:1412-1423. doi:10.1007/s12072-022-10395-8
18. Fan JG, Wei L, Zhuang H, et al. Guidelines of prevention and treatment of nonalcoholic fatty liver disease (2018, China). *J Dig Dis*. 2019;20(4):163-173.
19. **Lee H**, **Lim TS**, Kim SU, Kim HC. Long-term cardiovascular outcomes differ across metabolic dysfunction-associated fatty liver disease subtypes among middle-aged population. *Hepatol Int*. 2022;16:1308-1317. doi:10.1007/s12072-022-10407-7
20. Shah AG, Lydecker A, Murray K, et al. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol*. 2009;7(10):1104-1112.
21. Johnston PM, Patel J, Byrne DC. Causes of mortality in non-alcoholic fatty liver disease (NAFLD) and alcohol related fatty liver disease (AFLD). *Curr Pharm Des*. 2020;26(10):1079-1092.
22. Song P, Fang Z, Wang H, et al. Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study. *Lancet Glob Health*. 2020;8(5):e721-e729. doi:10.1016/S2214-109X(20)30117-0
23. Lechner K, von Schacky C, McKenzie AL, et al. Lifestyle factors and high-risk atherosclerosis: pathways and mechanisms beyond traditional risk factors. *Eur J of Prev Cardiol*. 2020;27(4):394-406.
24. **Wang X**, **Cheng S**, Lv J, et al. Liver biomarkers, genetic and lifestyle risk factors in relation to risk of cardiovascular disease in Chinese. *Front Cardiovasc Med*. 2022;9:938902.
25. Folsom AR, Kronmal RA, Detrano RC, et al. Coronary artery calcification compared with carotid intima-media thickness in the prediction of cardiovascular disease incidence: the multi-ethnic study of atherosclerosis (MESA). *Arch Intern Med*. 2008;168(12):1333-1339.
26. Gepner AD, Young R, Delaney JA, et al. Comparison of carotid plaque score and coronary artery calcium score for predicting cardiovascular disease events: the multi-ethnic study of atherosclerosis. *J Am Heart Assoc*. 2017;6(2):e005179. doi:10.1161/JAHA.116.005179
27. Zhou YY, Zhou XD, Wu SJ, et al. Nonalcoholic fatty liver disease contributes to subclinical atherosclerosis: a systematic review and meta-analysis. *Hepatol Commun*. 2018;2(4):376-392.
28. Yang W, Xu H, Yu X, Wang Y. Association between retinal artery lesions and nonalcoholic fatty liver disease. *Hepatol Int*. 2015;9(2):278-282.
29. **Yu X**, **Chen C**, Guo Y, et al. High NAFLD fibrosis score in non-alcoholic fatty liver disease as a predictor of carotid plaque development: a retrospective cohort study based on regular health check-up data in China. *Ann Med*. 2021;53(1):1621-1631.
30. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. *Nat Med*. 2011;17(11):1410-1422.
31. Brunt EM, Wong VWS, Nobili V, et al. Nonalcoholic fatty liver disease. *Nat Rev Dis Primers*. 2015;1(1):15080.
32. **Hernaez R**, **Lazo M**, Bonekamp S, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. *Hepatology*. 2011;54(3):1082-1090.
33. Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: clinical prediction rules and blood-based biomarkers. *J Hepatol*. 2018;68(2):305-315.

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: Wang X, Zhang R, Man S, et al. Metabolic-associated fatty liver disease in relation to site-specific and multiple-site subclinical atherosclerosis. *Liver Int*. 2023;43:1691-1698. doi:10.1111/liv.15591

HUMAN VACCINES & IMMUNOTHERAPEUTICS
2023, VOL. 19, NO. 2, 2222052
https://doi.org/10.1080/21645515.2023.2222052

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Assessing SARS-CoV-2 vaccine hesitancy among the people living with and without HIV from May to September 2022 in Blantyre, Malawi

Chifundo Mchawa[a,b#], Shan-Shan Zhang[c#], Wan-Xue Zhang[c], Yiguo Zhou[d], Ting-Ting Wei[e], Juan Du[e], Fuqiang Cui[a,c,d,e,f], and Qing-Bin Lu[a,c,d,e,f]

[a]Department of Laboratory Science and Technology & Vaccine Research Center, School of Public Health, Peking University, Beijing, China; [b]Department of HIV & AIDS, Populations Services International, Lilongwe, Malawi; [c]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China; [d]Department of Health Policy and Management, School of Public Health, Peking University, Beijing, China; [e]Global Center for Infectious Disease and Policy Research & Global Health and Infectious Diseases Group, Peking University, Beijing, China; [f]Ministry of Education, Key Laboratory of Epidemiology of Major Diseases (Peking University), Beijing, China

ABSTRACT

Vaccine hesitancy is a significant obstacle to the prevention and control of coronavirus disease 2019, especially among people with human immunodeficiency virus (HIV) in developing countries like Malawi, where HIV prevalence rate is high and limited data is available on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccine hesitancy among people living with HIV (PLHIV). This study was conducted among people aged ≥18 years at Mpemba health center, Blantyre. All PLHIV were interviewed using a structured questionnaire. All non-PLHIVs who were willing and were conveniently available were investigated. Multivariate logistic regression model and generalized linear model were used to assess the factors associated with SARS-CoV-2 vaccine hesitancy and knowledge, attitude, and trust. Totally 682 subjects were recruited with 341 PLHIV and 341 non-PLHIV. SARS-CoV-2 vaccine hesitancy rates were similar between PLHIV and non-PLHIV (56.0% vs. 57.2%, $p = .757$). In PLHIV, SARS-CoV-2 vaccine hesitancy was associated with education, occupation, and religion (all $p < .05$). In non-PLHIV, vaccine hesitancy was associated with sex, education, occupation, income, marital status, and residence (all $p < .05$). Higher knowledge, attitude, and trust scores were associated with a lower vaccine hesitancy rate in PLHIV (knowledge: OR = 0.79, 95% CI 0.65–0.97, $p = .022$; attitude: OR = 0.45, 95% CI 0.37–0.55, $p < .001$; trust: OR = 0.84, 95% CI 0.71–0.99, $p = .038$). SARS-CoV-2 vaccine hesitancy was high among PLHIV in Blantyre city, Malawi, which was a similar situation to non-PLHIV. Targeted efforts are needed to address these concerns and increase knowledge, trust, and positive attitudes toward the vaccine to reduce vaccine hesitancy against SARS-CoV-2 in PLHIV.

ARTICLE HISTORY
Received 26 March 2023
Revised 23 May 2023
Accepted 1 June 2023

KEYWORDS
Vaccine hesitancy; vaccination rate; SARS-CoV-2; people living with HIV; Malawi; knowledge; attitude; trust

Introduction

Vaccination is crucial in managing the pandemic of coronavirus disease 2019 (COVID-19); however, the scourge of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccine hesitancy is evident worldwide.[1] Therefore, deep diving into the factors driving vaccination uptake and the associated behaviors and attitudes is critical in expanding the coverage and subsequent controlling of the pandemic. Therefore, the use of innovative evidence-based interventions to address such gaps cannot be overemphasized. Malawi is a southern eastern African country with a population of about 17.5 million.[2] Up to 17 May 2023, there have been about 90,000 confirmed cases of COVID-19 and caused nearly three thousand deaths in Malawi.[3] During the COVID-19 pandemic, several epidemic peaks occurred in this country, that were June to August 2020, December 2020 to March 2021, June to September 2021, and December 2021 to February 2022, respectively, and the largest weekly confirmed cases was about seven thousands.[3] Moreover, in Malawi, the prevailing SARS-CoV-2 vaccination rates are alarmingly low to offer long-term control of the pandemic.[4] The SARS-CoV-2 vaccination rates for Malawi compared with other countries in the sub-Saharan region like South Africa, Kenya, and Mozambique are considerably lower, and far more from the majority level of the globe.[3,4] In this case, it is important to improve the situation of vaccination against SARS-CoV-2.

The human immunodeficiency virus (HIV) pandemic is mainly concentrated in the sub-Saharan Africa region, with approximately 25.5 million people living with HIV (PLHIV).[5] Malawi is one of the countries that worst influenced by the HIV epidemic with estimated adult HIV prevalence of 8.9% in 2019 nationwide and 17.7% in Blantyre City.[6] According to Malawi Population-based HIV Impact Assessment survey (2021), HIV prevalence in Malawi was still staggering at 9.2%, and 14.2% in Blantyre.[7] Research has shown that PLHIV are more susceptible to SARS-CoV-2 infection in

CONTACT Fuqiang Cui cuifuq@bjmu.edu.cn; Qing-Bin Lu qingbinlu@bjmu.edu.cn Department of Laboratorial Science and Technology & Vaccine Research Center, School of Public Health, Peking University, No. 38 Xue-Yuan Road, Haidian District, Beijing 100191, China.
[#]These authors contributed equally to this work.
Supplemental data for this article can be accessed on the publisher's website at https://doi.org/10.1080/21645515.2023.2222052.

comparison to general population and might exhibit worse outcomes compared with HIV-negative individuals.[8,9] Aside from the physical effects of SARS-CoV-2 infection, there are emotional and social aspects that affected even access to antiretroviral therapy (ART). For PLHIV, vaccine hesitancy might be more pronounced and therefore consequential poor SARS-CoV-2 vaccine coverage.[10] It is therefore imperative to protect PLHIV from SARS-CoV-2 infection. The extent of the SARS-CoV-2 vaccination in PLHIV in Malawi is still not known which unearths more research needs in this area.

Vaccine hesitancy might be widespread, particularly in specific groups such as PLHIV, and may consequently decrease SARS-CoV-2 vaccination rates.[11] Therefore, it is vital to comprehend the disposition to accept SARS-CoV-2 vaccination and its accompanying influences among PLHIV to mount evidence-based interventions that are effective in increasing the number of PLHIV vaccinated against SARS-CoV-2. Previous studies that explored SARS-CoV-2 vaccine willingness in PLHIV showed that around 70% of PLHIV were willing to get COVID-19 jabs in the United States and France.[12,13] On the contrary, a study in eight cities in China reported that 57.2% of PLHIV had a willingness for SARS-CoV-2 vaccination, lower than in the United States and France.[14] Lazarus et al. reported varied ranges of vaccine acceptance from almost 90% (in China) to less than 55% (in Russia) in 19 countries.[15] They declared that SARS-CoV-2 vaccine acceptance was 65.22% in Nigeria and it was 81.58% in South Africa. This evidence shows high levels of variability across different countries and multifaceted reasons could be behind this variability. To date, the levels of SARS-CoV-2 vaccine hesitancy in different marginalized groups in Malawi are still not known and more research studies are needed in different marginalized groups like the PLHIV. According to the fast-track cities study, Blantyre city is among the top 10 cities in the world with the highest HIV prevalence rate and Mpemba health center has the highest prevalence of HIV among the health facilities around Blantyre.[7,16] Hence, we conducted this cross-sectional study to figure out the status of SARS-CoV-2 vaccine hesitancy among PLHIV and explore the influencing factors from May to September 2022 in Blantyre, Malawi, and then put forward feasible measures to improve the situation based on the findings.

Methods

Study design and participants

A cross-sectional study was conducted from May to September 2022 at Mpemba health center in Traditional Authority Somba, in Blantyre district, which is a government-owned primary health care facility level institution and provides the following services: outpatient care, maternity, family planning, under-five care, ART clinic, and mobile clinics service for hard-to-reach places. Inclusion criteria of participants were as follows: All PLHIV aged ≥18 years old and accessing ART through the Mpemba ART clinic that were conveniently available at the time of the study were interviewed using a structured questionnaire. All non-PLHIVs aged ≥18 years who were willing and were conveniently available at the time the researcher conducted the study were also sampled. Exclusion criteria of participants were as follows: All PLHIV clients accessing ART through the Mpemba ART clinic without giving written consent were excluded from the study. All HIV-non-infected members without completed information were automatically excluded from the study. All individuals contraindicated from receiving SARS-CoV-2 vaccination were excluded.

Questionnaire

The questionnaire included the following parts: sociodemographic and health characteristics. ① Age, sex, educational attainment, level of income, marital status, years since HIV was diagnosed, and health status (whether they are satisfied with their current health status) were collected. ② The information on SARS-CoV-2 vaccination. This part focused on the history of the vaccination against SARS-CoV-2, whether they had encountered any adverse reactions to COVID-19, the reasons for willingness or hesitancy to SARS-CoV-2 vaccination, and whether the study participants or anybody in their home was diagnosed or succumbed to COVID-19. ③ The SARS-CoV-2 vaccine hesitancy scale referred to Zhao et al.[17] and consisted of three dimensions: vaccination knowledge (knowledge), vaccination attitude (attitude), and trust in confidence acquisition channels (trust). ④ The views of participants on the relationship between HIV and SARS-CoV-2 infection. An inquiry was made to establish whether study participants were vaccinated against SARS-CoV-2 or not, whether they had encountered any adverse reactions to COVID-19, and whether the study participants or anybody in their home was diagnosed or succumbed to COVID-19. The complete questionnaire was in Supplementary Materials and the scale score assignment was listed in Table S1. The overall Cronbach's α coefficient was 0.842, and the Cronbach's α coefficients of knowledge, attitude, and trust were 0.784, 0.832, and 0.940, respectively.

Data collection

Data were mainly collected using the structured questionnaire described above. The questionnaire was translated into Chichewa (Malawi's local dialect) to enhance good communication with the participants. The questionnaire was piloted by the researcher and 10 questionnaires were administered to the target population in the study. The pretesting focused on the following aspects; comprehension, wording, sequence, and general layout. Minor adjustments were made based on the feedback and a second pretest was done with a different group of respondents. No major feedback was given in the second piloting session and general observation was the ease of understanding the questions by the participants. Small mediums were defined as the people employing one to twenty employees according to the Micro, Small, and Medium Enterprises survey of Malawi in 2012.

Sampling and sample size

The researcher employed convenience sampling such that participants that were available in the ART clinic and members from the general population were sampled. The sample size was determined using the quick calculation formula for the cross-sectional study: $n = \frac{Z^2_{1-\alpha/2} \times pq}{d^2}$. In the formula, p is the expected prevalence rate, q = 1-p, d is the allowable error, $Z_{1-\alpha/2}$ is the statistic of significance test, and n is the sample size. We set the allowable error as 0.12×p, $\alpha = 0.05$, $Z_{1-\alpha/2} = 1.96$. Assuming the SARS-CoV-2 vaccine hesitancy was 55%, the sample size was calculated as 219. Considering that convenience sampling is not a probability sampling, we increased the sample size by 1/2, that was, 328 individuals were needed. Assuming vaccine hesitancy rates were similar in the two groups, the same number of samples would be investigated in the two groups with 656 subjects at least in total.

Ethical consideration

Ethical clearance and study approval were sought from the Blantyre district office by the researcher. Both the study approval and ethical clearance were granted by the Blantyre district health office and the ethics committee under the Blantyre district office before the commencement of data collection. Participants also gave consent to participate in the study by signing or thumb printing on consent forms. All data collected from the participants were treated with confidentiality as no names were used. All participants in the study were recruited on a voluntary basis and this notion was emphasized prior to the consent signing or thumb printing.

Definition of vaccine hesitancy against SARS-CoV-2

People who were vaccinated against SARS-CoV-2 and without any initial hesitancy were defined as no vaccine hesitancy, while people who were vaccinated but had vaccine hesitancy initially and those who were not vaccinated were defined as vaccine hesitancy.

Statistical analysis

Categorical variables were described by frequencies and proportions. Continuous variables were described as mean and standard deviation (SD). Descriptive statistics, including percentages and frequencies, were used to summarize the characteristics of the study participants. In addition, bivariate analysis using Chi-square was performed to examine the relationship between SARS-CoV-2 vaccine hesitancy and participant characteristics. A multivariate logistic regression model was conducted to evaluate factors that are associated with SARS-CoV-2 vaccine hesitancy, and the association between the scale scores and SARS-CoV-2 vaccine hesitancy. A multivariate analysis was conducted by using the generalized linear model (GLM) to influence factors of scale scores. The dependent variables were the scale scores of knowledge, attitude, and trust respectively, with the social demographic information as independent variables. A two-sided $p < .05$ was considered statistically significant. Data analysis was conducted using Microsoft Excel, STATA (version 17) and R 4.1.2 software.

Results

Demographic characteristics of the participants

There were 341 participants in the PLHIV group and 341 participants in the non-PLHIV group. The demographic characteristics and SARS-CoV-2 vaccination status in the non-PLHIV group and PLHIV group are displayed in Table 1. Compared to the non-PLHIV group, the PLHIV group had more males (41.6% vs. 33.7%, $p = .033$) and more rural residents (95.6% vs. 91.2%, $P = .021$). The distribution of age groups, education levels, occupation types, and monthly income were significantly different in the two groups (all $p < .05$). A higher education level and monthly income were observed in the non-PLHIV group, while the marital status and religion were comparable (both $p > .05$).

Comparison of vaccine hesitancy rate and vaccination rate

In the non-PLHIV group, 195 (57.2%) were had vaccine hesitancy and 146 (42.8%) had no vaccine hesitancy (Figure 1, Table S2). While 191 (56.0%) had vaccine hesitancy in the PLHIV group. Overall, the vaccine hesitancy status in people living with or without HIV infection was similar ($p = .757$). 173 (50.7%) participants were vaccinated in the non-PLHIV group and 168 (49.3%) were vaccinated in the PLHIV group. Overall, the vaccination status in people living with or without HIV infection was also comparable ($p = .702$) (Figure 1, Table S2).

Vaccine hesitancy rate in different subgroups of PLHIV

In PLHIV, the vaccine hesitancy rates against SARS-CoV-2 were associated with education, occupation, and religion (all $p < .05$) (Table 2). Participants who had primary and secondary education had the highest vaccination rate (43.3% and 46.7%, respectively) and the primary category had the highest vaccine hesitancy rate (53.4%). The participants who were occupationally farmers, professional/skilled laborers, and others had the top three highest vaccine hesitancy rates (48.2%, 21.5%, and 14.7%, respectively), while the unemployed had the lowest hesitancy rate (3.1%). The vaccine hesitancy rate of Christians was significantly higher than that of Christians (86.4% *vs.* 13.6%, $p = .038$) (Table 2).

Multivariate logistic regression model for SARS-CoV-2 vaccination hesitancy

A multivariate logistic regression model was conducted based on the demographic variables which had significant differences in univariate analysis in the PLHIV group. Meanwhile, age and sex were also included in the model as the key demographic characteristics. The multivariate logistic regression model showed that vaccine hesitancy was associated with age, education, occupation, and religion in the

Table 1. Demographic characteristics according to HIV status of respondents.

Characteristic, n (%)	Total N = 682	non-PLHIV n = 341	PLHIV n = 341	*p* value
Sex				.033
Female	425 (62.3)	226 (66.3)	199 (58.4)	
Male	257 (37.7)	115 (33.7)	142 (41.6)	
Age groups (years)				<.001
18–29	208 (30.5)	131 (38.4)	77 (22.6)	
30–39	293 (43.0)	127 (37.2)	166 (48.7)	
40–49	142 (20.8)	60 (17.6)	82 (24.0)	
Above 50	39 (5.7)	23 (6.7)	16 (4.7)	
Education				<.001
Never been to School	41 (6.0)	7 (2.1)	34 (10.0)	
Primary	291 (42.7)	124 (36.4)	167 (49.0)	
Secondary	306 (44.9)	175 (51.3)	131 (38.4)	
Tertiary	44 (6.5)	35 (10.3)	9 (2.6)	
Occupation				<.001
Farmers	247 (36.2)	91 (26.7)	156 (45.7)	
Small mediums	91 (13.3)	35 (10.3)	56 (16.4)	
Students	60 (8.8)	40 (11.7)	20 (5.9)	
Professional/Skilled workers	133 (19.5)	94 (27.6)	39 (11.4)	
Unemployed	104 (15.2)	49 (14.4)	55 (16.1)	
Other	47 (6.9)	32 (9.4)	15 (4.4)	
Monthly income				<.001
Less than K50,000	334 (49.0)	129 (37.8)	205 (60.1)	
K50,000-K100,000	167 (24.5)	99 (29.0)	68 (19.9)	
K100,000–500,000	150 (22.0)	88 (25.8)	62 (18.2)	
Above K500,000	31 (4.5)	25 (7.3)	6 (1.8)	
Marital status				.403
Single	204 (29.9)	97 (28.4)	107 (31.4)	
Married	478 (70.1)	244 (71.6)	234 (68.6)	
Residence				.021
Rural	637 (93.4)	311 (91.2)	326 (95.6)	
Urban	45 (6.6)	30 (8.8)	15 (4.4)	
Religion				.055
Christian	624 (91.5)	319 (93.5)	305 (89.4)	
Muslim	58 (8.5)	22 (6.5)	36 (10.6)	

HIV, human immunodeficiency virus; PLHIV, people living with HIV.

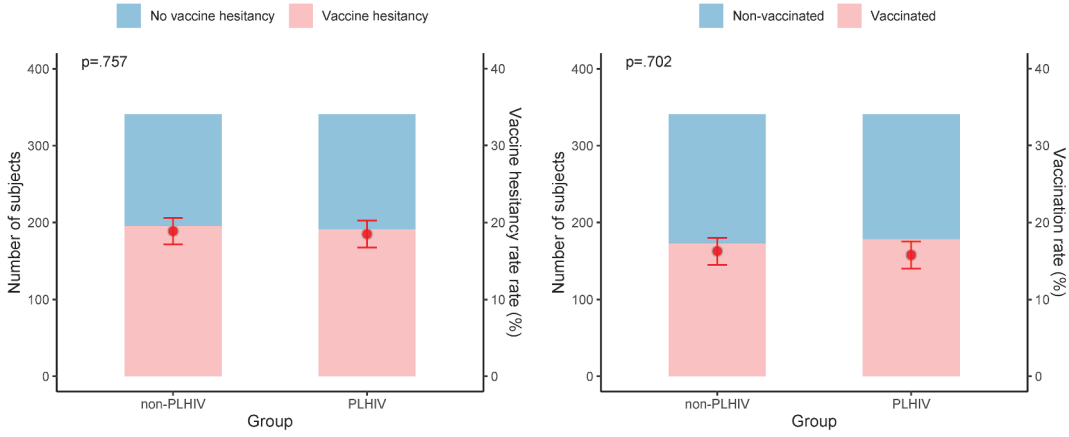


Figure 1. Comparison of vaccine hesitancy and vaccination rate between PLHIV and non-PLHIV. The bars represented the numbers, measured by the left y-axis. The dots represented the rates and the error bars represented the confidence intervals, measured by the right y-axis. PLHIV, people living with human immunodeficiency virus.

PLHIV group (all $p < .05$) (Figure 2, Table S3). Compared to the 18–29 years age group, respondents in the 40–49 years and above 50 years were less likely to have vaccine hesitancy against SARS-CoV-2 (40–49 years: OR = 0.46, 95% CI 0.22–0.96; above 50 years: OR = 0.23, 95% CI 0.07–0.80). Compared to the Never been school, respondents in the primary and secondary, and tertiary education categories were likely not to have vaccine hesitancy against SARS-CoV-2 (primary: OR = 0.29, 95% CI 0.11–0.76; secondary: 0.14, 95% CI 0.05–0.40; tertiary: OR = 0.02, 95% CI 0.01–0.22). Compared to farmers, participants in the small and medium enterprises were more likely to have vaccine hesitancy against SARS-CoV-2 (OR = 2.93, 95% CI 1.35–6.35). Compared to Christians, Muslims were more likely to have vaccine hesitancy against SARS-CoV-2 (OR = 2.67, 95% CI 1.09–6.53) (Table S3).

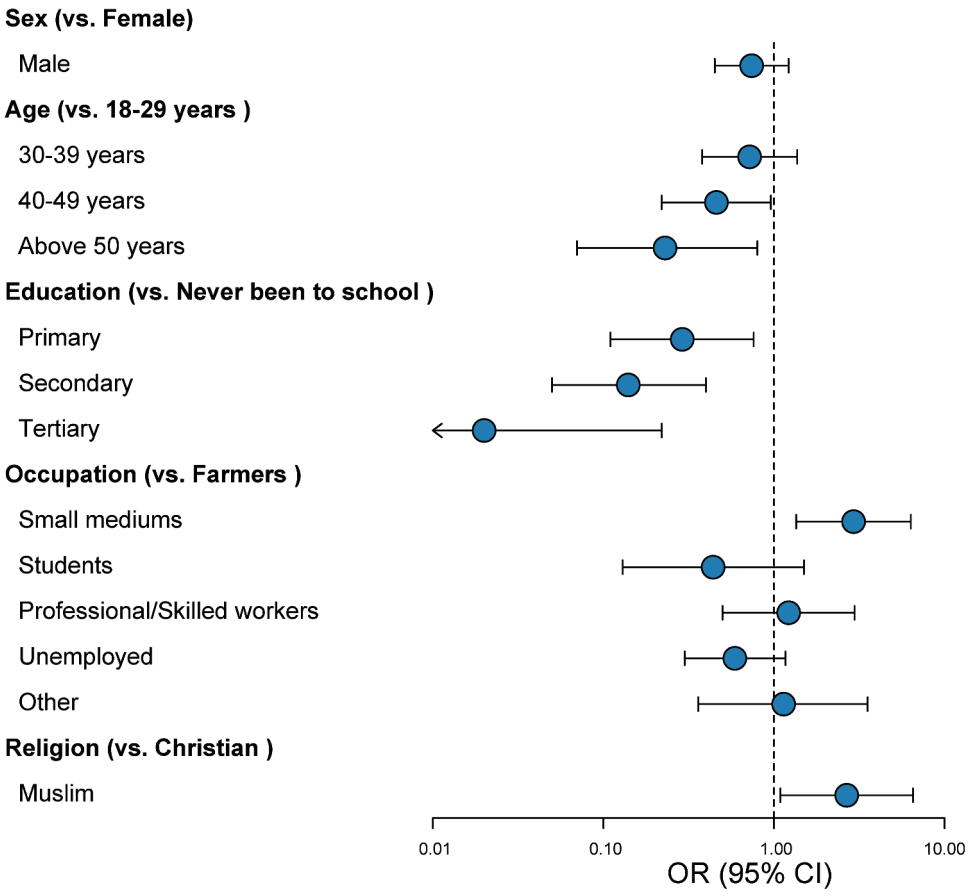


Figure 2. The association between vaccine hesitancy and demographics in PLHIV. The circle represented OR and the error bar represented 95% CI. OR, odds ratio. CI, confidence interval. PLHIV, people living with human immunodeficiency virus..

The association between knowledge, attitude, and trust scores and vaccine hesitancy

In the PLHIV group, the results showed that the status of vaccine hesitancy against SARS-CoV-2 was associated with the knowledge of SARS-CoV-2 vaccines and attitude toward the vaccination (Table 3). The higher the knowledge scores and the attitude scores were, the less likely the PLHIV had vaccine hesitancy against SARS-CoV-2 (knowledge: OR 0.80, 95% CI 0.67–0.97, $p = .013$; attitude: OR 0.62, 95% CI 0.53–0.72, $p < .001$).

GLMs of knowledge scale and demographic characteristics

The mean of the knowledge scores was 10.93 (SD 2.09) among all the participants and the knowledge level was associated with their education level, occupation, monthly income level, and residence in the PLHIV group ($p < .05$) (Table 4). Compared with participants never been to school, those with tertiary education had higher knowledge scores (β = 2.98, 95% CI 0.28–5.69, $p = .031$). Professional/skilled laborers had higher knowledge scores compared to farmers (β = 1.34, 95% CI 0.25–2.42, $p = .015$). The knowledge scores were lower in those with higher monthly income than those with a monthly income of less than K50,000 (K50,000-K100,000: β = −0.83, 95% CI −1.55, −0.11; K100,000–500,000: β = −1.22, 95% CI −2.18, −0.26; above K500,000: β = −3.85, 95% CI −7.04, −0.66; all $p < .05$). Participants residing in the urban area had a lower knowledge score than that in rural (β = −1.83, 95% CI −2.97, −0.68, $p = .002$) (Table 4).

GLMs of attitude scale and demographic characteristics

The mean of the attitude scores was 12.62 (SD 3.14) among all the participants and the attitude level was associated with their age, education level, occupation, monthly income level, residence, and religion in the PLHIV group ($p < .05$) (Table 5). Compared to participants aged 18–29 years, those aged 30–39 years, 40–49 years, and above 50 years had higher attitude scores (β = 0.95, 95% CI 0.09, 1.81; β = 1.6, 95% CI 0.61, 2.59; β = 2.64, 95% CI 1.01, 4.29). Compared with participants never been to school, those with secondary and tertiary education had higher attitude scores (β = 1.93, 95% CI 0.66, 3.19; β = 4.46, 95% CI 0.47, 8.46). Students, professionals/skilled workers, people unemployed, and others had higher scores compared to farmers (β = 2.39, 95% CI 0.64, 4.13; β = 2.63, 95% CI 1.14, 4.12; β = 1.59, 95% CI 0.65, 2.53; β = 1.95, 95% CI 0.31, 3.6). The attitude scores were lower in those with a monthly income of K50,000-K100,000 and K100,000–500,000 than those with less than K50,000 (β = −1.56, 95% CI −2.59, −0.53; β = −2.15, 95% CI −3.53, −0.78).

Table 2. The comparison of SARS-CoV-2 vaccine hesitancy rate in different subgroups among PLHIV.

Subgroups, n (%)	No Vaccine hesitancy	Vaccine hesitancy	*p* value
Sex			.059
Female	79 (39.7)	120 (60.3)	
Male	71 (50.0)	71 (50.0)	
Age groups (years)			.282
18–29	31 (40.3)	46 (59.7)	
30–39	69 (41.6)	97 (58.4)	
40–49	40 (48.8)	42 (51.2)	
Above 50	10 (62.5)	6 (37.5)	
Education			<.001
Never been to School	7 (20.6)	27 (79.4)	
Primary	65 (38.9)	102 (61.1)	
Secondary	70 (53.4)	61 (46.6)	
Tertiary	8 (88.9)	1 (11.1)	
Occupation			.006
Farmers	64 (41.0)	92 (59.0)	
Small mediums	8 (53.3)	7 (46.7)	
Students	22 (56.4)	17 (43.6)	
Professional/Skilled workers	15 (26.8)	41 (73.2)	
Unemployed	14 (70.0)	6 (30.0)	
Other	27 (49.1)	28 (50.9)	
Monthly income			.114
Less than K50,000	95 (46.3)	110 (53.7)	
K50,000-K100,000	25 (36.8)	43 (63.2)	
K100,000–500,000	25 (40.3)	37 (59.7)	
Above K500,000	5 (83.3)	1 (16.7)	
Marital status			.491
Single	50 (46.7)	57 (53.3)	
Married	100 (42.7)	134 (57.3)	
Residence			.070
Rural	140 (42.9)	186 (57.1)	
Urban	10 (66.7)	5 (33.3)	
Religion			.038
Christian	140 (45.9)	165 (54.1)	
Muslim	10 (27.8)	26 (72.2)	

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; PLHIV, people living with human immunodeficiency virus.

Table 3. The multivariate logistic regression model of knowledge, attitude and trust scores with vaccination hesitancy in the PLHIV group.

Scale	OR	95% CI	*p*-value
Knowledge	0.80	0.67–0.97	.022
Attitude	0.62	0.53–0.72	<.001
Trust	0.93	0.82–1.06	.267

The reference of the logistic regression model was no vaccine hesitancy (Y = 0). Age, sex, education, occupation, monthly income and religion were adjusted. PLHIV, people living with human immunodeficiency virus; OR, odds ratio; CI, confidence interval.

Participants who resided in urban had lower attitude scores than that in rural (β = −1.99, 95% CI −3.63, −0.35). Muslims had a lower attitude score compared to Christians (β = −1.44, 95% CI −2.48, −0.40) (Table 5).

GLMs of trust scale and demographic characteristics

The mean of the trust scores was 12.86 (SD 2.69) among all the participants and the trust level was associated with their age and religion in the PLHIV group (Table 6). Compared to participants aged 18–29 years, those aged 30–39 years, 40–49 years and above 50 years had higher trust scores (β = 1.02, 95% CI 0.25, 1.78; β = 1.10, 95% CI 0.22, 1.98; β = 2.20, 95% CI 0.74, 3.66; all $p < .05$). Muslims had a lower trust score compared to Christians (β = −1.04, 95% CI −1.97, −0.11, $p = .029$) (Table 6).

Self-perceived risk of COVID-19 infection in PLHIV

The comparison of self-perceived risk to COVID-19 infection in PLHIV with vaccine hesitancy or not was presented in Table S4. The three perceptions between PLHIV with no vaccine hesitancy and PLHIV with vaccine hesitancy were significantly different ($p < .001$). In those with vaccine hesitancy, significantly fewer people thought that 'HIV infection increases the risk of SARS-CoV-2 infection,' 'HIV infection increases the risk of severe or life-threatening SARS-CoV-2 infection,' or 'SARS-CoV-2 infection is a serious infection warranting vaccination.'

Discussion

In this study, we found that there was no significant difference of SARS-CoV-2 vaccine hesitancy rate between PLHIV and non-PLHIV in Blantyre, Malawi. Over half of PLHIV were observed to have vaccine hesitancy. In PLHIV, SARS-CoV-2 vaccine hesitancy was associated with age, education, occupation, and religion. Those having higher knowledge and attitude scores were less likely to have vaccine hesitancy against SARS-CoV-2 in PLHIV.

Regarding vaccine hesitancy and vaccination status, there were no significant differences between the non-PLHIV and PLHIV groups. This was contrary to Iliyasu et al., who found vaccine acceptance was lower among PLHIV compared to the general population.[18] A lower willingness for SARS-CoV-2

Table 4. The generalized linear model of knowledge scores and demographic characteristics in the PLHIV group.

Characteristic	Mean (SD)	β (95% CI)	*P* value
Sex			
Female	10.88 (2.12)	Reference	
Male	11.01 (2.04)	−0.02 (−0.48, 0.45)	.943
Age groups (years)			
18–29	10.82 (2.19)	Reference	
30–39	10.84 (2.13)	0.14 (−0.46, 0.74)	.639
40–49	11.10 (2.00)	0.48 (−0.21, 1.17)	.173
Above 50	11.56 (1.46)	0.89 (−0.26, 2.04)	.131
Education			
Never been to school	10.74 (2.56)	Reference	
Primary	10.68 (2.09)	0.13 (−0.64, 0.91)	.734
Secondary	11.29 (1.93)	0.87 (0, 1.75)	.052
Tertiary	11.22 (1.56)	2.98 (0.28, 5.69)	.031
Occupation			
Farmers	10.86 (2.21)	Reference	
Small mediums	10.61 (2.25)	0.27 (−0.60, 1.14)	.546
Students	11.85 (1.39)	0.71 (−0.51, 1.93)	.257
Professional/Skilled workers	11.44 (1.94)	1.3 (0.26, 2.35)	.015
Unemployed	10.67 (1.95)	−0.24 (−0.89, 0.42)	.482
Other	11.33 (1.18)	0.49 (−0.67, 1.64)	.409
Monthly income			
Less than K50,000	11.02 (2.03)	Reference	
K50,000-K100,000	10.69 (2.25)	−0.83 (−1.55, −0.11)	.025
K100,000–500,000	10.94 (2.15)	−1.22 (−2.18, −0.26)	.013
Above K500,000	10.67 (1.37)	−3.85 (−7.04, −0.66)	.019
Marital status			
Single	10.90 (2.11)	Reference	
Married	10.95 (2.08)	0.13 (−0.39, 0.65)	.627
Residence			
Rural	10.98 (2.07)	Reference	
Urban	9.93 (2.31)	−1.83 (−2.97, −0.68)	.002
Religion			
Christian	10.97 (2.03)	Reference	
Muslim	10.58 (2.49)	−0.52 (−1.25, 0.21)	.165

PLHIV, people living with human immunodeficiency virus; SD, standard deviation; CI, confidence interval.

Table 5. The generalized linear model of attitude scores and demographic characteristics in the PLHIV group.

Characteristic	Mean (SD)	β (95% CI)	*P* value
Sex			
Female	12.53 (3.15)	Reference	
Male	12.75 (3.14)	−0.01 (−0.67, 0.66)	.983
Age groups (years)			
18–29	12.04 (3.02)	Reference	
30–39	12.60 (3.16)	0.95 (0.09, 1.81)	.031
40–49	12.95 (3.17)	1.60 (0.61, 2.59)	.002
Above 50	13.88 (3.12)	2.64 (1.00, 4.29)	.002
Education			
Never been to school	11.88 (3.41)	Reference	
Primary	12.20 (3.21)	0.82 (−0.29, 1.93)	.149
Secondary	13.27 (2.95)	1.83 (0.58, 3.09)	.004
Tertiary	13.67 (1.5)	3.36 (−0.50, 7.23)	.089
Occupation			
Farmers	12.02 (3.17)	Reference	
Small mediums	11.75 (3.07)	0.73 (−0.51, 1.98)	.251
Students	14.20 (2.76)	2.39 (0.64, 4.13)	.008
Professional/Skilled workers	13.59 (2.75)	2.63 (1.14, 4.12)	.001
Unemployed	13.53 (2.97)	1.59 (0.65, 2.53)	.001
Other	14.13 (2.8)	1.95 (0.31, 3.60)	.021
Monthly income			
Less than K50,000	12.84 (3.16)	Reference	
K50,000-K100,000	12.01 (3.10)	−1.56 (−2.59, −0.53)	.003
K100,000–500,000	12.44 (3.21)	−2.15 (−3.53, −0.78)	.002
Above K500,000	13.67 (1.63)	−3.21 (−7.77, 1.35)	.168
Marital status			
Single	12.46 (3.05)	Reference	
Married	12.69 (3.19)	0.39 (−0.35, 1.13)	.307
Residence			
Rural	12.64 (3.18)	Reference	
Urban	12.07 (2.05)	−1.99 (−3.63, −0.35)	.018
Religion			
Christian	12.76 (3.13)	Reference	
Muslim	11.44 (3.07)	−1.44 (−2.48, −0.40)	.007

PLHIV, people living with human immunodeficiency virus; SD, standard; CI, confidence interval.

Table 6. The generalized linear model of trust scores and demographic characteristics in the PLHIV group.

Characteristic	Mean (SD)	β (95% CI)	*P* value
Sex			
Female	11.79 (2.62)	Reference	
Male	11.96 (2.8)	−0.06 (−0.66, 0.53)	.833
Age groups (years)			
18–29	11.09 (2.54)	Reference	
30–39	11.95 (2.68)	1.02 (0.25, 1.78)	.010
40–49	12.04 (2.78)	1.10 (0.22, 1.98)	.015
Above 50	13.56 (2.13)	2.20 (0.74, 3.66)	.003
Education			
Never been to school	11.76 (2.31)	Reference	
Primary	11.77 (2.73)	0.25 (−0.73, 1.23)	.617
Secondary	11.93 (2.74)	0.41 (−0.70, 1.52)	.467
Tertiary	12.88 (3.09)	3.07 (−0.36, 6.51)	.081
Occupation			
Farmers	12.11 (2.60)	Reference	
Small mediums	11.02 (2.05)	−0.89 (−1.99, 0.22)	.117
Students	12.89 (2.59)	1.40 (−0.2, 3.00)	.088
Professional/Skilled workers	11.84 (2.95)	−0.24 (−1.58, 1.09)	.719
Unemployed	11.24 (3.13)	−0.67 (−1.50, 0.17)	.119
Other	13.40 (2.35)	1.12 (−0.34, 2.58)	.134
Monthly income			
Less than K50,000	12.05 (2.74)	Reference	
K50,000-K100,000	11.19 (2.58)	−0.89 (−1.81, 0.03)	.058
K100,000–500,000	11.95 (2.59)	−0.10 (−1.32, 1.12)	.873
Above K500,000	11.80 (3.19)	−2.85 (−6.99, 1.28)	.177
Marital status			
Single	11.56 (2.87)	Reference	
Married	11.99 (2.61)	0.52 (−0.14, 1.18)	.122
Residence			
Rural	11.84 (2.70)	Reference	
Urban	12.27 (2.55)	−0.31 (−1.78, 1.16)	.679
Religion			
Christian	11.99 (2.61)	Reference	
Muslim	10.78 (3.17)	−1.04 (−1.97, −0.11)	.029

PLHIV, people living with human immunodeficiency virus; SD, standard deviation; CI, confidence interval.

vaccination in PLHIV compared with the general population was also reported in another study.[14] This disparity might due to the difference of definitions or divisions of vaccine hesitancy and acceptance or willingness, as well as the disparate demographic characteristics. Moreover, as to the reason for the similar vaccine hesitancy rates between PLHIV and non-PLHIV of our study in Malawi, the bias might exist due to the sample selection. Nevertheless, it is definitely worth noting that over half of all the participants reported vaccine hesitancy, which is a significant barrier to achieving herd immunity and controlling the COVID-19 pandemic in Blantyre, Malawi.

We found that vaccination hesitancy rates against SARS-CoV-2 in PLHIV were significantly associated with age, education, occupation, and religion. The findings indicated that participants in older age groups (40–49 years and above 50 years) were less likely to have vaccine hesitancy compared to younger adults (18–29 years). This finding was consistent with previous studies which indicated that older PLHIV are more likely to accept vaccination.[19] This could be because older adults may perceive themselves to be at greater risk of severe illness and complications from COVID-19 and may be more motivated to get vaccinated.[20] PLHIV with a higher educational level might act better in understanding the knowledge of the vaccination importance of SARS-CoV-2 vaccine and thus they performed a lower vaccine hesitancy. This was verified by the higher knowledge scores in well-educated PLHIV. Regarding the occupation, the small mediums were more likely to have vaccine hesitancy against SARS-CoV-2 in our study, might due to certain consideration. This phenomenon still needed further investigation to clarify. Additionally, despite that Dubé et al. found no significant differences in vaccine acceptance between different religious groups.[21] Muslims appeared to have more intense vaccine hesitancy against SARS-CoV-2 in our study, which might attribute to their trust in the vaccine or the healthcare providers was weak, explained by the negative association with trust scores. It has been reported that the vaccine hesitancy rate was higher among certain religious groups with specific beliefs around health and healing, as documented by Iliyasu et al.,[18] and religion were shown to be significant predictors of trust in healthcare providers in another study.[22]

The results of this study also showed that the attitude toward vaccination was linked to vaccine hesitancy. Previous research showed positive associations between knowledge, attitudes, and vaccine acceptance. Larson *et al.* found that positive attitudes toward vaccines were associated with higher vaccination rates,[23] while Brewer *et al.* reported that vaccine attitudes were a significant predictor of vaccine uptake.[24] In contrast, one study showed that vaccine hesitancy was related to negative attitudes toward vaccines and distrust of the healthcare system.[25] In addition, this study revealed that PLHIV with vaccine hesitancy against SARS-CoV-2 had lower perceptions of their risk of SARS-CoV-2 infection, which was consistent with Brisson *et al.* that identified vaccine hesitancy as being associated with a lower perceived susceptibility and the severity of infectious diseases.[26] This finding

highlighted the need to address misconceptions and improve knowledge about the risks and benefits of vaccination, particularly among vulnerable populations.

There were several limitations of this study. The results of this study cannot be generalized or the potential selection bias and the sample size was limited. In addition, the status of HIV for the participants was self-reported which might have been biased especially for the non-PLHIV group. Third, health surveillance assistants were recruited to help conducting the data collection part. The health surveillance assistants provide health care at primary level in Malawi, that is, they do most of health awareness, provide vaccines and treat simple ailments in the community. Thus, they probably relate to the respondents but not a close relationship of one-to-one correspondence. Lastly, the study did not consider the influence of the vaccination environment which was an important determinant of vaccine uptake. However, the findings of this study might provide a more comprehensive understanding of the reasons for vaccine hesitancy against SARS-CoV-2 among PLHIV and help identify strategies to improve vaccine uptake.

Conclusion

SARS-CoV-2 vaccine hesitancy rate was high among PLHIV in Blantyre city, Malawi, which was a similar situation to non-PLHIV. Tailored efforts are needed to address these concerns and increase knowledge, trust, and positive attitude toward the vaccine to reduce vaccine hesitancy against SARS-CoV-2 in PLHIV.

Acknowledgments

We appreciate the assistance of Mpemba health center, Blantyre and all the participants in this study.

Author contributions

Chifundo Mchawa performed the investigation and collected the data. Chifundo Mchawa and Shan-Shan Zhang analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. Wan-Xue Zhang, Yiguo Zhou, Ting-Ting Wei and Juan Du prepared figures and/or tables and approved the final draft. Qing-Bin Lu and Fuqiang Cui conceived and designed the experiments, authored, or reviewed drafts of the paper, and approved the final draft.

Disclosure statement

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

Funding

This work was supported by the National Key Research and Development Program of China [2021YFC2301604] and Fundamental Research Funds for the Central Universities and Peking University Health Science Center [BMU2021YJ041].

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

1. Marzo RR, Sami W, Alam MZ, Acharya S, Jermsittiparsert K, Songwathana K, Pham NT, Respati T, Faller EM, Baldonado AM, et al. Hesitancy in COVID-19 vaccine uptake and its associated factors among the general adult population: a cross-sectional study in six Southeast Asian countries. Trop Med Health. 2022;50(1):4. doi:10.1186/s41182-021-00393-1.
2. Mzumara GW, Chawani M, Sakala M, Mwandira L, Phiri E, Milanzi E, Phiri MD, Kazanga I, O'Byrne T, Zulu EM, et al. The health policy response to COVID-19 in Malawi. BMJ Glob Health. 2021;6(5):e006035. doi:10.1136/bmjgh-2021-006035.
3. World Health Organization. WHO Coronavirus (COVID-19). [accessed 2023 May 22]. https://covid19.who.int.
4. World Health Organization. Coronavirus disease (COVID-19) and people living with HIV. [accessed 2023 Mar 13]. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-covid-19-and-people-living-with-hiv.
5. Okano JT, Sharp K, Valdano E, Palk L, Blower S. HIV transmission and source-sink dynamics in sub-Saharan Africa. Lancet HIV. 2020;7(3):e209–e14. doi:10.1016/S2352-3018(19)30407-2.
6. Burke RM, Henrion MYR, Mallewa J, Masamba L, Kalua T, Khundi M, Gupta-Wright A, Rylance J, Gordon SB, Masesa C, et al. Incidence of HIV-positive admission and inpatient mortality in Malawi (2012–2019). AIDS. 2021;35(13):2191–9. doi:10.1097/QAD.0000000000003006.
7. The Malawi Population-based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021). [accessed 2023 Mar 26]. https://phia.icap.columbia.edu/malawi-final-report-2020-2021/.
8. Ssentongo P, Heilbrunn ES, Ssentongo AE, Advani S, Chinchilli VM, Nunez JJ, Du P. Epidemiology and outcomes of COVID-19 in HIV-infected individuals: a systematic review and meta-analysis. Sci Rep. 2021;11(1):6283. doi:10.1038/s41598-021-85359-3.
9. Bhaskaran K, Rentsch CT, MacKenna B, Schultze A, Mehrkar A, Bates CJ, Eggo RM, Morton CE, Bacon SCJ, Inglesby P, et al. HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. Lancet HIV. 2021;8(1):e24–e32. doi:10.1016/S2352-3018(20)30305-2.
10. Kabir Sulaiman S, Sale Musa M, Isma'il Tsiga-Ahmed F, Muhammad Dayyab F, Kabir Sulaiman A, Dabo B, Idris Ahmad S, Abubakar Haruna S, Abdurrahman Zubair A, Hussein A, et al. COVID-19 vaccine hesitancy among people living with HIV in a low-resource setting: a multi-center study of prevalence, correlates and reasons. Vaccine. 2023;41(15):2476–84. doi:10.1016/j.vaccine.2023.02.056.
11. Razai MS, Chaudhry UAR, Doerholt K, Bauld L, Majeed A. Covid-19 vaccination hesitancy. BMJ. 2021;373:n1138. doi:10.1136/bmj.n1138.
12. Teixeira da Silva D, Biello K, Lin WY, Valente PK, Mayer KH, Hightow-Weidman L, Bauermeister JA. COVID-19 vaccine acceptance among an online sample of sexual and gender minority men and transgender women. Vaccines (Basel). 2021;9(3):204. doi:10.3390/vaccines9030204.
13. Vallee A, Fourn E, Majerholc C, Touche P, Zucman D. COVID-19 vaccine hesitancy among French people living with HIV. Vaccines (Basel). 2021;9(4):302. doi:10.3390/vaccines9040302.
14. Wu S, Ming F, Xing Z, Zhang Z, Zhu S, Guo W, Zou S, Liu J, Liu Y, Liang K. COVID-19 vaccination willingness among people living with HIV in Wuhan, China. Front Public Health. 2022;10:883453. doi:10.3389/fpubh.2022.883453.
15. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, Kimball S, El-Mohandes A. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med. 2021;27(2):225–8. doi:10.1038/s41591-020-1124-9.

16. AIDS MMoHDoHa, Malawi HIV estimates. [accessed 2023 Mar 26]. https://dms.hiv.health.gov.mw/dataset/malawi-2022-district-hiv-estimates-naomi-model.
17. Tianshuo Z, Hanyu L, Bingfeng H, Bei L, Jiang L, Juan D, Ninghua H, Qingbin L, Yaqiong L, Fuqiang C. Evaluation of the reliability and validity of a vaccine hesitancy scale on knowledge, attitude, trust and vaccination environment (KATE-S) in Chinese parents. Vaccine. 2022;40(21):2933–9. doi:10.1016/j.vaccine.2022.03.068.
18. Iliyasu Z, Kwaku AA, Umar AA, Tsiga-Ahmed F, Nass NS, Abdullahi HM, Amole TG, Salihu HM, Aliyu MH. Predictors of COVID-19 vaccine acceptability among patients living with HIV in Northern Nigeria: a mixed methods study. Curr HIV Res. 2022;20(1):82–90. doi:10.2174/1570162X19666211217093223.
19. Schwarzinger M, Watson V, Arwidson P, Alla F, Luchini S. COVID-19 vaccine hesitancy in a representative working-age population in France: a survey experiment based on vaccine characteristics. Lancet Public Health. 2021;6(4):e210–e21. doi:10.1016/S2468-2667(21)00012-8.
20. Govere-Hwenje S, Jarolimova J, Yan J, Khumalo A, Zondi G, Ngcobo M, Wara NJ, Zionts D, Bogart LM, Parker RA, et al. Willingness to accept COVID-19 vaccination among people living with HIV in a high HIV prevalence community. BMC Public Health. 2022;22(1):1239. doi:10.1186/s12889-022-13623-w.
21. Dube E, Gagnon D, Nickels E, Jeram S, Schuster M. Mapping vaccine hesitancy–country-specific characteristics of a global phenomenon. Vaccine. 2014;32(49):6649–54. doi:10.1016/j.vaccine.2014.09.039.
22. Guidry JPD, Miller CA, Perrin PB, Laestadius LI, Zurlo G, Savage MW, Stevens M, Fuemmeler BF, Burton CW, Gultzow T, et al. Between healthcare practitioners and clergy: evangelicals and COVID-19 vaccine hesitancy. Int J Environ Res Public Health. 2022;19(17):11120. doi:10.3390/ijerph191711120.
23. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine. 2014;32(19):2150–9. doi:10.1016/j.vaccine.2014.01.081.
24. Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A. Increasing vaccination: putting psychological science into action. Psychol Sci Public Interest. 2017;18(3):149–207. doi:10.1177/1529100618760521.
25. MacDonald NE, Smith J, and Appleton M. Risk perception, risk management and safety assessment: what can governments do to increase public confidence in their vaccine system? Biologicals. 2012;40(5):384–8. doi:10.1016/j.biologicals.2011.08.001.
26. Brisson M, Edmunds WJ. Economic evaluation of vaccination programs: the impact of herd-immunity. Med Decis Making. 2003;23(1):76–82. doi:10.1177/0272989X02239651.

JAMA Network Open

Original Investigation | Psychiatry

Esketamine vs Midazolam in Boosting the Efficacy of Oral Antidepressants for Major Depressive Disorder

A Pilot Randomized Clinical Trial

Chunfeng Xiao, MD; Jia Zhou, MSc; Anning Li, MD, PhD; Ling Zhang, MD, PhD; Xuequan Zhu, MSc; Jingjing Zhou, MD, PhD; Yongdong Hu, MD, PhD; Yunying Zheng, MD; Jing Liu, RN; Qiying Deng, RN; Haibo Wang, PhD; Gang Wang, MD, PhD

Abstract

IMPORTANCE Loss of a previously effective response while still using adequate antidepressant treatment occurs in a relatively high proportion of patients with major depressive disorder (MDD); therefore, there is a need to develop novel effective treatment strategies.

OBJECTIVE To assess the efficacy and safety of a single subanesthetic dose of esketamine in boosting the efficacy of oral antidepressants for treating fluctuating antidepressant response in MDD.

DESIGN, SETTING, AND PARTICIPANTS This single-center, double-blind, midazolam-controlled pilot randomized clinical trial was conducted at Beijing Anding Hospital, Capital Medical University in China. The study enrolled participants aged 18 years and older with fluctuating antidepressant response, defined as patients with MDD experiencing fluctuating symptoms after symptom relief and stabilization. Patient recruitment was conducted from August 2021 to January 2022, and participants were followed-up for 6 weeks. Data were analyzed as intention-to-treat from July to September 2022.

INTERVENTIONS All participants in the esketamine-treated group received intravenous esketamine at 0.2 mg/kg in 40 minutes. Participants in the midazolam control group received intravenous midazolam at 0.045 mg/kg in 40 minutes.

MAIN OUTCOMES AND MEASURES The primary outcome was the response rate at 2 weeks, defined as a 50% reduction in Montgomery-Åsberg Depression Rating Scale (MADRS). Secondary outcomes included response rate at 6 weeks, remission rates at 2 and 6 weeks, and change in MADRS and Clinical Global Impression–Severity score from baseline to 6 weeks; remission was defined by a MADRS score of 10 or lower.

RESULTS A total of 30 patients (median [IQR] age, 28.0 [24.0-40.0] years; 17 [56.7%] female) were randomized, including 15 patients randomized to midazolam and 15 patients randomized to esketamine; 29 patients completed the study. Response rates at 2 weeks were significantly higher in the esketamine-treated group than in the midazolam control group (10 patients [66.7%] vs 1 patient [6.7%]; $P < .001$). Participants treated with esketamine experienced significantly greater reduction in MADRS score from baseline to 2 weeks compared with those treated with midazolam (mean [SD] reduction, 15.7 [1.5] vs 3.1 [1.3]; $P < .001$). No serious adverse events were observed in this trial, and no psychotogenic effects and clinically significant manic symptoms were reported.

(continued)

Key Points

Question Can esketamine boost the efficacy of oral antidepressants in patients with major depressive disorder with fluctuating symptoms (ie, reemergence of initial symptoms or episode or onset of a new episode) occurring during adequate antidepressant treatment?

Findings In this randomized clinical trial of 30 patients with fluctuating antidepressant response, more patients who received a single subanesthetic dose of esketamine experienced a reduction of depression severity by at least 50% compared with patients who received midazolam, The difference remained statistically significant at the final follow-up at 6 weeks.

Meaning These findings suggest that esketamine could be used as a potentially effective and safe treatment option for fluctuating antidepressant response.

+ Visual Abstract

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

CONCLUSIONS AND RELEVANCE This pilot randomized clinical trial found that a single subanesthetic dose of esketamine could boost the efficacy of oral antidepressants in treating fluctuating antidepressant response, with a good safety profile.

TRIAL REGISTRATION Chinese Clinical Trial Registry Identifier: ChiCTR2100050335

JAMA Network Open. 2023;6(8):e2328817. doi:10.1001/jamanetworkopen.2023.28817

Introduction

Depression is a common chronic and disabling condition that affects an estimated 300 million people worldwide and is associated with personal, societal, and economic burdens.[1-3] Depression has high rates of recurrence and disability, with approximately half of patients experiencing a relapse or a recurrence.[4-6]

Pharmacotherapy is currently the leading strategy in the treatment of major depressive disorder (MDD).[7,8] Treatment guidelines recommend that patients with MDD continue antidepressant therapy for 4 to 9 months after successful acute phase treatment to prevent relapse and recurrence and 2 years or more of maintenance treatment at a full therapeutic dose for patients with an increased risk of MDD recurrence.[9,10] However, rates of relapse and recurrence remain high even among adherent patients with MDD.[11-13]

Unlike relapse and recurrence, which occur during the continuation or maintenance phase, loss of a previously effective response while still receiving adequate antidepressant treatment (ADT) occurs in a relatively high proportion of patients with MDD, with rates ranging from 9% to 57%.[14] This concerning phenomenon is known as *ADT tachyphylaxis* or *antidepressant tolerance*.[15] Tachyphylaxis is a pharmacological term defined as the rapid appearance of a progressive decrease in response to a given dose after administering an active substance.[16] This metric is unconducive to our examination of this phenomenon from a clinical perspective in practice, especially when this phenomenon is not well understood. Instead, we used the term *fluctuating antidepressant response in MDD* (FAD) to refer to the occurrence of fluctuating symptoms, including re-emergence of initial symptoms or episodes or the onset of a new episode during adequate ADT in patients with MDD.

The causes of FAD remain unclear, and there is a lack of effective treatments.[14] Available options are limited to increasing the current antidepressant dose, drug holidays, decreasing current antidepressant dose, changing antidepressant drugs, augmenting treatment strategies, and combination treatment strategies. These strategies might be successful after several attempts, but they increase the risk of unanticipated adverse effects in the meantime.[15,17] FAD may contribute to treatment-resistant depression (TRD) development.[18,19] Therefore, finding novel effective treatment strategies is essential to prevent TRD and adverse outcomes in patients with MDD.

Ketamine, a racemic mixture of 2 enantiomers (esketamine and R-ketamine), is a noncompetitive N-methyl-D-aspartate (NMDA) receptor antagonist that has been used in clinical practice as a standard anesthetic for nearly half a century.[20] It has been demonstrated that 40-minute administration of 0.5 mg/kg ketamine or 0.2 mg/kg esketamine produced a more pronounced antidepressant effect than other subanesthetic intravenous doses.[21-24] Ketamine has an affinity for multiple receptors, including the NMDA receptor, opioid receptor, and alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor, so it can initiate cascading effects acting in concert.[22,25-27] Since ketamine can act on the same receptors as current antidepressants, combining ketamine with existing antidepressants may be more beneficial for patients with MDD.[26] A study by Hu et al[27] reported that a single subanesthetic dose of ketamine could shorten antidepressant onset. Moreover, a recent clinical study found that ketamine may interact with existing antidepressants. Another study by Popova et al[28] found that switching patients with TRD from an ineffective antidepressant to a flexibly dosed esketamine nasal spray plus a newly initiated antidepressant was

more effective than the newly initiated antidepressant plus placebo. Thus, ketamine might potentially boost the efficacy of oral antidepressants. Therefore, we conducted a double-blind, midazolam-controlled randomized clinical trial to explore the efficacy and safety of a single dose subanesthetic esketamine plus ongoing adequate ADT among patients with FAD.

Methods

Trial Design

This single-center, double-blind, midazolam-controlled randomized clinical trial was conducted at Beijing Anding Hospital, Capital Medical University, Beijing, China, and approved by a local independent ethics committee. All participants provided their written informed consent. This study followed the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline. The trial protocol and statistical analysis plan are provided in Supplement 1.

Patients

Key inclusion criteria were age 18 to 60 years and diagnosed MDD according to the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). Symptoms were measured with 2 questions on a visual analogue scale (VAS; range, 0-10; lower score indicates worse symptoms): question 1, "How did you feel about your state during the period when your condition was stable, under the treatment of your current medication?" and question 2, "How do you feel about your state during the recent period when you are not feeling well, under the treatment of your current medication?" Patients were eligible for inclusion if they were experiencing symptom fluctuation (VAS question 2 score, ≤5) after symptom relief and stabilization (VAS question 1 score, ≥8). For inclusion, the patient's most recent depressive episode had to have been stabilized by a single antidepressant without any adjustments in the type and dosage, and the duration of drug interruption during this period had to be less than 7 days. Finally, patients must have had a score of at least 14 on the 17-item Hamilton Rating Scale for Depression (range, 0-52, with higher score indicating more severe depression symptoms). The key exclusion criteria were (1) a history of TRD, schizophrenia, bipolar disorder, obsessive-compulsive disorder, drug or alcohol dependence, and ketamine or its enantiomers exposure, or current MDD episode with psychotic features; (2) contraindications to esketamine or midazolam; (3) clinically significant anomalies in laboratory test results at screening and any concern of certain conditions with trial risks. Detailed eligibility criteria are described in Supplement 1.

Randomization and Blinding

Eligible patients were randomly assigned in a 1:1 ratio to receive either intravenous esketamine or intravenous midazolam for 40 minutes using a computer-generated block randomization scheme (block size, 4). An independent assistant who was not involved in the study was responsible for packing and labeling the prescribed medication. Study investigators, staff, participants, and all outcome assessors were blinded to the treatment assignment.

Interventions

The infusion process was performed at Beijing Anding Hospital, Capital Medical University. All participants were admitted to the research ward the day before infusion and started 6 hours preoperative fasting and water deprivation. On the infusion day, 0.2 mg/kg esketamine or 0.045 mg/kg midazolam was administered intravenously according to randomization group. Both medications were dissolved in 50 mL of normal saline and administered for 40 minutes using a syringe pump. Safety checks were performed before, during, and after infusion. Detailed information on the safety check procedure is presented in the trial protocol in Supplement 1. Participants were permitted to discharge after completing the 24 hours follow-up visit after infusion. All participants in both groups maintained their original drug regimen throughout the study.

Efficacy Assessments

Efficacy was assessed by the Montgomery-Åsberg Depression Rating Scale (MADRS; range, 0-60; higher scores indicating more severe depressive symptoms) and Clinical Global Impression–Severity (CGI-S; range, 1-7; higher score indicates greater severity of illness) at 40 minutes; 2, 4, and 24 hours; and 1, 2, 4, and 6 weeks after infusion. Recalling time of MADRS general version is 1 week. In this regard, the structured interview guide for MADRS was used at 4 and 24 hours with its corresponding version (only these versions were available before this study), since such guides have previously been shown to increase the reliability of given scales and to ensure the accuracy of assessment a short time after infusion.[29]

Safety Assessments

General side effects were measured by the Udvalg for Kliniske Undersogelser side effect rating scale (range, 0-144; higher score indicates more severe side effects) and the Frequency, Intensity, and Burden of Side Effects Rating Scale (range, 0-18; higher score indicates more severe side effects). Psychotogenic effects were measured with the 4-item positive symptom subscale of Brief Psychiatric Rating Scale consisting of suspiciousness, hallucinations, unusual thought content, and conceptual disorganization (range, 4-28; higher score indicates more severe psychotic symptoms). Dissociative effects were measured with the Clinician-Administered Dissociative States Scale (range, 0-92; higher score indicates more severe dissociative symptoms), and manic symptoms were assessed using the Young Mania Rating Scale (range, 0-60; higher score indicates more severe manic symptoms). The Columbia-Suicide Severity Rating Scale (range, 0-42; higher score indicating greater suicidal risk) was also rated throughout the trial for safety purposes.

Outcomes

The primary outcome was the response rate at 2 weeks, with response defined as a 50% reduction in the MADRS. Secondary outcomes included the response rate at 6 weeks, remission rates at 2 and 6 weeks, and a change in MADRS and CGI-S score from baseline to 6 weeks, with remission defined by a MADRS score of 10 or lower. Safety outcomes included general side effects, psychotogenic effects, dissociative effects, manic symptoms, and suicide risk.

Sample Size Calculation

Since this study was a pilot study, no formal sample size calculation was performed. We planned to recruit 30 participants (15 participants in each group). Effect size data from this pilot study will be used to estimate the sample size of a future randomized clinical trial.

Statistical Analysis

Analyses were performed as intention-to-treat, including all randomized participants who completed the assessments at baseline and received study medication (full analysis set [FAS]). Missing data for MADRS scores were imputed using the last observation carried forward. Continuous variables are presented as mean (SD) or median (IQR). Categorical variables are presented as numbers and percentages. We used χ^2 tests, Fisher exact test, independent-sample *t* test, or Wilcoxon rank-sum test, as appropriate, to compare the differences in baseline variables between the intervention and control groups. Treatment effects (ie, response and remission rates) were evaluated using the χ^2 test. Sensitivity analysis was also performed for the primary outcome using the logistic regression model with baseline MADRS score and the number of MDD episodes as covariates. Secondary outcomes, the reduction in MADRS score and CGI-S from baseline over time, were analyzed using independent-sample *t* tests or Wilcoxon rank-sum test. Sensitivity analysis was also performed using an analysis of covariance model with baseline MADRS score and the number of MDD episodes as covariates. Moreover, to compare the difference between groups over time, a 2-way repeated measures analysis of variance was performed accounting for group, time, and the interaction term between group and time. All tests were 2-sided, and $P < .05$ was considered statistically significant. All the statistical

analyses were performed with SAS statistical software, version 9.4 (SAS Institute). Data were analyzed from July to September 2022.

Results

Participants

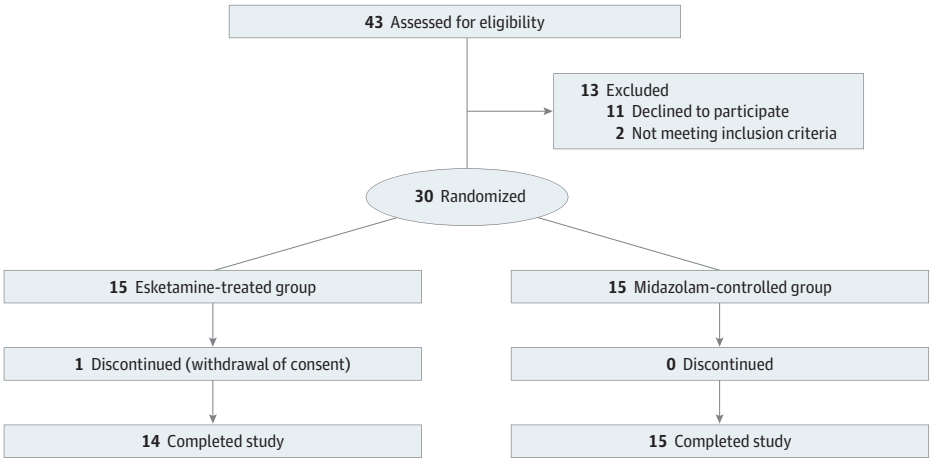
From August 2021 to January 2022, 43 participants were evaluated for eligibility, and 30 participants (median [IQR] age, 28.0 [24.0-40.0] years; 17 [56.7%] female) with FAD were enrolled (**Figure 1**). All the enrolled participants were included in the FAS for final analysis, and 29 participants (96.7%) completed the trial. There was 1 participant in the esketamine-treated group who dropped out at the end of 1 week. They received 0.2 mg/kg esketamine infusion and left the city due to a job change.

The 2 groups were generally balanced in terms of baseline demographic and clinical characteristics, except for the proportion of participants with MDD of the first episode, mean number of MDD episodes, and mean MADRS score (**Table 1**). The median (IQR) duration of ongoing ADT with their current antidepressant was 24.9 (14.9-51.0) weeks, 8.6 (4.0-32.9) weeks stable after symptom relief, and 5.8 (3.1-10.3) weeks with fluctuating symptoms. Among all participants, 15 (50.0%) were experiencing their first episode of MDD, 5 (16.7%) had a family history of psychiatric illness, and 19 (63.3%) were receiving selective serotonin receptor inhibitors, followed by 9 (30.0%) receiving serotonin and norepinephrine reuptake inhibitors, and 2 (6.7%) receiving other types of antidepressants.

Primary Outcome

At 2 weeks, the response rate was significantly higher in the group treated with esketamine than in participants in the midazolam control group (10 participants [66.7%] vs 1 participants [6.7%]; OR, 28.00; 95% CI, 2.82 to 277.96). Significant between-group differences were also observed at other visits, including 1, 4, and 6 weeks (**Figure 2**). However, there was no significant difference in response rate between the intervention and control groups at 24 hours (10 participants [66.7%] vs 5 participants [33.3%]; $P = .07$). In sensitivity analysis using logistic regression, the difference in response rate between the esketamine-treated and midazolam control groups at 2 weeks remained statistically significant after adjusting for the baseline MADRS score and the number of MDD episodes (OR, 14.17; 95% CI, 1.14 to 175.96; $P = .04$).

Figure 1. Participant Recruitment Flowchart



One patient in the esketamine-treated group dropped out at the end of 1 week. They received 0.2 mg/kg esketamine infusion and left the city due to job changes.

Table 1. Participant Demographic and Clinical Characteristics at Baseline

Characteristic	Participants, No. (%)	
	Midazolam-controlled group (n = 15)	Esketamine-treated group (n = 15)
Sex		
Female	8 (53.3)	9 (60.0)
Male	7 (46.7)	6 (40.0)
Ethnicity		
Han	15 (100)	13 (86.7)
Other[a]	0	2 (13.3)
Education level, y		
>12	12 (80.0)	13 (86.7)
≤12	3 (20.0)	2 (13.3)
Marital status		
Married	5 (33.3)	6 (40.0)
Unmarried	10 (66.7)	9 (60.0)
Alcohol history		
No	11 (73.3)	11 (73.3)
Yes	4 (26.7)	4 (26.7)
Smoking history		
No	10 (66.7)	11 (73.3)
Yes	5 (33.3)	4 (26.7)
Family history of psychiatric illness		
No	12 (80.0)	13 (86.7)
Yes	3 (20.0)	2 (13.3)
First-episode MDD		
No	11 (73.3)	4 (26.7)
Yes	4 (26.7)	11 (73.3)
Antidepressant use		
Bupropion	1 (6.7)	0 (0.0)
Escitalopram oxalate	4 (26.7)	4 (26.7)
Duloxetine	1 (6.7)	2 (13.3)
Fluvoxamine	1 (6.7)	2 (13.3)
Fluoxetine	1 (6.7)	1 (6.7)
Mirtazapine	0	1 (6.7)
Sertraline	3 (20.0)	3 (20.0)
Venlafaxine	4 (26.7)	2 (13.3)
Class of antidepressant		
SSRIs	9 (60.0)	10 (66.7)
SNRIs	5 (33.3)	4 (26.7)
Other[b]	1 (6.7)	1 (6.7)
Age, median (IQR), y	26.0 (24.0-37.0)	30.0 (25.0-40.0)
BMI, mean (SE)	24.1 (5.9)	23.7 (3.7)
Age at onset of MDD, mean (SE), y	23.9 (8.1)	28.5 (11.7)
Total course of MDD, median (IQR), y	5.0 (1.8-12.6)	1.9 (0.6-6.7)
MDD episodes, mean (SE), No.	3.0 (1.0-6.0)	1.0 (1.0-2.0)
Duration of recent episode, median (IQR), wk	48.0 (29.0-67.0)	61.0 (26.0-103.0)
Duration of ADT for recent episode, median (IQR), wk	30.3 (11.7-52.7)	59.6 (20.9-99.0)
Duration of symptoms with current antidepressant medication, median (IQR), wk		
Ongoing ADT	26.4 (11.7-45.7)	23.1 (14.9-85.4)
Maintaining stability after symptom relief	11.0 (3.3-31.6)	8.0 (4.0-36.0)
Fluctuating symptoms	6.6 (3.1-14.1)	4.7 (2.3-10.0)
Screening HAMD-17 score, mean (SE)[c]	18.9 (4.2)	20.1 (2.6)

(continued)

Table 1. Participant Demographic and Clinical Characteristics at Baseline (continued)

Characteristic	Participants, No. (%)	
	Midazolam-controlled group (n = 15)	Esketamine-treated group (n = 15)
Baseline MADRS score, mean (SE)[d]	21.3 (5.0)	26.1 (6.4)
Screening VAS-1 score, median (IQR)[e]	8.0 (8.0-9.0)	8.0 (8.0-8.0)
Screening VAS-2 score, median (IQR)[e]	2.8 (1.2)	2.7 (1.1)

Abbreviations: ADT, antidepressant treatment; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); HAMD-17, 17-item Hamilton Rating Scale for Depression; MADRS, the Montgomery-Åsberg Depression Rating Scale; MDD, major depressive disorder; SNRIs, serotonin and norepinephrine reuptake inhibitors; SSRIs, selective serotonin reuptake inhibitors; VAS, Visual Analogue Scale score.

[a] Two patients were included in other ethnicities: 1 identified as Hui ethnicity, and the other identified as Man ethnicity.

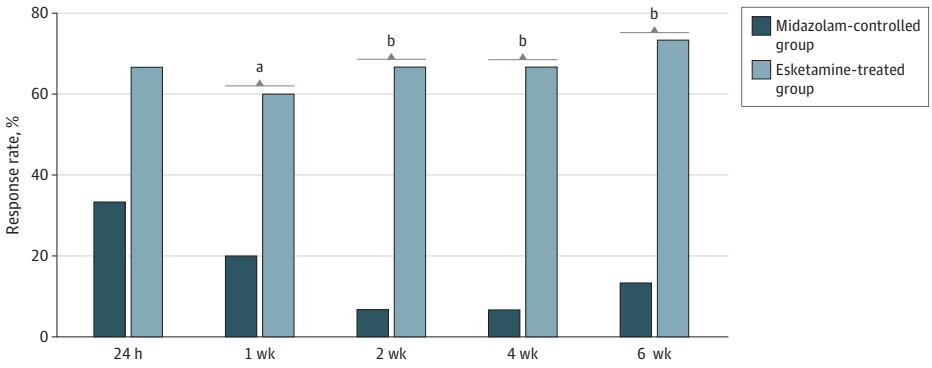
[b] One patient was receiving a norepinephrine–dopamine reuptake inhibitor (bupropion), and the other was receiving a noradrenergic and specific serotonergic antidepressant (mirtazapine).

[c] Range, 0 to 52; higher total score indicates more severe depression.

[d] Range, 0 to 60; higher score indicates more severe depressive symptoms.

[e] Range, 0 to 10. A score of 0 indicates "Seeing my normal state as a reference, my state is horrible during this period"; 10, "Seeing my normal state as a reference, my state is quite good (close to my normal state)" during this period. VAS-1 was "How did you feel about your state during the period when your condition was stable, under the treatment of your current medication?"; VAS-2 was "How do you feel about your state during the recent period when you are not feeling well, under the treatment of your current medication?"

Figure 2. The Response Rates at Different Visits by Intervention and Control Group



Response was defined as a 50% reduction in the Montgomery-Åsberg Depression Rating Scale (MADRS; range, 0-60; higher scores indicating more severe depressive symptoms). Missing data for 1 patient's MADRS score were imputed using the last observation carried forward.

[a] $P < .05$.

[b] $P < .001$.

Secondary Outcomes

Participants in the esketamine-treated group had higher remission rates than the midazolam control group at 4 hours (10 participants [66.7%] vs 3 participants [20.0%]; $P = .01$) and 4 weeks (8 participants [53.3%] vs 1 participants [6.7%]; $P = .005$). There was no statistically significant difference between groups at other visits (eTable 1 in Supplement 2).

Mean reduction in MADRS score from baseline to 2 weeks among participants treated with esketamine was significantly greater than among those treated with midazolam (mean [SD] reduction, 15.7 [1.5] vs 3.1 [1.3]; $P < .001$). This between-group difference was also observed at all other visits (**Figure 3**). Additionally, results of analysis of covariance indicated that differences between groups in the reduction in MADRS score from baseline to 6 weeks remained statistically significant after adjusting for the baseline MADRS score and the number of MDD episodes, except for at 24 hours. After adjustment for the covariates, the differences in the least squares mean reduction between the groups were −8.74 (95% CI, −13.11 to- −4.37; $P < .001$) at 2 weeks and −5.55 (95% CI, −10.08 to −1.03; $P = .02$) at 6 weeks (eFigure in Supplement 2). The results of 2-way repeated-measures analysis of variance indicated that the difference between groups over time was statistically significant, referring to interaction between factors (time × group) ($F = 8.27$; $P < .001$).

The reduction in CGI-S score from baseline to 6 weeks was significantly greater in the esketamine-treated group than in the midazolam control group at each visit (eTable 2 in Supplement 2).

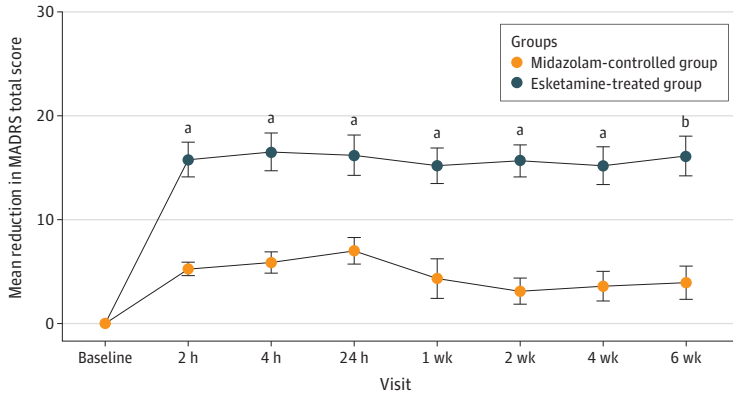
Safety

No serious adverse events were observed in this trial, no psychotogenic effects were reported on the Brief Psychiatric Rating Scale, and no clinically significant manic symptoms were reported on the Young Mania Rating Scale. All participants scored 0 points in Clinician-Administered Dissociative States Scale at 2, 4, and 24 hours. At 40 minutes, participants in the esketamine-treated group showed higher level of dissociation compared with the midazolam control group (median [IQR] score, 0 [0-0] vs 6 [3-13]; $P < .001$). For general side effects assessed by Udvalg for Kliniske Undersogelser, which were defined as an increase of at least 1 on each item, 2 participants (13.3%) had increased dream activity in the group treated with esketamine, and 4 participants (26.7%) had tension or inner unrest and 3 participants (20.0%) had increased dream activity in the midazolam control group (**Table 2**).

Discussion

This double-blind, pilot randomized clinical trial assessed the efficacy and safety of a subanesthetic dose of esketamine for treating FAD and found that esketamine was more effective than midazolam, with a good safety profile. Efficacy duration of a single subanesthetic dose of esketamine for treating MDD, including TRD, is approximately 10 to 14 days, with limited studies suggesting longer-lasting effects.[27] The most recent evidence indicates that the efficacy of esketamine as an add-on therapy in combination with oral antidepressants does not extend beyond 4 weeks. In a double-blind, active-controlled randomized clinical trial by Chen et al,[30] the group receiving esketamine exhibited a clinically meaningful treatment effect at 24 hours compared with the placebo group, but the effect had diminished by day 28.[30] Likewise, a phase 2b study in Japan involving participants with TRD indicated no significant difference between the esketamine and placebo groups at day 28.[31] In this study, the mean reduction in MADRS score from baseline to 4 weeks was significantly higher in participants treated with esketamine, which contradicts the findings of previous studies showing a loss of efficacy within 4 weeks. This indicates that a single subanesthetic dose of esketamine has a positive effect on oral antidepressants in patients with FAD, since the expected spillover efficacy was aided by oral antidepressants. Furthermore, this effect could last for an extended period, as the mean reduction in MADRS score from 4 to 6 weeks in the esketamine-treated group remained significantly higher than in the control group. This interesting phenomenon suggests that a subanesthetic dose of esketamine could boost the efficacy of antidepressants, therefore providing preliminary evidence

Figure 3. The Mean Reduction of Montgomery-Åsberg Depression Rating Scale (MADRS) Scores From Baseline to 6 Weeks at Each Visit by Intervention and Control Group



MADRS (the Montgomery-Åsberg Depression Rating Scale; range: 0-60; higher scores indicating more severe depressive symptoms). Error bars indicate SEs. Missing data of MADRS score for 1 patient were imputed using the last observation carried forward.

[a] $P < .001$.

[b] $P < .01$.

to confirm our hypothesis that esketamine therapy is a promising strategy for the treatment of FAD. The simplicity and high success rate of esketamine therapy would allow patients to continue using their current antidepressant regimen and avoid adverse effects associated with changes in the medication regimen, for example, unanticipated adverse effects and a reduced willingness to seek treatment due to repeated attempts.

The mechanism of ketamine and its enantiomers in exerting antidepressant effects is unclear, and few studies have reported the underlying interaction between ketamine and antidepressants. A study by Dean et al[32] summarized the differences in efficacy between ketamine and other glutamate receptor modulators in the treatment of MDD, reporting that ketamine had higher response efficacy compared with midazolam at 24 hours, with very low–certainty evidence at 1 and 2 weeks, and no significant difference was observed at 4 weeks.[32] In our study, the rates of response at 1, 2, 4, and 6 weeks were significantly higher in the esketamine-treated group compared with the midazolam control group. A study by Hu et al[33] found that a single subanesthetic dose of ketamine plus newly initiated escitalopram was associated with significantly lower MADRS scores from 2 hours to 2 weeks. However, the mean reduction in MADRS score from baseline at each visit was significantly higher among participants treated with esketamine than among those treated with midazolam. The possible reasons for these inconsistencies could be the different types of participants and the unknown interaction of esketamine and antidepressants.[33] Several reports have suggested that FAD may result from an evolving drug sensitization that causes a pharmacokinetic or pharmacodynamic tolerance, or both, to the previously effective concentration or drug actions.[34,35] For patients with FAD receiving the same drug with a loss of efficacy, the underlying mechanism of ketamine's effect on antidepressants might involve some unidentified interaction with antidepressants to alter antidepressant sensitization and improve tolerance to antidepressants. Although there is no direct evidence to suggest that ketamine and antidepressants interact, ketamine may act on the same receptors, with the most evidence for serotonin (eg, 5-hydroxytryptophan).[26]

Additionally, ketamine can alleviate depressive symptoms by restoring the lost prefrontal synaptic spines, which may partially explain the sustained antidepressant effects observed in our study.[36] By modifying the structure of individual nervous system, ketamine may allow patients to restore antidepressant sensitivity. Based on our findings, it is hypothesized that ketamine or its enantiomer also exerts antidepressant effects via a reticular receptor-associated system involving

Table 2. Reported Frequency of Side Effects and Adverse Events Assessed With Udvalg for Kliniske Undersogelser by Intervention and Control Groups

Side effect	Participants, No.	
	Midazolam control group	Esketamine-treated group
Psychiatric		
Asthenia or lassitude	0	1
Sleepiness or sedation	0	1
Tension or inner unrest	4	0
Reduced duration of sleep	0	1
Increased dream activity	3	2
Emotional indifference	1	1
Neurologic: tremors	2	0
Autonomic		
Polyuria or polydipsia	1	0
Palpitations or tachycardia	1	0
Increased sweating	1	0
Other		
Weight loss	0	1
Reduced sexual desire	1	0

several antidepressant receptors, which may include interaction with multiple antidepressant receptors.

Limitations

This study has several limitations. First, symptom fluctuation was evaluated retrospectively using the VAS tool, and the definition of FAD was empirical without a reference standard, which may have introduced bias and affected data accuracy. However, proposing a definition of FAD was fundamental to the needs of clinical practice. An improved definition of FAD in the future requires adequate support of higher-quality evidence. Additionally, another potential weakness of the study was its small sample size in a single center, which may limit the generalizability of the results. What is more, including patients with first-episode depression and patients with recurrent depression in the sample may limit the internal and external validity. Furthermore, collecting data during the COVID-19 pandemic may have impacted the results of this study. Additionally, a significant difference in dissociation levels at 40 minutes was observed and blinding assessment was lacking, which could potentially affect the integrity of blinding.

Conclusion

This double-blind randomized clinical trial found that a single subanesthetic dose of esketamine could boost the efficacy of oral antidepressants in the treatment of FAD with a good safety profile. Future larger trials with longer follow-up are required to confirm these findings.

ARTICLE INFORMATION

Accepted for Publication: June 30, 2023.

Published: August 14, 2023. doi:10.1001/jamanetworkopen.2023.28817

Corresponding Authors: Gang Wang, MD, PhD, Beijing Anding Hospital, Capital Medical University, 5 Ankang Hutong, Beijing 100088, China (gangwangdoc@ccmu.edu.cn); Haibo Wang, PhD, Peking University Clinical Research Institute, Peking University Frist Hospital, 38 Xueyuan St, Beijing 100191, China (hbwang2005@163.com).

Author Affiliations: Beijing Key Laboratory of Mental Disorders, National Clinical Research Center for Mental Disorders & National Center for Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China (Xiao, Jia Zhou, Li, Zhang, Zhu, Jingjing Zhou, Zheng, Liu, Deng, G. Wang); Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China (Xiao, Jia Zhou, Li, Zhang, Zhu, Jingjing Zhou, Zheng, Liu, Deng, G. Wang); Unit of Psychological Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China (Hu); Peking University Clinical Research Institute, Peking University First Hospital, Beijing, China (H. Wang); Key Laboratory of Epidemiology of Major Diseases, Peking University, Ministry of Education (H. Wang).

Author Contributions: Dr G. Wang had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Xiao, Jia Zhou, and Dr. Li are co–first authors.

Concept and design: Xiao, Zhang, Zhu, Zheng, H. Wang, G. Wang.

Acquisition, analysis, or interpretation of data: Xiao, Jia Zhou, Li, Zhang, Zhu, Jingjing Zhou, Hu, Liu, Deng, H. Wang, G. Wang.

Drafting of the manuscript: Xiao, Jia Zhou, Zheng, H. Wang.

Critical review of the manuscript for important intellectual content: Li, Zhang, Zhu, Jingjing Zhou, Hu, Liu, Deng, H. Wang, G. Wang.

Statistical analysis: Jia Zhou, H. Wang.

Obtained funding: G. Wang.

Administrative, technical, or material support: Li, Zhang, Zhu, Jingjing Zhou, Hu, Liu, Deng, H. Wang, G. Wang.

Supervision: Li, Zhang, Deng, G. Wang.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by research grants from the Sci-Tech Innovation 2030 Major Project of Brain Science and Brain-inspired Intelligence Technology (grant No. 2021ZD0200600) and Beijing Scholar 2021 (grant No. 063).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 3.

Additional Contributions: We thank the patients and their families for participating in this trial.

REFERENCES

1. Cheng X, Wang Q, Wang R, et al. Prevalence of depressive disorders and associated demographic characteristics in Shandong: an epidemiological investigation. *J Affect Disord*. 2022;311:198-204. doi:10.1016/j.jad.2022.05.084

2. Thapar A, Eyre O, Patel V, Brent D. Depression in young people. *Lancet*. 2022;400(10352):617-631. doi:10.1016/S0140-6736(22)01012-1

3. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018;75(4):336-346. doi:10.1001/jamapsychiatry.2017.4602

4. Moriarty AS, Meader N, Snell KI, et al. Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. *Cochrane Database Syst Rev*. 2021;5(5):CD013491. doi:10.1002/14651858.CD013491.pub2

5. Beshai S, Dobson KS, Bockting CL, Quigley L. Relapse and recurrence prevention in depression: current research and future prospects. *Clin Psychol Rev*. 2011;31(8):1349-1360. doi:10.1016/j.cpr.2011.09.003

6. Burcusa SL, Iacono WG. Risk for recurrence in depression. *Clin Psychol Rev*. 2007;27(8):959-985. doi:10.1016/j.cpr.2007.02.005

7. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7

8. LeBlanc A, Herrin J, Williams MD, et al. Shared decision making for antidepressants in primary care: a cluster randomized trial. *JAMA Intern Med*. 2015;175(11):1761-1770. doi:10.1001/jamainternmed.2015.5214

9. Deshauer D, Moher D, Fergusson D, Moher E, Sampson M, Grimshaw J. Selective serotonin reuptake inhibitors for unipolar depression: a systematic review of classic long-term randomized controlled trials. *CMAJ*. 2008;178(10):1293-1301. doi:10.1503/cmaj.071068

10. Davidson JR. Major depressive disorder treatment guidelines in America and Europe. *J Clin Psychiatry*. 2010;71(suppl E1):e04. doi:10.4088/JCP.9058se1c.04gry

11. Lewis G, Marston L, Duffy L, et al. Maintenance or discontinuation of antidepressants in primary care. *N Engl J Med*. 2021;385(14):1257-1267. doi:10.1056/NEJMoa2106356

12. Duffy L, Clarke CS, Lewis G, et al. Antidepressant medication to prevent depression relapse in primary care: the ANTLER RCT. *Health Technol Assess*. 2021;25(69):1-62. doi:10.3310/hta25690

13. Akerblad AC, Bengtsson F, von Knorring L, Ekselius L. Response, remission and relapse in relation to adherence in primary care treatment of depression: a 2-year outcome study. *Int Clin Psychopharmacol*. 2006;21(2):117-124. doi:10.1097/01.yic.0000199452.16682.b8

14. Kinrys G, Gold AK, Pisano VD, et al. Tachyphylaxis in major depressive disorder: A review of the current state of research. *J Affect Disord*. 2019;245:488-497. doi:10.1016/j.jad.2018.10.357

15. Targum SD. Identification and treatment of antidepressant tachyphylaxis. *Innov Clin Neurosci*. 2014;11(3-4):24-28.

16. Nichols JJ. Optometry and vision science. In: Stedmans; *Stedman's Medical Dictionary*. 27th ed. Lippincott Williams & Wilkins; 2000.

17. Kudlow PA, McIntyre RS, Lam RW. Early switching strategies in antidepressant non-responders: current evidence and future research directions. *CNS Drugs*. 2014;28(7):601-609. doi:10.1007/s40263-014-0171-5

18. Zimmerman M, Thongy T. How often do SSRIs and other new-generation antidepressants lose their effect during continuation treatment: evidence suggesting the rate of true tachyphylaxis during continuation treatment is low. *J Clin Psychiatry*. 2007;68(8):1271-1276. doi:10.4088/JCP.v68n0814

19. Amsterdam JD, Shults J. Does tachyphylaxis occur after repeated antidepressant exposure in patients with bipolar II major depressive episode? *J Affect Disord*. 2009;115(1-2):234-240. doi:10.1016/j.jad.2008.07.007

20. Andrade C. Ketamine for depression, 3: does chirality matter? *J Clin Psychiatry*. 2017;78(6):e674-e677. doi:10.4088/JCP.17f11681

21. Peyrovian B, McIntyre RS, Phan L, et al. Registered clinical trials investigating ketamine for psychiatric disorders. *J Psychiatr Res*. 2020;127:1-12. doi:10.1016/j.jpsychires.2020.03.020

22. Su TP, Chen MH, Li CT, et al. Dose-related effects of adjunctive ketamine in Taiwanese patients with treatment-resistant depression. *Neuropsychopharmacology*. 2017;42(13):2482-2492. doi:10.1038/npp.2017.94

23. Fava M, Freeman MP, Flynn M, et al. Double-blind, placebo-controlled, dose-ranging trial of intravenous ketamine as adjunctive therapy in treatment-resistant depression (TRD). *Mol Psychiatry*. 2020;25(7):1592-1603. doi:10.1038/s41380-018-0256-5

24. Singh JB, Fedgchin M, Daly E, et al. Intravenous esketamine in adult treatment-resistant depression: a double-blind, double-randomization, placebo-controlled study. *Biol Psychiatry*. 2016;80(6):424-431. doi:10.1016/j.biopsych.2015.10.018

25. Williams NR, Schatzberg AF. NMDA antagonist treatment of depression. *Curr Opin Neurobiol*. 2016;36:112-117. doi:10.1016/j.conb.2015.11.001

26. Zanos P, Moaddel R, Morris PJ, et al. Ketamine and ketamine metabolite pharmacology: insights into therapeutic mechanisms. *Pharmacol Rev*. 2018;70(3):621-660. doi:10.1124/pr.117.015198

27. Han Y, Chen J, Zou D, et al. Efficacy of ketamine in the rapid treatment of major depressive disorder: a meta-analysis of randomized, double-blind, placebo-controlled studies. *Neuropsychiatr Dis Treat*. 2016;12:2859-2867. doi:10.2147/NDT.S117146

28. Popova V, Daly EJ, Trivedi M, et al. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. *Am J Psychiatry*. 2019;176(6):428-438. doi:10.1176/appi.ajp.2019.19020172

29. Williams JB, Kobak KA. Development and reliability of a structured interview guide for the Montgomery Asberg Depression Rating Scale (SIGMA). *Br J Psychiatry*. 2008;192(1):52-58. doi:10.1192/bjp.bp.106.032532

30. Chen X, Hou X, Bai D, et al. Efficacy and safety of flexibly dosed esketamine nasal spray plus a newly initiated oral antidepressant in adult patients with treatment-resistant depression: a randomized, double-blind, multicenter, active-controlled study conducted in China and USA. *Neuropsychiatr Dis Treat*. 2023;19:693-707. doi:10.2147/NDT.S391096

31. Takahashi N, Yamada A, Shiraishi A, Shimizu H, Goto R, Tominaga Y. Efficacy and safety of fixed doses of intranasal esketamine as an add-on therapy to oral antidepressants in Japanese patients with treatment-resistant depression: a phase 2b randomized clinical study. *BMC Psychiatry*. 2021;21(1):526. doi:10.1186/s12888-021-03538-y

32. Dean RL, Hurducas C, Hawton K, et al. Ketamine and other glutamate receptor modulators for depression in adults with unipolar major depressive disorder. *Cochrane Database Syst Rev*. 2021;9(9):CD011612. doi:10.1002/14651858.CD011612.pub3

33. Hu YD, Xiang YT, Fang JX, et al. Single I.V. ketamine augmentation of newly initiated escitalopram for major depression: results from a randomized, placebo-controlled 4-week study. *Psychol Med*. 2016;46(3):623-635. doi:10.1017/S0033291715002159

34. Byrne SE, Rothschild AJ. Loss of antidepressant efficacy during maintenance therapy: possible mechanisms and treatments. *J Clin Psychiatry*. 1998;59(6):279-288. doi:10.4088/JCP.v59n0602

35. Fava GA. Do antidepressant and antianxiety drugs increase chronicity in affective disorders? *Psychother Psychosom*. 1994;61(3-4):125-131. doi:10.1159/000288880

36. Moda-Sava RN, Murdock MH, Parekh PK, et al. Sustained rescue of prefrontal circuit dysfunction by antidepressant-induced spine formation. *Science*. 2019;364(6436):eaat8078. doi:10.1126/science.aat8078

SUPPLEMENT 1.

Trial Protocol and Statistical Analysis Plan

SUPPLEMENT 2.

eTable 1. Rates of Remission at Different Visits by Intervention and Control Group
eFigure. The Least-Square Mean Reduction of MADRS Scores From Baseline to 6 Weeks at Each Visit by Intervention and Control Group
eTable 2. Reduction of CGI-S Score From Baseline at Different Visits by Intervention and Control Group

SUPPLEMENT 3.

Data Sharing Statement

GeroScience
https://doi.org/10.1007/s11357-023-00891-6

ORIGINAL ARTICLE

Associations of healthy aging index and all-cause and cause-specific mortality: a prospective cohort study of UK Biobank participants

Zhenhuang Zhuang · Yimin Zhao · Ninghao Huang · Yueying Li · Wenxiu Wang · Zimin Song · Xue Dong · Wendi Xiao · Jinzhu Jia · Zhonghua Liu · Lu Qi · Tao Huang

Received: 28 March 2023 / Accepted: 24 July 2023

Abstract The healthy aging index (HAI) has been recently developed as a surrogate measure of biological age. However, to what extent the HAI is associated with all-cause and cause-specific mortality and whether this association differs in younger and older adults remains unknown. We aimed to quantify the association between the HAI and mortality in a population of UK adults. In the prospective cohort study, data are obtained from the UK Biobank. Five HAI components (systolic blood pressure, reaction time, cystatin C, serum glucose, forced vital capacity) were scored 0 (healthiest), 1, and 2 (unhealthiest) according to sex-specific tertiles or clinically relevant cut-points and summed to construct the HAI (range 0–10). Cox proportional hazard regression models were used to estimate the associations of the HAI with the risk of all-cause and cause-specific mortality. 387,794 middle-aged and older participants were followed up for a median of 8.9 years (IQR 8.3–9.5). A total of 14,112 all-cause deaths were documented. After adjustments, each 1-point increase in the HAI was related to a higher risk of all-cause mortality (hazards ratio [HR], 1.17; 95%CI, 1.15–1.18). Such association was stronger among adults younger than 60 years (1.19, 1.17–1.21) than that among those 60 years and older (1.15, 1.14–1.17) (P interaction < 0.001). For each unit increment of the HAI, the multivariate-adjusted HRs for risk of death were 1.28 (1.25–1.31) for cardiovascular diseases, 1.09 (1.07–1.10) for cancer, 1.36 (1.29–1.44) for digestive disease, 1.42 (1.35–1.48) for respiratory disease, 1.42 (1.33–1.51) for infectious diseases, and

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11357-023-00891-6.

Z. Zhuang · Y. Zhao · N. Huang · Y. Li · W. Wang · Z. Song · X. Dong · W. Xiao · T. Huang (✉)
Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing, China
e-mail: huangtao@bjmu.edu.cn

J. Jia
Department of Biostatistics, School of Public Health, Peking University, Beijing, China

Z. Liu
Department of Biostatistics, Columbia University, New York, NY, USA

L. Qi (✉)
Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA
e-mail: lqi1@tulane.edu

L. Qi
Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

T. Huang
Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China

T. Huang
Center for Intelligent Public Health, Institute for Artificial Intelligence, Peking University, Beijing, China

Published online: 01 August 2023

1.15 (1.09–1.21) for neurodegenerative disease, respectively. Our findings indicate that the HAI is positively associated with all-cause and cause-specific mortality independent of chronological age. Our results further underscore the importance of effective early-life interventions to slow aging and prevent premature death.

Keywords Healthy aging · Mortality · Cohort study · Biological age

Introduction

Rapid population aging has posed severe threats to health and social care systems [1]. Between 2019 and 2050, the number of persons aged 65 or over worldwide is projected to more than double (1.5 billion), which will outnumber adolescents and youth aged 15 to 24 years (1.3 billion) [2]. Nevertheless, between-person variations in the pace of aging may exist among individuals of the same chronological age, which manifest as differences in susceptibility to disease and death [3, 4]. Therefore, it is of paramount importance to study the biological determinants of the between-person variations in "healthy aging."

While several aging indicators have been proposed in the UK Biobank, such as the biological age based on 72 biomarkers [5] and the frailty index based on 49 items [6], which have provided multisystem approaches to research and prevention of diseases of aging, surrogate aging measures with relative affordability and practicality are needed to be applicable to the clinical settings. The healthy aging index (HAI), a modified Physiological Index of Comorbidity developed by Newman et al. [7], has been associated with increased mortality risk in older adults (≥60 years) in the United States (US) and China [8–13]. Using only five noninvasive clinical tests (systolic blood pressure, reaction time, cystatin C, serum glucose, forced vital capacity), the HAI not only captures subclinical hypofunction across multiple organ systems, but also provides incremental value for death prediction beyond clinically diagnosed chronic diseases among elders [12]. The global proportion of older individuals is increasing, and this phenomenon is more pronounced in the UK than in the US and China, which may result in a growing risk of adverse age-related outcomes in the UK [14]. Although the HAI has been applied in the UK Biobank recently and has a significant impact on major vascular events [15], the association of the HAI with all-cause and cause mortality risk remains unclear.

In the present study, we aimed to assess whether an adaptation of the HAI using data from UK Biobank (a large prospective cohort of over 500,000 middle-aged people) can be used to identify people at increased risk of all-cause and cause-specific mortality across a wide range of ages; and to further investigate whether the effect of HAI on mortality risk could be modified by sociodemographic characteristics, lifestyle factors, and health status.

Method

Study participants

In this prospective cohort study, we sourced data from the UK Biobank. Details of the design and survey methods of the UK Biobank have been described elsewhere [16–18]. In brief, UK Biobank, a general population cohort study, recruited more than 500,000 middle-aged participants between April 2007 and December 2010 (5.5% response rate for around 9 million people who received invitations). The baseline survey was done in 22 assessment centers across England, Wales, and Scotland, where participants completed a touch screen questionnaire, took physical measurements, and provided biological samples.

From the initial sample of 502,527 participants, we excluded those with missing information on all five HAI components [SBP, reaction time, cystatin C, serum glucose, and FVC; $n=113{,}372$), and other key covariates [race/ethnicity; $n=1361$], leaving 387,794 participants included in the present study. Written informed consent was provided by all participants, and the study was approved by the North West Multicenter Research Ethics Committee. We conducted this research using the UK Biobank Resource under Application Number 44430. Ethical approval was also obtained from the Ethical Committee of Peking University (Beijing, China).

Index components

Systolic blood pressure

Two SBP measurements were taken in a seated position after a 2-min rest using an appropriate cuff and an Omron HEM-7015IT digital BP monitor, and in this study, the average was calculated. For individuals with missing automated SBP readings, we used the mean of these 2 manual values for imputation. We used sex-specific tertiles of SBP values to classify participants into three categories: 0 = < 135 mmHg for men and < 127 mmHg for women; 1 = 135–150 mmHg for men and 127–146 mmHg for women; and 2 = ≥ 150 mmHg for men and ≥ 146 mmHg for women. Participants who reported a physician diagnosis of hypertension or were taking BP-lowering medications were also grouped in the unhealthiest category (component score = 2).

Reaction time

Although previous studies of HAI have identified the Digit Symbol Substitution Test (DSST) as a potential indicator of cognitive performance [8, 13], only a subset of participants who accepted the invitation to the follow-up assessment have completed the DSST in the UK Biobank [19]. Given the strong correlation between DSST score and reaction time, we proposed that reaction time could substitute for DSST, both of which fall within the cognitive domain of processing speed. Although the replacement measurement may not represent a phenotype identical to that represented by the original components, they are nonetheless indicators of neurological health. Reaction times were collected via a touchscreen at the baseline assessment, in which participants were asked to complete a computerized version of the card game "Snap," pressing a button as quickly as possible when two cards displayed on the screen matched. Scores consisted of the average time to give a correct response. Sex-specific tertiles of reaction time were applied to classify participants into three groups: 0 = < 493 ms for men and < 509 ms for women; 1 = 493–574 ms for men and 509–591 ms for women; and 2 = ≥ 574 ms for men and ≥ 591 ms for women.

Cystatin C

Cystatin C was measured by latex-enhanced immunoturbidimetric analysis on a Siemens ADVIA 1800, which was scored using sex-specific tertiles of the cystatin C values: 0 = < 0.87 mg/dL for men and < 0.81 mg/dL for women; 1 = 0.87–0.98 mg/dL for men and 0.81–0.91 mg/dL for women; and 2 = ≥ 0.98 mg/dL for men and ≥ 0.91 mg/dL for women.

Serum glucose

Random glucose was measured by hexokinase analysis on a Beckman Coulter AU5800 for all participants with an available blood sample at baseline. Given the majority (> 95%) of the participants who took the random test fasted < 8 h, we classified participants into 3 groups according to clinical cutoff points of random glucose as follows: 0 = ≤ 140 mg/dL, 1 = 140–200 mg/dL, and 2 = ≥ 200 mg/dL [20]. In addition, participants who reported a diagnosis of diabetes or who were using medication for diabetes were coded as 2.

Forced vital capacity

FVC value was calculated from blow using a Vitalograph Pneumotrac 6800. We divided participants into three FVC categories using sex-specific tertiles: 0 = ≥ 4750 mL for men and ≥ 3400 mL mg/dL for women; 1 = 4000–4750 mL for men and 2870–3400 mL for women; and 2 = < 4000 mL for men and < 2870 mL for women.

All the above HAI components are known risk factors for death [21–27]. The sum of these five scores together gave a total HAI score ranging from 0 to 10, with a higher value indicating a worse, unhealthier aging status. The summary of cutoffs used in this study can be seen in eTable 1. With reference to previous studies [8, 9, 13, 28], we further categorized HAI scores into the following four levels: 0–2 (healthiest), 3–4, 5–6, and 7–10 (unhealthiest).

Mortality status follow-up

Long-term follow-up for mortality was done in the UK Biobank through comprehensive data linkage.

Date and cause of death were obtained from the National Health Service (NHS) Information Centre for participants from England and Wales, and from the NHS Central Register, Scotland, for participants from Scotland. Details of information about the linkage procedure can be found at http://content.digital.nhs.uk/services. Linkage recorded all deaths occurring until 14 February 2018 for England and 31 December 2016 for Scotland. The outcomes of this study were all-cause mortality and cause-specific mortality (cardiovascular diseases, cancer, respiratory disease, digestive disease, neurodegenerative disease, infectious disease, and all other causes), based on the 10th revision of the International Classification of Diseases (ICD-10) (eTable 2).

Covariates

This study utilized the following potential confounding factors: sociodemographic characteristics [age, sex, ethnicity, UK Biobank assessment center, social deprivation (Townsend deprivation index), and highest education levels]; anthropometric indicators [body mass index (BMI)]; lifestyle behaviors [smoking status (current, former, or never), alcohol intake, physical activity, and diet], personal health and medical history [cardiovascular diseases, cancer, respiratory disease, digestive disease, neurodegenerative disease, infectious disease, and medication use (anti-hypertensive medication, cholesterol-lowering medication, hormone replacement therapy, oral contraceptive pill, or none)], and biomarkers [high-density lipoprotein, low-density lipoprotein, triglycerides, and total cholesterol].

Daily physical activity levels were calculated by multiplying the metabolic equivalent tasks (METs) value for a specific type of physical activity by hours spent on that activity per week and then adding up the MET hours for all activities. Overall dietary pattern was assessed by healthy diet score including the following factors: fruit intake at least three pieces each day; vegetable intake at least four tablespoons each day; fish intake at least twice each week; processed meat intake no more than twice each week; and unprocessed red meat intake no more than twice each week. We added each healthy dietary factor to calculate the final diet score ranging from 0 to 5 [29, 30]. Information on prevalent diseases was obtained through self-reported and hospital inpatient records (eTable 3). For continuous variables, missing values were imputed sex-specific median values, while categorical variables were handled with missing indicators.

Statistical analysis

Baseline characteristics were described across four categories of HAI (i.e., 0–2, 3–4, 5-–6, and 7–10) as means (SD) for continuous variables and percentages for categorical variables. In addition, we plotted the mean HAI and the prevalence of unhealthiest aging status by age and sex to examine the associations of the HAI with age.

In the full cohort, we used Cox proportional hazard regression models to estimate the hazard ratios (HRs) and 95% confidence intervals (CIs) of the risk of all-cause and cause-specific mortality associated with the HAI, which was included in the models as a continuous or a categorical variable. Follow-up person-years were calculated from baseline until the date of death, or end of follow-up, whichever occurred first. No violations to the proportional hazards assumption were observed by inspection of the plots of the Schoenfeld residual. We undertook separate models by sequential inclusion of three groups of covariates to examine the different potential confounding effects on the associations between HAI levels and mortality risk. Model 1 included basic sociodemographic characteristics (sex, age, UK Biobank assessment center, and race/ethnicity), followed by model 2 that additionally included other sociodemographic characteristics (Townsend deprivation index and highest education level), anthropometric indicators (BMI), and lifestyle factors (smoking status, alcohol intake, and physical activity). Model 3 additionally included chronic health conditions (eTable 3) and medication use (anti-hypertensive medication, cholesterol-lowering medication, hormone replacement therapy, oral contraceptive pill, or none). Notably, we adjusted for all the chronic conditions in the all-cause analysis, and for only one disease in the corresponding cause-specific analyses. To visually explore non-linear associations between HAI and mortality risk, we used restricted cubic splines for HAI score with 3 knots at the 25th, 50th, and 75th percentiles of the distribution, stratified by sex and baseline age. We also applied a Kaplan–Meier survival curve to compare survival probabilities after baseline recruitment between different groups of HAI

stratified by sex and baseline age. In addition, interaction terms of potential covariates (age, sex, smoking status, alcohol intake, level of physical activity, and BMI) were tested using the likelihood ratio test, which involved comparing models with and without interaction terms, and we performed subgroup analyses. Considering that most studies of the HAI used 60 years as the cutoff of age, we conducted subgroup analysis by age (≥ 60 vs. < 60 years) for convenience of comparison, with 60 years old also being the median age. Additionally, we also calculated the population attributable risk percent (PAR%) to estimate the proportion of death that theoretically would not have occurred if all participants were in a low-risk group, in the case of assuming causality.

To examine the robustness of our findings, we also conducted several sensitivity analyses: additionally adjusting for dietary factors; excluding individuals with one or more of these diseases at baseline from the analysis, either separately or together; excluding those who had ever used medications; excluding current smokers and current drinker. In order to avoid bias from the imputation of missing covariates, participants with missing values were also removed in another sensitivity analysis. In addition, we excluded those who had died during the first 5 years of follow-up to minimize the influence of reverse causality.

All statistical analyses were conducted using Stata version 16.0. All *p* values were two-sided and we defined the level of statistical significance as a *P* value less than 0.05.

Patient and public involvement

No patients or the public were involved in setting the research question or the outcome measures, nor were they involved in developing plans for the design or implementation of the study. No patients were asked to advise on the interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

Population characteristics

Of the 502,527 participants recruited to UK Biobank, 387,794 (77%) who had complete data on HAI and key covariates were included, with a median follow-up of 8.9 years (IQR 8.3–9.5; total follow-up 3,425,798 person-years). Table 1 summarizes the main characteristics of the participants by quarters of HAI. Among 387,794 participants (mean age 56.4 years, 45.8% men), 90,804 (23.4%) had a score of 0–2, 122,574 (31.6%) had a score of 3–4, 113,505 (29.3%) had a score of 5–6, and 60,911 (15.7%) had a score of 7–10. Participants with lower HAI were more likely to be women, White people, younger, non-smoker, and more educated, and less likely to use medications, and to have lower Townsend index, and a lower prevalence of comorbidities, but higher level of physical activity. There were no significant differences between the included and the whole populations from UK Biobank in terms of most baseline characteristics.

The HAI showed a right-skewed distribution towards older age, either in combination or separately (eFigs. 1 and 2). The mean HAI across all participants was 4.23 (SD 2.16). Both the mean HAI and the prevalence of unhealthiest aging status increased with age. The prevalence of unhealthiest aging status increased from 2319 (2.5%) in 92,794 people aged < 50 years to 12,827 (9.9%) in 129,870 people aged 50–60 years, and to 45,765 (27.7%) in 165,130 people aged > 60 years. As shown in Fig. 1, women had a higher mean HAI between the ages 40 and 59 years, while men had a higher prevalence of unhealthiest aging status across all age subgroups.

Association of healthy aging index with all-cause mortality

During follow-up, 14,112 all-cause deaths were documented. Absolute mortality rates according to HAI categories were 1.40, 2.86, 4.92, and 9.42 deaths per 1000 person-years for participants with a score of 0–2, 3–4, 5–6, and 7–10, respectively (Table 2). As shown in Table 2, each HAI component was significantly associated with all-cause mortality; after further adjustment, the magnitude of associations was slightly attenuated in models 2 and 3. Among all components, serum random glucose and cystatin C were most strongly associated with mortality. When all 5 components were combined, the multivariate-adjusted HRs for participants scoring 3–4, 5–6, and 7–10, as compared

Table 1 Baseline characteristics of study participants according to HAI

Characteristics	HAI				Samples included in the study ($n=387,794$)	UK Biobank Full samples ($n=502,527$)
	0–2 Healthy ($n=38,694$)	3–4 ($n=57,358$)	5–6 ($n=52,580$)	7–10 Unhealthy ($n=29,030$)		
White (%)	97.0%	95.7%	94.2%	91.4%	94.9%	94.6%
Male (%)	42.6%	46.8%	46.3%	47.7%	45.8%	45.6%
Age at baseline (year)	50.4 (7.2)	55.3 (7.7)	59.4 (7.0)	62.2 (5.9)	56.4 (8.1)	56.5 (8.1)
Townsend deprivation index	− 1.6 (2.9)	− 1.5 (3.0)	− 1.3 (3.1)	− 0.8 (3.3)	− 1.4 (3.1)	− 1.3 (3.1)
College or university degree (%)	46.0%	37.8%	29.9%	22.0%	33.1%	32.7%
Body mass index (kg/m^2)	26.2 (3.4)	27.3 (3.8)	28.4 (4.2)	29.9 (4.8)	27.4 (4.7)	27.4 (4.8)
Physical activity (MET h/week)	2907.5 (2951.7)	2891.7 (2988.2)	2817.1 (2959.4)	2533.4 (2786.9)	2663.1 (2710.6)	2650.2 (2713.1)
Smoking status (%)						
Never	58.4%	52.5%	46.1%	38.9%	55.2%	54.8%
Past	31.2%	35.7%	41.2%	47.2%	34.6%	34.6%
Current	10.5%	11.8%	12.7%	13.9%	10.2%	10.6%
Alcohol consumption (g/day)	21.2 (20.3)	22.0 (22.5)	21.5 (22.6)	18.9 (22.2)	14.8 (18.2)	14.7 (18.4)
Medications (%)						
Cholesterol-lowering medication	5.9%	14.5%	28.0%	50.2%	17.0%	17.6%
Blood pressure medication	1.1%	6.8%	14.6%	19.4%	10.0%	10.2%
Insulin	0.0%	0.1%	0.2%	0.5%	0.2%	0.2%
Chronic conditions (%)						
Diabetes	0.2%	1.3%	5.5%	31.2%	5.4%	5.4%
Hypertension	3.9%	21.2%	43.2%	69.1%	28.0%	28.7%
Cardiovascular disease	2.3%	5.3%	10.6%	21.3%	6.2%	6.9%
Cancer	4.1%	6.3%	8.2%	10.1%	8.8%	9.3%
Respiratory disease	0.7%	1.2%	2.3%	4.0%	1.7%	2.3%
Digest disease	0.1%	0.3%	0.4%	0.6%	0.2%	0.3%
Neurodegenerative disease	3.0%	3.7%	4.8%	6.7%	4.3%	4.7%
Infectious diseases	0.2%	0.3%	0.6%	1.2%	0.4%	0.5%
Single item of HAI						
Systolic blood pressure (mmHg)	130.8 (12.7)	141.1 (17.1)	148.5 (18.1)	151.4 (18.9)	139.7 (19.6)	139.8 (19.7)
Random glucose (mg/dl)	4.9 (0.7)	5.0 (0.9)	5.2 (1.3)	5.9 (2.3)	5.1 (1.2)	5.1 (1.2)

Table 1 (continued)

Characteristics	HAI				Samples included in the study ($n=387,794$)	UK Biobank Full samples ($n=502,527$)
	0–2 Healthy ($n=38,694$)	3–4 ($n=57,358$)	5–6 ($n=52,580$)	7–10 Unhealthy ($n=29,030$)		
Forced vital capacity (liters)	5.2 (0.9)	4.6 (0.9)	4.1 (0.8)	3.6 (0.7)	3.7 (1.1)	3.7 (1.1)
Cystatin C (mg/dl)	0.8 (0.1)	0.9 (0.1)	1.0 (0.2)	1.1 (0.2)	0.9 (0.2)	0.9 (0.2)
Reaction time (ms)	477.4 (67.6)	528.9 (99.4)	576.5 (116.8)	635.7 (125.4)	558.2 (116.9)	559.6 (118.0)

Mean (SD) is presented for continuous variables and percentage for categorical variables. Percentages may not total 100 because of rounding. All exposures were associated with HAI, with $P<0.001$ for trends across categories

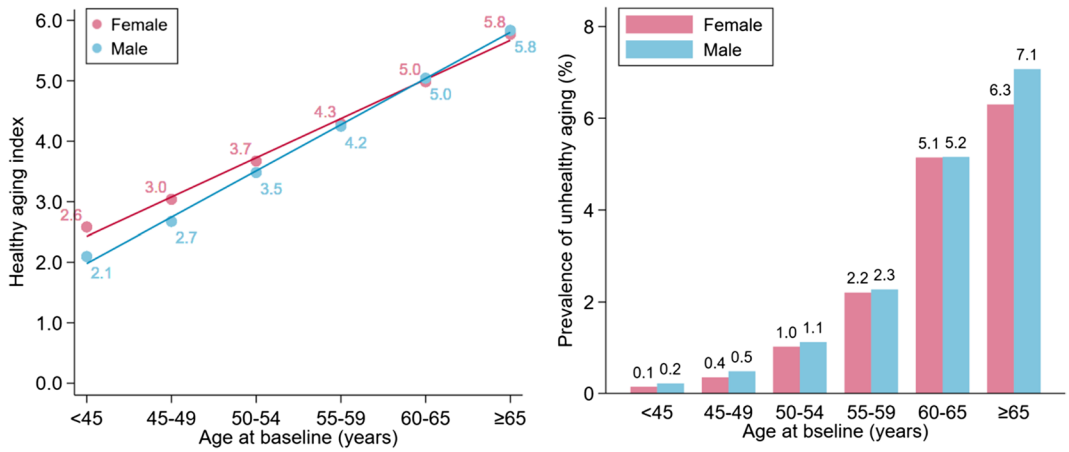


Fig. 1 Mean HAI and prevalence of unhealthy aging by age and sex. **A** The data points represent the mean value of the HAI per each 5-year age group and the lines represent the fitted curve of the HAI. **B** The histogram represents the prevalence of unhealthy aging with a HAI of 7–10 per each 5-year age group

with those scoring 0–2, were 1.31 (1.22, 1.40), 1.60 (1.49, 1.72), and 2.28 (2.11, 2.46) (P for trend < 0.001). Each 1-point increase in HAI was associated with a 17% higher risk of all-cause mortality (HR 1.17; 95%CI 1.15, 1.18). As presented in Kaplan–Meier survival curves, there was a graded decrease in the survival probabilities as HAI increased, and the difference between the HAI categories in survival probability over time was larger in men and increased with baseline age (eFig. 3). For participants with healthy aging status (scoring 3–10), the PAR of all-cause mortality was 29.8% (95% CI 25.6%, 33.8%), suggesting that approximately one-third of the deaths might have been prevented if all participants had not been in the healthy aging status, regardless of chronological age. Estimates of the PAR for the alternative levels of HAI or individual components can be found in eTable 4. In addition, we found potential evidence of a non-linear dose-response association between HAI and mortality risk (P non-linearity < 0.0001, eFig. 4).

Association of healthy aging index with cause-specific mortality

Over the follow-up period, 2848 (0.7%) participants died from cardiovascular disease, 8107 (2.1%)

Table 2 Associations between single and combined HAI components and risk of all-cause mortality

	Events per 1000 person-year (95%CI)	HR (95% CI)		
		Model 1	Model 2	Model 3
Systolic blood pressure†				
0	2.64 (2.54, 2.74)	(Reference)	(Reference)	(Reference)
1	3.10 (2.98, 3.22)	0.94 (0.89, 0.99)	0.95 (0.90, 1.00)	0.97 (0.92, 1.02)
2	5.66 (5.54, 5.78)	1.13 (1.08, 1.18)	1.14 (1.09, 1.19)	1.14 (1.09, 1.19)
Random glucose†				
0	3.74 (3.68, 3.81)	(Reference)	(Reference)	(Reference)
1	6.43 (5.31, 7.56)	1.41 (1.18, 1.68)	1.40 (1.18, 1.67)	1.39 (1.17, 1.66)
2	10.34 (9.88, 10.81)	1.80 (1.72, 1.90)	1.75 (1.66, 1.84)	1.65 (1.57, 1.75)
Forced vital capacity†				
0	2.32 (2.24, 2.41)	(Reference)	(Reference)	(Reference)
1	3.53 (3.43, 3.64)	1.08 (1.03, 1.13)	1.05 (1.00, 1.10)	1.04 (0.99, 1.09)
2	6.55 (6.41, 6.70)	1.50 (1.43, 1.57)	1.37 (1.31, 1.44)	1.30 (1.24, 1.37)
Cystatin C†				
0	2.31 (2.22, 2.40)	(Reference)	(Reference)	(Reference)
1	3.34 (3.23, 3.45)	1.14 (1.08, 1.20)	1.10 (1.05, 1.16)	1.10 (1.04, 1.15)
2	6.88 (6.73, 7.04)	1.77 (1.69, 1.85)	1.60 (1.53, 1.68)	1.53 (1.46, 1.61)
Reaction time†				
0	2.77 (2.68, 2.86)	(Reference)	(Reference)	(Reference)
1	3.93 (3.81, 4.04)	1.06 (1.02, 1.11)	1.04 (1.00, 1.09)	1.04 (0.99, 1.09)
2	5.79 (5.65, 5.93)	1.29 (1.23, 1.35)	1.23 (1.18, 1.28)	1.20 (1.15, 1.26)
Total HAI score				
0–2	1.40 (1.32, 1.48)	(Reference)	(Reference)	(Reference)
3–4	2.86 (2.76, 2.96)	1.42 (1.32, 1.52)	1.32 (1.23, 1.41)	1.31 (1.22, 1.40)
5–6	4.92 (4.78, 5.06)	1.92 (1.79, 2.06)	1.66 (1.55, 1.78)	1.60 (1.49, 1.72)
7–10	9.42 (9.16, 9.68)	3.17 (2.96, 3.41)	2.51 (2.33, 2.71)	2.28 (2.11, 2.46)
P for trend	-	< 0.001	< 0.001	< 0.001
Per score point	-	1.23 (1.22, 1.24)	1.19 (1.18, 1.20)	1.17 (1.15, 1.18)

†All of the 5 individual components were included in the model simultaneously

Values are hazard ratios (95% confidence intervals) unless stated otherwise

Model 1: adjusted for sex, age, UK Biobank center, and race/ethnicity

Model 2: model 1 + education, TDI, body mass index, physical activity, smoking status, and alcohol intake

Model 3: model 2 + medications, and chronic conditions

from cancer, 530 (0.1%) from digestive disease, 757 (0.2%) from respiratory disease, 369 (0.1%) from infectious disease, and 575 (0.2%) from neurodegenerative disease. A strong graded increase in the risk of cause-specific mortality was also observed across the HAI categories (Table 3). For each unit increment of the HAI, the corresponding HRs for risk of death were 1.28 (1.25, 1.31) for cardiovascular diseases, 1.09 (1.07, 1.10) for cancer, 1.36 (1.29, 1.44) for digestive disease, 1.42 (1.35, 1.48) for respiratory disease, 1.42 (1.33, 1.51) for infectious diseases, 1.15 (1.09, 1.21) for neurodegenerative disease, and 1.33 (1.28, 1.38) for all other causes. The PAR of cause-specific mortality ranged from 17.1 to 86.7% for participants who had a relatively higher HAI of 3–10, and the reclassification of HAI groups yielded similar estimates (eTable 4). Furthermore, we found linear associations with the risk of death for cardiovascular diseases (*P* non-linearity = 0.059), cancer (*P* non-linearity = 0.100), digestive disease (*P* non-linearity = 0.720), neurodegenerative disease (*P* non-linearity = 0.822), and infectious disease (*P*

Table 3 Association between HAI and cause-specific mortality

Cause of death	HAI				*P* for trend	Each unit increment
	0–2 Healthy ($n=90,804$)	3–4 ($n=122,574$)	5–6 ($n=113,505$)	7–10 Unhealthy ($n=60,911$)		
Cardiovascular diseases						
Events per 1000 person-year (95%CI)	0.17 (0.14, 0.20)	0.43 (0.39, 0.47)	0.95 (0.89, 1.01)	2.45 (2.32, 2.59)	-	-
Model 1 [HR (95% CI)]	(Reference)	1.83 (1.51, 2.22)	3.26 (2.70, 3.93)	7.25 (5.99, 8.77)	< 0.001	1.42 (1.39, 1.45)
Model 2 [HR (95% CI)]	(Reference)	1.60 (1.31, 1.94)	2.48 (2.05, 3.00)	4.69 (3.85, 5.71)	< 0.001	1.33 (1.30, 1.36)
Model 3 [HR (95% CI)]	(Reference)	1.59 (1.31, 1.93)	2.37 (1.95, 2.86)	4.03 (3.31, 4.92)	< 0.001	1.28 (1.25, 1.31)
Cancer						
Events per 1000 person-year (95%CI)	0.99 (0.92, 1.06)	1.85 (1.77, 1.94)	2.90 (2.79, 3.00)	4.54 (4.36, 4.72)	-	-
Model 1 [HR (95% CI)]	(Reference)	1.24 (1.14, 1.35)	1.47 (1.35, 1.60)	1.96 (1.79, 2.14)	< 0.001	1.12 (1.11, 1.14)
Model 2 [HR (95% CI)]	(Reference)	1.17 (1.07, 1.27)	1.31 (1.20, 1.42)	1.62 (1.48, 1.78)	< 0.001	1.09 (1.08, 1.11)
Model 3 [HR (95% CI)]	(Reference)	1.16 (1.06, 1.26)	1.29 (1.18, 1.41)	1.59 (1.44, 1.74)	< 0.001	1.09 (1.07, 1.10)
Digestive disease						
Events per 1000 person-year (95%CI)	0.04 (0.02, 0.05)	0.08 (0.06, 0.09)	0.20 (0.17, 0.23)	0.42 (0.36, 0.48)	-	-
Model 1 [HR (95% CI)]	(Reference)	1.88 (1.23, 2.89)	4.59 (3.04, 6.93)	9.42 (6.18, 14.37)	< 0.001	1.47 (1.41, 1.54)
Model 2 [HR (95% CI)]	(Reference)	1.56 (1.01, 2.41)	3.47 (2.28, 5.26)	6.29 (4.06, 9.73)	< 0.001	1.41 (1.34, 1.49)
Model 3 [HR (95% CI)]	(Reference)	1.47 (0.96, 2.27)	3.10 (2.04, 4.70)	5.12 (3.30, 7.95)	< 0.001	1.36 (1.29, 1.44)
Respiratory disease						
Events per 1000 person-year (95%CI)	0.01 (0.00, 0.02)	0.10 (0.08, 0.12)	0.25 (0.22, 0.28)	0.74 (0.66, 0.81)	-	-
Model 1 [HR (95% CI)]	(Reference)	6.18 (3.00, 12.72)	11.59 (5.68, 23.63)	27.76 (13.60, 56.67)	< 0.001	1.50 (1.44, 1.56)
Model 2 [HR (95% CI)]	(Reference)	5.67 (2.75, 11.67)	9.75 (4.77, 19.93)	21.08 (10.26, 43.31)	< 0.001	1.46 (1.39, 1.52)
Model 3 [HR (95% CI)]	(Reference)	5.71 (2.77, 11.77)	9.23 (4.51, 18.91)	18.44 (8.95, 38.00)	< 0.001	1.42 (1.35, 1.48)
Infectious diseases						
Events per 1000 person-year (95%CI)	0.02 (0.01, 0.03)	0.06 (0.04, 0.07)	0.12 (0.10, 0.14)	0.33 (0.28, 0.38)	-	-
Model 1 [HR (95% CI)]	(Reference)	2.27 (1.28, 4.03)	4.18 (2.38, 7.32)	10.18 (5.79, 17.93)	< 0.001	1.47 (1.39, 1.55)
Model 2 [HR (95% CI)]	(Reference)	2.18 (1.22, 3.87)	3.80 (2.16, 6.71)	8.67 (4.84, 15.54)	< 0.001	1.45 (1.36, 1.54)

Table 3 (continued)

Cause of death	HAI				*P* for trend	Each unit increment
	0–2 Healthy (*n* = 90,804)	3–4 (*n* = 122,574)	5–6 (*n* = 113,505)	7–10 Unhealthy (*n* = 60,911)		
Model 3 [HR (95% CI)]	(Reference)	2.19 (1.23, 3.90)	3.74 (2.11, 6.61)	8.08 (4.49, 14.54)	< 0.001	1.42 (1.33, 1.51)
Neurodegenerative disease						
Events per 1000 person-year (95%CI)	0.04 (0.03, 0.06)	0.13 (0.11, 0.15)	0.21 (0.19, 0.24)	0.35 (0.30, 0.40)	-	-
Model 1 [HR (95% CI)]	(Reference)	1.48 (1.01, 2.17)	1.53 (1.05, 2.24)	1.93 (1.31, 2.84)	< 0.001	1.11 (1.06, 1.16)
Model 2 [HR (95% CI)]	(Reference)	1.57 (1.07, 2.30)	1.71 (1.16, 2.51)	2.27 (1.52, 3.40)	< 0.001	1.14 (1.09, 1.20)
Model 3 [HR (95% CI)]	(Reference)	1.57 (1.07, 2.30)	1.75 (1.19, 2.57)	2.29 (1.53, 3.44)	< 0.001	1.15 (1.09, 1.21)
All other causes						
Events per 1000 person-year (95%CI)	0.13 (0.11, 0.16)	0.24 (0.21, 0.27)	0.36 (0.32, 0.39)	0.82 (0.74, 0.89)	-	-
Model 1 [HR (95% CI)]	(Reference)	1.73 (1.37, 2.17)	2.53 (2.01, 3.18)	5.76 (4.54, 7.31)	< 0.001	1.37 (1.33, 1.42)
Model 2 [HR (95% CI)]	(Reference)	1.61 (1.28, 2.02)	2.20 (1.73, 2.78)	4.59 (3.57, 5.90)	< 0.001	1.34 (1.29, 1.38)
Model 3 [HR (95% CI)]	(Reference)	1.60 (1.27, 2.02)	2.16 (1.70, 2.74)	4.38 (3.40, 5.65)	< 0.001	1.33 (1.28, 1.38)

Values are hazard ratios (95% confidence intervals) unless stated otherwise

Model 1: adjusted for sex, age, UK Biobank center, and race/ethnicity

Model 2: model 1 + education, TDI, body mass index, physical activity, smoking status, and alcohol intake

Model 3: model 2 + medications, and chronic conditions

non-linearity = 0.741), and non-linear associations with respiratory disease (*P* non-linearity = 0.022) and all other cause (*P* non-linearity < 0.001) (eFig. 4).

Subgroup analyses

For all-cause mortality, the effect estimates were stronger among younger adults (< 60 years) than among older adults (≥ 60 years; *P* interaction < 0.0001; Fig. 2). Compared with the reference group (HAI 0–2), the group with the HAI of 7–10 had a 1.15-fold greater risk of mortality in adults aged ≥ 60 years, while the HRs decreased to 1.19 in adults aged < 60 years. The association was also significantly different between men and women, with men generally having higher HRs associated with HAI than women (*P* interaction < 0.0001; Fig. 3). In analyses of cause-specific mortality, stratified analyses by age showed that the association of the HAI persisted in both age groups, but the association was larger in participants aged < 60 years than participants aged ≥ 60 years for cancer mortality (Fig. 2). In addition, there was a significant difference in the association of the HAI with the risk of mortality from cardiovascular diseases (*P* interaction = 0.047) and cancer (*P* interaction < 0.0001) between men and women, the difference in the effect estimates was not clinically meaningful (Fig. 3).

Notably, significant interactions between HAI and smoking status were observed in mortality outcome (*P* = 0.026 for interaction in all

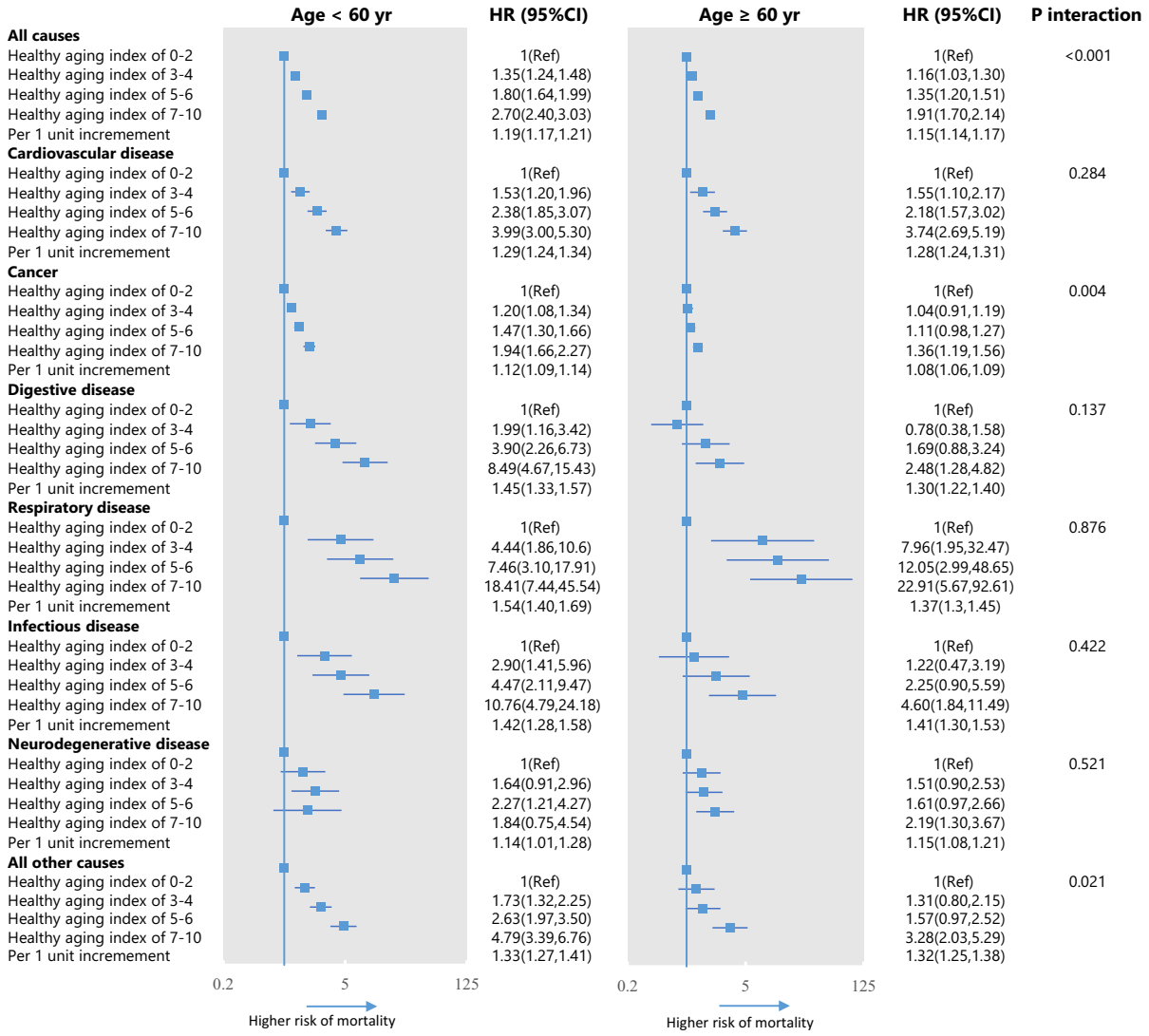


Fig. 2 The association of HAI with all-cause and cause-specific mortality by age. Models were adjusted for age, sex, UKB center, education level, TDI, alcohol intake, smoking status, physical activity, medications (lowing cholesterol, hormone, oral pills), and chronic conditions (cardiovascular disease, cancer, respiratory disease, digestive disease, neurogenerative disease, infectious disease)

participants; $P < 0.0001$ for interaction in women) (eTable 5). The test for the interaction item of physical activity was also statistically significant ($P = 0.004$ for interaction in all participants; $P = 0.021$ for interaction in women). In subgroup analyses with stratification by these covariates, the effect sizes between HAI and all-cause mortality were somewhat stronger among regular smokers (versus never or former smokers) and adults with high MET hours (versus low or moderate MET hours). However, the associations between HAI and mortality risk were similar across subgroups stratified according to alcohol intake and BMI (eTable 5).

Sensitivity analyses

In the sensitivity analyses, the associations of the HAI with all-cause and cause-specific mortality did not change appreciably with additional adjustment

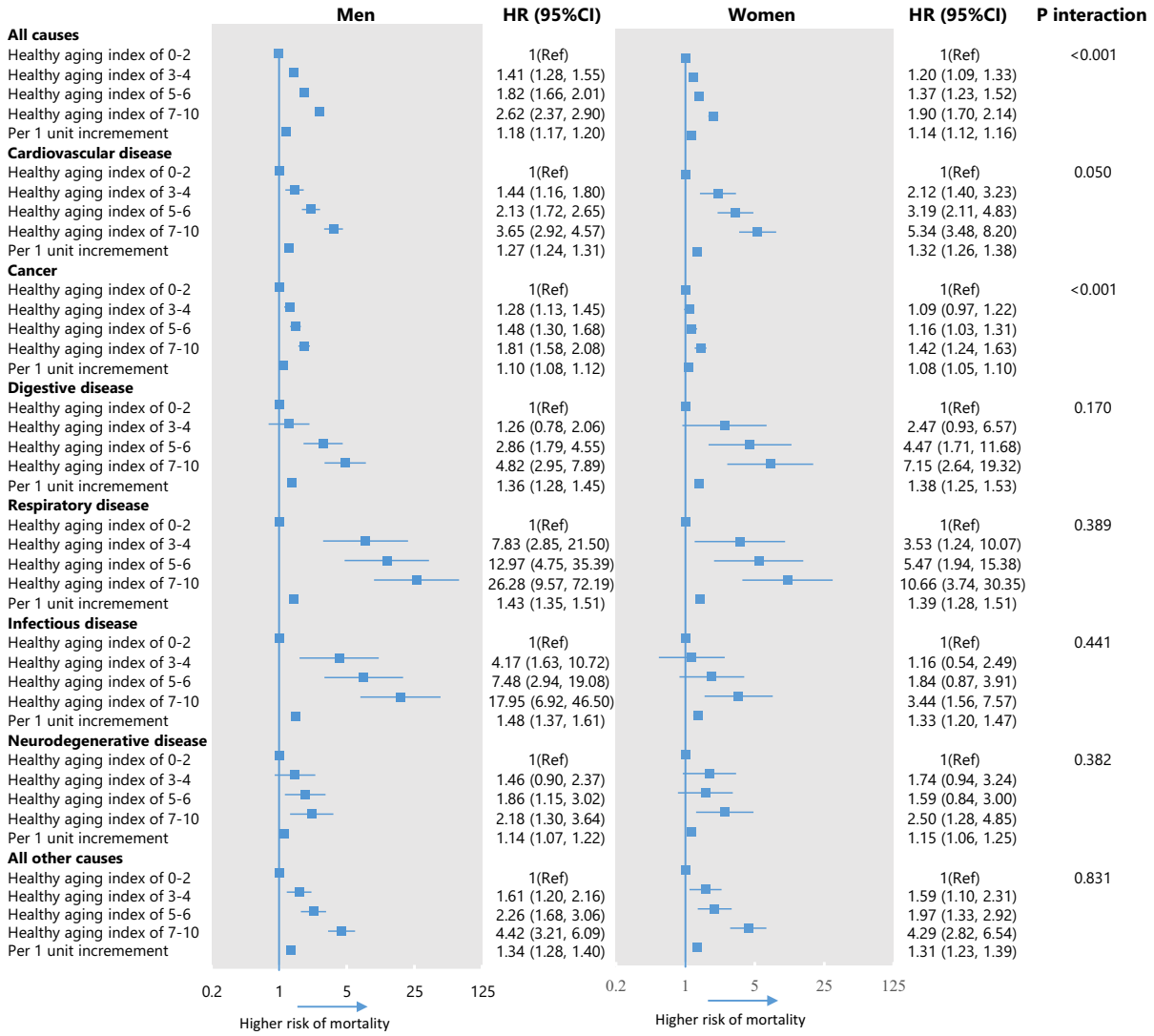


Fig. 3 The association of HAI with all-cause and cause-specific mortality by sex. Models were adjusted for age, UKB center, education level, TDI, alcohol intake, smoking status, physical activity, medications (lowing cholesterol, hormone, oral pills), and chronic conditions (cardiovascular disease, cancer, respiratory disease, digestive disease, neurogenerative disease, infectious disease)

for blood lipids, or additional adjustment for dietary factors (eTable 6). After excluding participants with diabetes, hypertension, cardiovascular disease, cancer, respiratory disease, digestive disease, neurodegenerative disease, or infectious disease at baseline, the association of the HAI with all-cause mortality and cause-specific mortality showed small to moderate changes (eTable 7). The most apparent attenuation of associations occurred in mortality risk after participants who had any of the five aforementioned diseases at baseline were removed from the analyses, but these significant associations persisted. Among relatively healthy participants, the HR for all-cause mortality per unit increment in the HAI was 1.12 (1.09, 1.14). The exclusion of people from the analyses who were current smokers, current drinkers, with missing covariates, or died during the first 5 years of follow-up only changed the HRs slightly.

Discussion

In this large-scale prospective UK cohort, we constructed a 5-item HAI to characterize healthy aging levels. Higher HAI was significantly associated with an increased risk of all-cause and cause-specific mortality with the highest risk from respiratory diseases and the lowest risk from cancer, even after adjustment for chronological age and other known risk factors. In addition, we found that the effect sizes decreased with chronological age, indicating that interventions to slow the biological aging to prevent premature death at an earlier stage of life course might be more effective.

The strength of association of the HAI with all-cause mortality reported in the present study is in line with previous research in smaller cohorts of older adults (> 60 years), primarily in the US and China [8–13]. Our finding of a HR for all-cause mortality of 1.15 for participants aged > 60 years and 1.19 for those aged < 60 years, for one point higher HAI, in our fully adjusted model, is in agreement with observations from the Framingham Offspring Study of 934 subjects (> 60 years), which reported that per unit higher of the 5-item HAI was, after adjustment, associated with a HR of 1.15 for all-cause mortality [10]. Furthermore, our findings for all-cause mortality are also comparable to those reported in the National Health and Nutrition Examination Survey, the Cardiovascular Health Study, and the Health, Aging, and Body Composition Study, with a HR of 1.19, 1.17, and 1.19, respectively. [8, 11, 13] However, China Health and Retirement Longitudinal Study ($n = 3740$) suggested that each point increase in HAI was associated with a 28% higher mortality risk in older adults (> 60 years) [12]. A potential reason for this observation may be that the measurements of HAI in the Wu et al. [8] study included C-reactive protein additionally, which is a measurement of overall inflammation rather than specific to a particular organ system as the traditional definition of HAI, leading to an overestimated effect on mortality. While another Chinese cohort, the Rugao Longevity and Ageing Study (n = 1719), has found that the HRs per unit increment of the HAI for all-cause mortality were 1.11 (95% CI 1.05–1.18), partially due to the selection criteria of participants aged 70 years or older [9]. These studies, including ours, collectively showed that public efforts to understand and promote a healthy aging status in the whole population will lead to an overall reduction in mortality risk.

Compared with all-cause mortality, we found that the HAI was more strongly associated with mortality from cardiovascular diseases, digestive diseases, respiratory diseases, infectious diseases, and all other causes. To our knowledge, only one study has analyzed the relationships between HAI and cause-specific mortality, suggesting that each point increase in HAI was associated with a 23% higher cardiovascular mortality risk while no significant association with cancer mortality [8]. This is partially similar to our finding that per unit higher HAI was associated, in the fully adjusted model, with a HR of 1.28 for mortality risk from cardiovascular diseases. However, in our study, the association between the HAI and the risk of death from cancer remained even after we excluded participants with cancer at baseline or those who died during the first 5 years of follow-up. Findings on cancer mortality warrant further confirmation because of inconsistency between study results [8]. The excess risk for cause-specific mortality may be partially due to reduced immunocompetence, progressive deterioration of organ function, or decreased likelihood of cure among people with higher HAI, leading to the development of diseases that caused death. For instance, some of the excess mortality from digestive diseases may be owing to gut dysbiosis associated with healthy aging status, which triggers the innate immune response and chronic low-grade inflammation, leading to many age-related degenerative pathologies, all together contributing to the development of digestive diseases [31–33]. As a whole, our findings support the notion that healthy aging status, which could be reflected by HAI, may have protean protective effects on all these specific systems to promote longevity.

Although the HAI has been applied in the UK Biobank with a positive effect on major vascular events [15], this is the first attempt to associate the HAI with mortality in the European population, which includes a considerably younger age range than previous studies [8–13]. Little is known about whether HAI is related to mortality risk in y ADDIN EN.CITE [15]ounger populations. We observed that in these associations for all-cause and cancer mortality, the effect estimates were moderately stronger among younger adults (< 60 years) than that among older adults (≥ 60 years). Therefore, the use of a

continuous HAI as a surrogate for biological age is more sensitive in younger adults who are at the lower end of the aging spectrum, which is meaningful for the prevention of premature death and the extension of healthy life expectancy. As longevity is a complex trait determined by a combination of genetics, environment, and stochasticity, accelerated aging may play a more critical role in younger participants (< 60 years) who lives longer, while deaths in elders (≥ 60 years) may occur more stochastically [34]. Such findings suggest that more studies focusing on healthy aging across the age spectrum and the development of relevant screening and intervention strategies are needed.

It is well known that women live longer than men, while the reasons are still largely unknown. Sex gap in healthy aging and survival prospects was also noticed in this study, which are often attributed to cultural differences (e.g., lifestyle factors and poor social status) in common thinking [35, 36]. However, after we adjusted for these confounders, the association between healthy aging and death was still stronger in men than in women. We assumed that biological factors, such as sex hormones and telomere shortening, may at least partly contribute to such sex differences [37–39]. For instance, estrogen has been related to lower risks of various life-threatening diseases, while testosterone has been associated with higher risks [40]. In addition, compared with women, men have greater telomere shortening, which is associated with accelerated aging [38, 41]. Further studies focusing on the underlying mechanism of the sex differences in unhealthy aging and death are warranted.

The finding that the HAI was able to stratify the risk of mortality even among apparently healthy adults, with no clinically diagnosed chronic conditions, is also novel. The HAI, which measures the burden of subclinical disease in multiple physiological systems, may therefore provide a complementary explanation of the variation in survival in clinically healthy adults. Compared with the biological age based on 72 biomarkers [5] and the frailty index based on 49 items [6], which have shown significant associations with mortality risk, the HAI has better performance with relative affordability and practicality. We also found that unhealthy lifestyle factors such as smoking and low physical activity level may further increase the detrimental impacts of HAI on mortality risk, indicating the positive influence of smoking cessation and regular exercise to delay aging [42–44]. Therefore, in a clinical setting, HAI may be a marker which tracks the effect of aging before diseases occurred to stratify risk, especially among younger adults and men as well as individuals who smoke regularly and lack exercise.

The present study has several strengths. First, this study was a large-scale prospective cohort study ($n = 387{,}794$) based on data from a geographically and socioeconomically diverse population in the UK. The inclusion of 222,664 adults younger than 60 years with a median follow-up of 8.9 years allowed a thorough analysis of cause-specific mortality in younger adults. Second, since measures of several potential confounding factors (personal, lifestyle, and chronic conditions) were available, we were able to adjust for these factors in our models. Thirdly, several sensitivity analyses were performed to confirm our findings. Finally, we constructed the HAI using objectively measured traits, which have been suggested to be more robust than self-reported health measures obtained from questionnaire information. It is worth noting that there are intricate connections between various health markers. Therefore, it is important to establish the HAI combining these various markers, which could be associated with human health or disease risk in a more concerted manner compared to just looking at these concepts individually. From the perspective of public health, the use of the simple scoring algorithm makes the results of epidemiological research easier to be interpreted and translated into practice, thereby being more informative for the general population. Despite these strengths, our study had some limitations. First, the assessment of HAI components was conducted only at one time point at baseline. The healthy aging status might have altered over the follow-up period, which could result in an underestimate of the magnitude of the true association. Second, although in our study we considered a variety of potential confounding factors, including personal variables, lifestyle factors, and history of chronic conditions, residual and unmeasured confounding might have influenced the association. Third, a large proportion of participants were excluded based on missing data in the present study, which may lead to selection bias. However, we have conducted a sensitivity

analysis and found that the baseline characteristics of these participants excluded are similar to our analytic sample (eTable 8). Finally, as a prospective cohort, the UK Biobank was not representative of the general UK population. Therefore, the generalizability to the broader UK population or the populations of other races and ethnicities outsides the UK should be made with caution.

Our findings from the large prospective cohort study of the UK population for the first time show the independent associations between the HAI and the risk of all-cause and cause-specific mortality, with a stronger association in younger adults (< 60 years) than that in older adults (≥ 60 years). Our results suggest that the HAI is a better predictor of mortality in younger adults than in those older. Further research is needed to explore the use of these routine clinical measures in clinical settings in identifying individuals whose biological age exceeds their chronological age, especially among younger adults. The identification of high-risk individuals may facilitate the development of effective intervention strategies to prevent premature death and improve the quality of life.

Author contribution T. H. and L. Q. designed the research. Z. Z. and T. H. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Z. Z. wrote the paper and performed the data analysis. All authors contributed to the statistical analysis, critically reviewed the manuscript during the writing process, and approved the final version to be published. Z. Z. and T. H. are the guarantors for the study.

Funding The study was supported by grants from the National Key R&D Program of China (2020YFC2003401), the National Natural Science Foundation of China (82173499), and the High-Performance Computing Platform of Peking University. The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the article for publication.

Data availability UKB data are available in a public, open-access repository. This research has been conducted using the UKB Resource under Application Number 44430. The UKB data are available on application to the UK Biobank (http://www.ukbiobank.ac.uk/).

Declarations

Ethics approval The UKB study was approved by the National Information Governance Board for Health and Social Care in England and Wales, the Community Health Index Advisory Group in Scotland, and the North West Multicenter Research Ethics Committee. All participants gave written informed consent. This UKB study was also approved by the Ethical Committee of Peking University (Beijing, China).

Competing interests The authors declare no competing interests.

References

1. Chang AY, Skirbekk VF, Tyrovolas S, Kassebaum NJ, Dieleman JL. Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. Lancet Public Health. 2019;4:e159–67. https://doi.org/10.1016/s2468-2667(19)30019-2.
2. United Nations, Department of Economic and Social Affairs, Population Division (2019) World Population Prospects 2019: highlights. ST/ESA/SER.A/423.
3. Hamczyk MR, Nevado RM, Barettino A, Fuster V, Andrés V. Biological versus chronological aging: JACC Focus Seminar. J Am Coll Cardiol. 2020;75:919–30. https://doi.org/10.1016/j.jacc.2019.11.062.
4. Jylhävä J, Pedersen NL, Hägg S. Biological age predictors. EBioMedicine. 2017;21:29–36. https://doi.org/10.1016/j.ebiom.2017.03.046.
5. Chan MS, Arnold M, Offer A, Hammami I, Mafham M, Armitage J, Perera R, Parish S. A Biomarker-based biological age in UK Biobank: composition and prediction of mortality and hospital admissions. J Gerontol A Biol Sci Med Sci. 2021;76:1295–302. https://doi.org/10.1093/gerona/glab069.
6. Williams DM, Jylhävä J, Pedersen NL, Hägg S. A frailty index for UK Biobank participants. J Gerontol A Biol Sci Med Sci. 2019;74:582–7. https://doi.org/10.1093/gerona/gly094.
7. Newman AB, Boudreau RM, Naydeck BL, Fried LF, Harris TB. A physiologic index of comorbidity: relationship to mortality and disability. J Gerontol A Biol Sci Med Sci. 2008;63:603–9. https://doi.org/10.1093/gerona/63.6.603.
8. Wu C, Smit E, Sanders JL, Newman AB, Odden MC. A modified healthy aging index and its association with mortality: the National Health and Nutrition Examination Survey, 1999–2002. J Gerontol A Biol Sci Med Sci. 2017;72:1437–44. https://doi.org/10.1093/gerona/glw334.
9. Zhang H, Zhu Y, Hao M, Wang J, Wang Z, Chu X, et al. The modified healthy ageing index is associated with mortality and disability: the Rugao Longevity and Ageing Study. Gerontology. 2021;67:572–80. https://doi.org/10.1159/000513931.
10. McCabe EL, Larson MG, Lunetta KL, Newman AB, Cheng S, Murabito JM. Association of an index of healthy aging with incident cardiovascular disease and mortality in a community-based sample of older adults. J Gerontol A Biol Sci Med Sci. 2016;71:1695–701. https://doi.org/10.1093/gerona/glw077.
11. Sanders JL, Minster RL, Barmada MM, Matteini AM, Boudreau RM, Christensen K, et al. Heritability of and mortality prediction with a longevity phenotype: the healthy aging index. J Gerontol A Biol Sci Med Sci. 2014;69:479–85. https://doi.org/10.1093/gerona/glt117.
12. Wu C, Newman AB, Dong BR, Odden MC. Index of healthy aging in Chinese older adults: China Health

and Retirement Longitudinal Study. J Am Geriatr Soc. 2018;66:1303–10. https://doi.org/10.1111/jgs.15390.
13. Sanders JL, Boudreau RM, Penninx BW, Simonsick EM, Kritchevsky SB, Satterfield S, Harris TB, Bauer DC, Newman AB. Association of a modified physiologic index with mortality and incident disability: the Health, Aging, and Body Composition Study. J Gerontol A Biol Sci Med Sci. 2012;67:1439–46. https://doi.org/10.1093/gerona/gls123.
14. Office for National Statistics. Population estimates for UK, England and Wales, Scotland and Northern Ireland: Mid 2020. Newport: Office for National Statistics; 2021.
15. Huang N, Zhuang Z, Song Z, Wang W, Li Y, Zhao Y, et al. Associations of modified healthy aging index with major adverse cardiac events, major coronary events, and ischemic heart disease. J Am Heart Assoc. 2023;12:e026736. https://doi.org/10.1161/jaha.122.026736.
16. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779. https://doi.org/10.1371/journal.pmed.1001779.
17. Allen NE, Sudlow C, Peakman T, Collins R. UK Biobank data: come and get it. Sci Transl Med. 2014;6:224ed224. https://doi.org/10.1126/scitranslmed.3008601.
18. Collins R. What makes UK Biobank special? Lancet. 2012;379:1173–4. https://doi.org/10.1016/s0140-6736(12)60404-8.
19. Fawns-Ritchie C, Deary IJ. Reliability and validity of the UK Biobank cognitive tests. PLoS One. 2020;15:e0231627. https://doi.org/10.1371/journal.pone.0231627.
20. Classification and diagnosis of diabetes. standards of medical care in diabetes-2021. Diabetes Care. 2021;44:S15-s33. https://doi.org/10.2337/dc21-S002.
21. Brunström M, Carlberg B. Association of blood pressure lowering with mortality and cardiovascular disease across blood pressure levels: a systematic review and meta-analysis. JAMA Intern Med. 2018;178:28–36. https://doi.org/10.1001/jamainternmed.2017.6015.
22. Metter EJ, Schrager M, Ferrucci L, Talbot LA. Evaluation of movement speed and reaction time as predictors of all-cause mortality in men. J Gerontol A Biol Sci Med Sci. 2005;60:840–6. https://doi.org/10.1093/gerona/60.7.840.
23. Batterham PJ, Bunce D, Mackinnon AJ, Christensen H. Intra-individual reaction time variability and all-cause mortality over 17 years: a community-based cohort study. Age Ageing. 2014;43:84–90. https://doi.org/10.1093/ageing/aft116.
24. Hart A, Blackwell TL, Paudel ML, Taylor BC, Orwoll ES, Cawthon PM, Ensrud KE. Cystatin C and the risk of frailty and mortality in older men. J Gerontol A Biol Sci Med Sci. 2017;72:965–70. https://doi.org/10.1093/gerona/glw223.
25. Emberson JR, Haynes R, Dasgupta T, Mafham M, Landray MJ, Baigent C, Clarke R. Cystatin C and risk of vascular and nonvascular mortality: a prospective cohort study of older men. J Intern Med. 2010;268:145–54. https://doi.org/10.1111/j.1365-2796.2010.02214.x.
26. Sorkin JD, Muller DC, Fleg JL, Andres R. The relation of fasting and 2-h postchallenge plasma glucose concentrations to mortality: data from the Baltimore Longitudinal Study of Aging with a critical review of the literature. Diabetes Care. 2005;28:2626–32. https://doi.org/10.2337/diacare.28.11.2626.
27. Lee HM, Le H, Lee BT, Lopez VA, Wong ND. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. Eur Respir J. 2010;36:1002–6. https://doi.org/10.1183/09031936.00042410.
28. O'Connell MDL, Marron MM, Boudreau RM, Canney M, Sanders JL, Kenny RA, Kritchevsky SB, Harris TB, Newman AB. Mortality in relation to changes in a healthy aging index: the Health, Aging, and Body Composition Study. J Gerontol A Biol Sci Med Sci. 2019;74:726–32. https://doi.org/10.1093/gerona/gly114.
29. Pazoki R, Dehghan A, Evangelou E, Warren H, Gao H, Caulfield M, Elliott P, Tzoulaki I. Genetic predisposition to high blood pressure and lifestyle factors: associations with midlife blood pressure levels and cardiovascular events. Circulation. 2018;137:653–61. https://doi.org/10.1161/circulationaha.117.030898.
30. Wang M, Zhou T, Li X, Ma H, Liang Z, Fonseca VA, Heianza Y, Qi L. Baseline vitamin D status, sleep patterns, and the risk of incident type 2 diabetes in data from the UK Biobank Study. Diabetes Care. 2020;43:2776–84. https://doi.org/10.2337/dc20-1109.
31. Salazar N, Valdés-Varela L, González S, Gueimonde M, de Los Reyes-Gavilán CG. Nutrition and the gut microbiome in the elderly. Gut Microbes. 2017;8:82–97. https://doi.org/10.1080/19490976.2016.1256525.
32. Clements SJ, Carding SR. Diet, the intestinal microbiota, and immune health in aging. Crit Rev Food Sci Nutr. 2018;58:651–61. https://doi.org/10.1080/10408398.2016.1211086.
33. Kim S, Jazwinski SM. The gut microbiota and healthy aging: a mini-review. Gerontology. 2018;64:513–20. https://doi.org/10.1159/000490615.
34. Carmona JJ, Michan S. Biology of healthy aging and longevity. Rev Invest Clin. 2016;68:7–16.
35. Luy M, Gast K. Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy. Gerontology. 2014;60:143–53. https://doi.org/10.1159/000355310.
36. Rochelle TL, Yeung DK, Bond MH, Li LM. Predictors of the gender gap in life expectancy across 54 nations. Psychol Health Med. 2015;20:129–38. https://doi.org/10.1080/13548506.2014.936884.
37. Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. Nat Rev Cancer. 2016;16:330–9. https://doi.org/10.1038/nrc.2016.30.
38. Barrett EL, Richardson DS. Sex differences in telomeres and lifespan. Aging Cell. 2011;10:913–21. https://doi.org/10.1111/j.1474-9726.2011.00741.x.
39. Santoro A, Ostan R, Candela M, Biagi E, Brigidi P, Capri M, Franceschi C. Gut microbiota changes in the extreme decades of human life: a focus on centenarians. Cell Mol Life Sci. 2018;75:129–48. https://doi.org/10.1007/s00018-017-2674-y.
40. Ostan R, Monti D, Gueresi P, Bussolotto M, Franceschi C, Baggio G. Gender, aging and longevity in humans:

an update of an intriguing/neglected scenario paving the way to a gender-specific medicine. Clin Sci (Lond). 2016;130:1711–25. https://doi.org/10.1042/cs20160004.
41. Yip BW, Mok HO, Peterson DR, Wan MT, Taniguchi Y, Ge W, Au DW. Sex-dependent telomere shortening, telomerase activity and oxidative damage in marine medaka Oryzias melastigma during aging. Mar Pollut Bull. 2017;124:701–9. https://doi.org/10.1016/j.marpolbul.2017.01.021.
42. Sanchez-Sanchez JL, Izquierdo M, Carnicero-Carreño JA, García-García FJ, Rodríguez-Mañas L. Physical activity trajectories, mortality, hospitalization, and disability in the Toledo Study of Healthy Aging. J Cachexia Sarcopenia Muscle. 2020;11:1007–17. https://doi.org/10.1002/jcsm.12566.
43. Stenholm S, Head J, Kivimäki M, Kawachi I, Aalto V, Zins M, et al. Smoking, physical inactivity and obesity as predictors of healthy and disease-free life expectancy between ages 50 and 75: a multicohort study. Int J Epidemiol. 2016;45:1260–70. https://doi.org/10.1093/ije/dyw126.
44. Daskalopoulou C, Koukounari A, Wu YT, Terrera GM, Caballero FF, de la Fuente J, et al. Healthy ageing trajectories and lifestyle behaviour: the Mexican Health and Aging Study. Sci Rep. 2019;9:11041. https://doi.org/10.1038/s41598-019-47238-w.

OXFORD

human reproduction

Human Reproduction, 2023, **00(0)**, 1–9
https://doi.org/10.1093/humrep/dead149
Original Article

Reproductive endocrinology

Uptake of the core outcome set on polycystic ovary syndrome before and after its publication

Wenqiang Li [1,2,†], Guoliang Li [1,3,†], Hongbin Chi [4], Haining Wang [5], and Lin Zeng [1,2,*]

[1]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, People's Republic of China
[2]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, People's Republic of China
[3]School of Basic Medical Sciences, Peking University Health Science Center, Beijing, People's Republic of China
[4]Department of Obstetrics and Gynecology, Center for Reproductive Medicine, Peking University Third Hospital, Beijing, People's Republic of China
[5]Department of Endocrinology and Metabolism, Peking University Third Hospital, Beijing, People's Republic of China

*Correspondence address. Research Center of Clinical Epidemiology, Peking University Third Hospital, 49 Huayuan North Road, Haidian District, Beijing 100191, People's Republic of China. E-mail: zlwhy@163.com https://orcid.org/0000-0001-8707-5854
†These authors consider that the first two authors should be regarded as joint First Authors.

ABSTRACT

STUDY QUESTION: Does the core outcome set (COS) on polycystic ovary syndrome (PCOS) impact the selection of research outcomes?

SUMMARY ANSWER: Following the publication of the COS on PCOS, an increasing number of trials are reporting both the generic domain and body mass index; however, the uptake of this COS has not been as extensive as expected.

WHAT IS KNOWN ALREADY: The COS on PCOS included 33 core outcomes in the following seven domains: the generic (3), metabolic (8), reproductive (7), pregnancy (10), psychological (3), oncological (1), and long-term (1). This was done to improve consistency in outcome selection and definition. However, thus far, no studies have investigated the effectiveness of this COS in the above-mentioned tasks.

STUDY DESIGN, SIZE, DURATION: A methodological study based on the trial registries, including 395 eligible clinical trials registered between 1 January 2018 and 21 September 2022.

PARTICIPANTS/MATERIALS, SETTING, METHODS: A total of 1258 registered clinical studies on PCOS were retrieved from the World Health Organization International Clinical Trials Registry Platform. Of those, 395 were selected according to the inclusion and exclusion criteria, and divided into two groups based on the publication date of the COS on PCOS (4 February 2020): pre-publication and post-publication. The practical uptake of this COS was explored after data collation, assessment, comparison of the uptake of core outcomes or domains before and after the publication of this COS, and correlation analysis between the domains.

MAIN RESULTS AND THE ROLE OF CHANCE: There were 26 out of 33 core outcomes and five out of seven domains reported in the 395 trials. The highest uptake was observed for the reproductive domain and the reproductive hormonal profile (63.0% and 38.7%, respectively). After the publication of the COS on PCOS, the uptake of the generic domain and body mass index increased from 24.1% to 35.8% ($P = 0.011$) and 17.8% to 26.5% ($P = 0.039$), respectively. The total number of reported core outcomes in the generic domain met statistical significance ($P = 0.012$). Moreover, multivariable analyses still supported the above finding in the generic domain. Correlation analysis showed that most of the domains were positively correlated with each other. However, the pregnancy domain was negatively correlated with the metabolic domain. Reasons responsible for the unsatisfactory uptake may be the absence of specific definitions of core outcomes, as well as the lack of awareness among researchers regarding this COS.

LIMITATIONS, REASONS FOR CAUTION: Due to the lack of standardized definition of outcomes, it was difficult to avoid some subjectivity in the process of consistency assessment.

WIDER IMPLICATIONS OF THE FINDINGS: Two years after its publication, there was no substantial improvement in the uptake of the COS on PCOS. This suggests that this COS may require further revision, refinement, and promotion to improve the comparability of PCOS studies.

STUDY FUNDING/COMPETING INTEREST(S): This work was funded by Beijing Municipal Health Science and Technology Achievements and Appropriate Technology Promotion Project (BHTPP2022069), and the special fund of Beijing Key Clinical Specialty Construction Project. The authors do not have conflicts of interest to declare.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: uptake of the core outcome set / polycystic ovary syndrome / assessment of consistency / clinical trials / International Clinical Trials Registry Platform

Received: March 23, 2023. **Revised:** June 19, 2023. **Editorial decision:** July 12, 2023.

Introduction

Minimizing bias when designing clinical trials is necessary for the direct comparison of the effects of different interventions. Therefore, the selection of appropriate outcomes is crucial (Williamson *et al.*, 2012; Smith *et al.*, 2019a). However, reporting outcomes in clinical studies is generally inconsistent, unimportant, or incomplete, thereby resulting in a waste of resources (Thornley and Adams, 1998; Chalmers and Glasziou, 2009; Williamson *et al.*, 2020). A core outcome set (COS) refers to an agreed standard set of outcomes that should be measured and reported, as a minimum, in all clinical trials for specific areas of health or health care (Williamson *et al.*, 2012; Clarke and Williamson, 2016; WHO Working Group on the Clinical Characterisation and Management of COVID-19 Infection, 2020). The main function of COSs was to reduce risk of reporting bias, improve consistency between similar studies, facilitate integration and comparison of data, and simplify the selection of study outcomes (Kirkham *et al.*, 2010; Dwan *et al.*, 2013; Smith *et al.*, 2015; Clarke and Williamson, 2016; Topjian *et al.*, 2020).

Polycystic ovary syndrome (PCOS) is a key disorder of concern and an active topic of scientific research in the reproductive, metabolic, and endocrine disciplines (Escobar-Morreale, 2018). The COS on PCOS comprises 33 core outcomes in the following seven domains: 3 generic, 8 metabolic, 7 reproductive, 10 pregnancy, 3 psychological, 1 oncological, and 1 long-term outcomes. This COS was developed to assist researchers in the selection of outcomes in practice. The developers of this COS expected researchers to tailor the reporting of outcomes according to their research questions. The objective was to cover all relevant outcomes in this COS while justifying the lack of reporting for any remaining outcomes (Al Wattar *et al.*, 2020). The World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) provides information on PCOS clinical trials collected from various registries, with data reflecting the design, planned completion time and outcomes selection of clinical trials. It can help in efficient evaluation of the uptake of COSs and thus has been used in some studies (Kirkham *et al.*, 2017; Smith *et al.*, 2019a,b).

In recent years, an increasing number of researchers have been focusing on COSs. A substantial increase in the development of COSs has been recorded (Williamson *et al.*, 2020). However, thus far, studies on the uptake of COSs are lacking (Hughes *et al.*, 2021). Previous studies showed that despite the rapid development of COSs, the clarity of reporting of core outcomes was suboptimal and numerous COSs still require improvement (Kirkham *et al.*, 2016; Gargon *et al.*, 2019; Goren *et al.*, 2023). Moreover, the low utilization rate causes a waste of resources; this contradicts the original intention of COSs development. Therefore, it is necessary to ensure the representativeness, universality, and practicability of COSs. Based on systematic assessment, this study investigated the uptake of the COS on PCOS research before and after its publication. The assessment can offer developers the opportunity to review their research results, and avoid situations in which a COS was developed but not effectively utilized.

Materials and methods

Research data acquisition

Using the keyword 'polycystic ovary syndrome', clinical studies registered between 18 October 1999 and 28 August 2022 were retrieved from the WHO ICTRP Search Portal.

The inclusion criteria were (i) registered clinical studies involving patients with PCOS, (ii) interventional clinical trials, and (iii) clinical trials registered from 1 January 2018 to 21 September 2022 (date of retrieval). The exclusion criteria were (i) studies which included non-PCOS populations and (ii) studies missing information on key outcomes.

Finally, eligible clinical trial registrations were ultimately included in this analysis. We collected relevant information (trial ID, public title, time of data registration, web links, study type, study design, country or region, primary outcomes, secondary outcomes, etc.). The purpose was to establish a database through Research Electronic Data Capture (REDCap: Vanderbilt University, Nashville, TN, USA), as well as to retrieve and extract the original registration data in registries through the web links to update the information directly provided by the WHO ICTRP.

Assessment for the uptake of the COS on PCOS

The outcomes of these clinical trial registry entries were evaluated in REDCap to determine whether they were recommended by the COS on PCOS (including three options, namely consistent, inconsistent, and derivable outcomes). Derivable outcomes refer to some outcomes in trial registrations that are not exactly consistent with the outcomes of this COS, but can be associated with them through a certain conceptual extension. For example, only reporting 'blood pressure' does not conform to 'hypertension' in this COS; however, it can be included in the derivable outcomes of hypertension according to the above criteria. Independent evaluations by two assessors were conducted to explore the consistency between the reported all outcomes in PCOS clinical trial registrations and this COS. Disagreements were resolved through group discussion; for unresolved disagreements and some undefined outcomes, a consensus was reached through consultation of experts.

According to the publication date of the COS on PCOS (4 February 2020), the eligible trials were divided into two groups, namely the pre-publication (Pre-Pub) group and post-publication (Post-Pub) group. Among the three options of all core outcomes, consistent outcomes were regarded as the reporting of core outcomes, while the others were regarded as not reporting outcomes. As randomization, blinding and prospective registration may affect the uptake of the COS, we also included these three factors in the analysis. To explore the adoption of this COS, the number of clinical trials reporting each core outcome or domain, and the total number of reported core outcomes were calculated. Clinical trials that reported at least one core outcome from a core domain were considered to have adopted the domain. The total number of reported core domains was categorized into four categories (i.e. 0, 1, 2, ≥3); subsequently, the number of trials for each category was calculated. In addition, the pairwise correlation between domains was analyzed to determine the domains that researchers prefer to report simultaneously.

Statistical analysis

The study database was established using REDCap. Statistical analysis was conducted using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Measurement data were presented as the median with upper and lower quartiles. The Mann–Whitney *U* test was used for comparisons between the groups in terms of the total number of reported core outcomes in all domains and each of them. Enumeration data were expressed as numbers and percentages. The chi-squared test was used for comparisons between two groups in terms of the number of trials reporting each core outcome or domain, and the number of trials classified into the four aforementioned categories. Multivariable logistic regression and a generalized linear model were used to explore the influence

of four factors on the uptake of this COS. Spearman correlation was used to explore the pairwise correlation between the domains. Figures were plotted using OriginPro 2022 (OriginLab Corp., Northampton, MA, USA). Two-tailed *P*-values ≤0.05 indicated statistically significant differences.

To evaluate the robustness of our findings, we conducted sensitivity analysis by adjusting the evaluation criteria and the time of trial registration. The former was achieved by selecting the consistent and derivable outcomes as the reporting of core outcomes and the inconsistent outcomes as not reporting outcomes. The latter was achieved through exploring annual variations in the distribution of the total number of reported core outcomes in all domains and each of them. The procedures described above were repeated once.

Results

The screening process for PCOS registered clinical studies is shown in Fig. 1. A total of 1258 PCOS-related studies were retrieved from the WHO ICTRP. Finally, 395 trials were selected and divided into the Pre-Pub and Post-Pub groups (191 and 204 interventional trials, respectively).

The characteristics of the 395 eligible clinical trials are presented in Table 1. Of those, 281 (71.1%) were randomized and 114 (28.9%) were non-randomized controlled trials. The median sample size was 78 (Q_1, Q_3: 52, 128). According to the data, Iran had the largest number of clinical trial registrations (n = 115, 29.1%). Notably, the vast majority of clinical trials were single-center studies (n = 386, 97.7%). The minimum and the maximum age of the study population were 18 years (Q_1, Q_3: 18, 20) and 40 years (Q_1, Q_3: 36, 44), respectively. There were no statistically significant differences between two groups in any of the characteristics.

Overall, 26 of the 33 core outcomes of the COS on PCOS were reported. Table 2 lists the uptake of this COS in included trials. The reporting percentages of core outcomes ranged from 0.3% to 38.7%. Nine outcomes had an uptake percentage >10%, including four reproductive outcomes, three metabolic outcomes, and two generic outcomes. The reproductive hormonal profile exhibited the highest uptake percentage (38.7%). The uptake of 12 outcomes increased following the publication of this COS. BMI was more frequently used in the Post-Pub group compared with the Pre-Pub group (17.8% versus 26.5%, respectively, $P = 0.039$). No significant intergroup differences were noted in terms of the uptake of the remaining core outcomes.

Figure 2A shows that the total number of reported core outcomes in all domains was slightly higher after than before the publication of the COS on PCOS (2 (Q_1, Q_3: 1, 3) versus 2 (Q_1, Q_3: 1, 4), $P = 0.742$). There were significant increases in the generic domain after the publication of this COS ($P = 0.012$), as well as in the

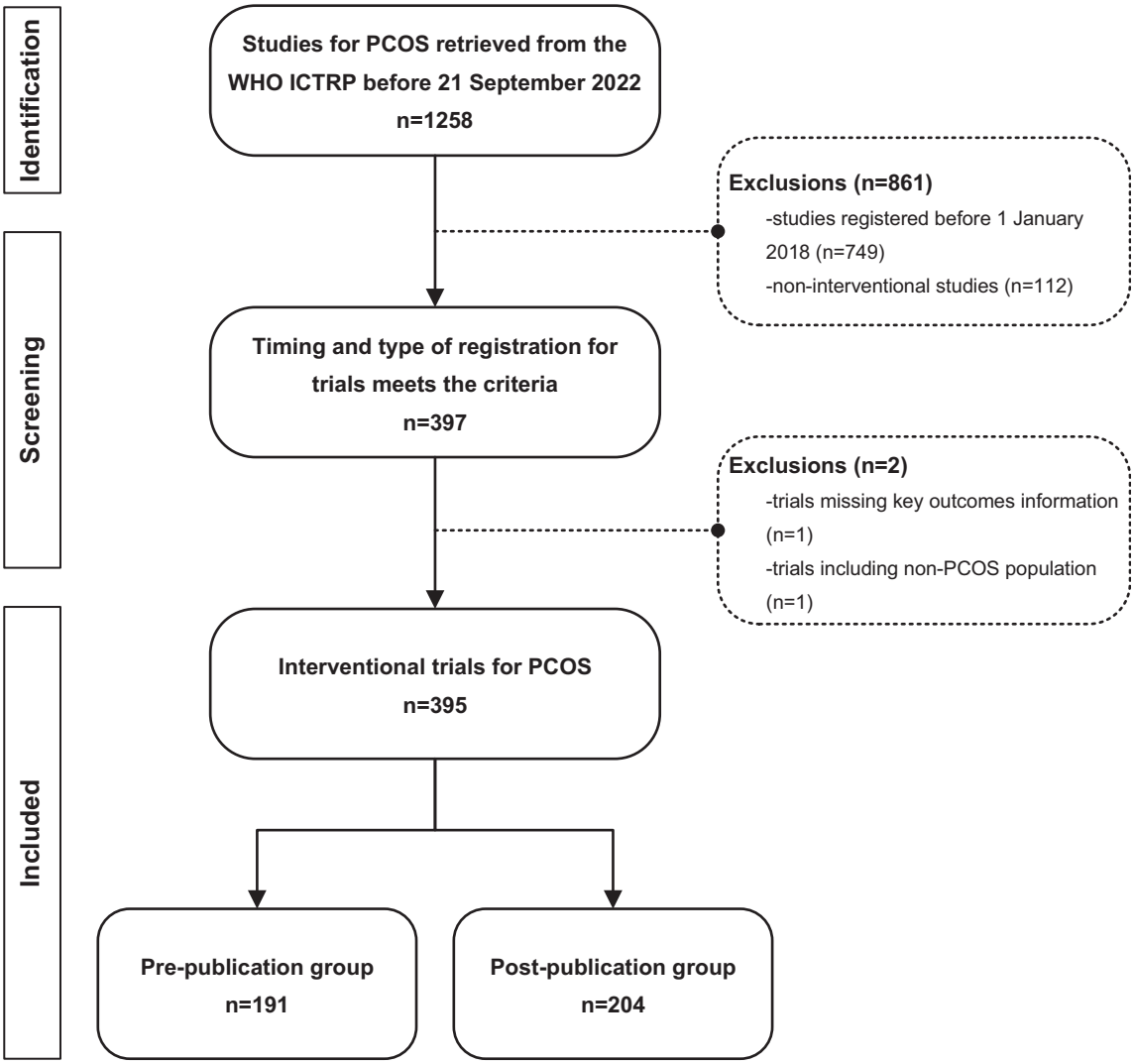


Figure 1. Flowchart of the screening process for PCOS clinical trial registrations retrieved from WHO ICTRP. PCOS, polycystic ovary syndrome; WHO, World Health Organization; ICTRP, International Clinical Trials Registry Platform.

Table 1. Characteristics of the 395 clinical trial registrations on PCOS.

Characteristics	Pre-Pub n = 191	Post-Pub n = 204	Total n = 395	P-value
Study design, n (%)*				
RCT	143 (74.9)	138 (67.6)	281 (71.1)	0.113
Non-RCT	48 (25.1)	66 (32.4)	114 (28.9)	
Median sample size (Q_1, Q_3)†	80 (60, 120)	75 (50, 140)	78 (52, 128)	0.970
Country or region, n (%)*				
Iran	62 (32.5)	53 (26.0)	115 (29.1)	0.069
China	40 (20.9)	57 (27.9)	97 (24.6)	
Egypt	19 (9.9)	16 (7.8)	35 (8.9)	
India	16 (8.4)	19 (9.3)	35 (8.9)	
USA	18 (9.4)	8 (3.9)	26 (6.6)	
Others[a]	36 (18.8)	51 (25.0)	87 (22.0)	
Median age of participants (Q_1, Q_3), years†				
Minimum (n = 323)[b]	18 (18, 20)	18 (18, 19)	18 (18, 20)	0.461
Maximum (n = 320)[c]	40 (38, 44)	40 (35, 44)	40 (36, 44)	0.540

[a] Countries or regions with five or fewer registered trials were classified as others. There were three and six international multicenter studies in the Pre-Pub and Post-Pub groups, respectively.
[b] 30 and 42 trials had missing data in the Pre-Pub and Post-Pub groups, respectively.
[c] 31 and 44 trials had missing data in the Pre-Pub and Post-Pub groups, respectively.
* χ^2 test was used.
† Mann–Whitney *U* test was used.
PCOS, polycystic ovary syndrome; Pre-Pub, pre-publication; Post-Pub, post-publication; Q_1, first quartile; Q_3, third quartile; RCT, randomized controlled trial.

Table 2. Uptake of the COS on PCOS in the clinical trial registrations before versus after publication (outcome level).†

Outcomes	Pre-Pub n = 191 n (%)	Post-Pub n = 204 n (%)	Total n = 395 n (%)	P-value
Generic outcomes				
BMI	34 (17.8)	54 (26.5)	88 (22.3)	0.039
Quality of life	20 (10.5)	29 (14.2)	49 (12.4)	0.259
Treatment satisfaction	1 (0.5)	5 (2.5)	6 (1.5)	0.217*
Metabolic outcomes				
Waist circumference	18 (9.4)	28 (13.7)	46 (11.6)	0.183
Insulin resistance	36 (18.8)	42 (20.6)	78 (19.7)	0.664
Impaired glucose tolerance	17 (8.9)	18 (8.8)	35 (8.9)	0.979
Hypertension	1 (0.5)	0 (0)	1 (0.3)	0.484*
Coronary heart disease	0 (0)	1 (0.5)	1 (0.3)	>0.999*
Lipid profile	50 (26.2)	47 (23.0)	97 (24.6)	0.469
Venous thromboembolic disease	1 (0.5)	0 (0)	1 (0.3)	0.484*
Reproductive outcomes				
Viable pregnancy	24 (12.6)	30 (14.7)	54 (13.7)	0.536
Hyperandrogenism	75 (39.3)	71 (34.8)	146 (37.0)	0.358
Menstrual regularity	38 (19.9)	46 (22.5)	84 (21.3)	0.519
Reproductive hormonal profile	75 (39.3)	78 (38.2)	153 (38.7)	0.833
Ovulation stimulation success and number of stimulated follicles ≥12 mm	4 (2.1)	5 (2.5)	9 (2.3)	>0.999*
Incidence and severity of ovarian hyperstimulation syndrome	7 (3.7)	11 (5.4)	18 (4.6)	0.411
Pregnancy outcomes				
Live birth	14 (7.3)	16 (7.8)	30 (7.6)	0.847
Miscarriage	14 (7.3)	9 (4.4)	23 (5.8)	0.216
Stillbirth	1 (0.5)	0 (0)	1 (0.3)	0.484*
Preterm birth	5 (2.6)	6 (2.9)	11 (2.8)	0.845
Gestational weight gain	1 (0.5)	0 (0)	1 (0.3)	0.484*
Gestational diabetes	4 (2.1)	2 (1.0)	6 (1.5)	0.436*
Hypertensive disease in pregnancy	4 (2.1)	2 (1.0)	6 (1.5)	0.436*
Baby birthweight	5 (2.6)	2 (1.0)	7 (1.8)	0.271*
Psychological outcomes				
Depression	17 (8.9)	18 (8.8)	35 (8.9)	0.979
Anxiety	13 (6.8)	11 (5.4)	24 (6.1)	0.557

† χ^2 test was used.
* Fisher's exact test was used.
PCOS, polycystic ovary syndrome; COS, core outcome set; Pre-Pub, pre-publication; Post-Pub, post-publication; BMI, body mass index.

metabolic domain between the randomization versus non-randomization group (P = 0.016) and between the blinding versus non-blinding group (P = 0.015) (Fig. 2; Supplementary Figs S1, S2, S3, and S4).

Among the included trials, five of the seven domains of the COS on PCOS were reported (Table 3). The reproductive domain was the most commonly reported (63.0%). Fewer trials reported the pregnancy and psychological domains (10.6% and 9.4%).

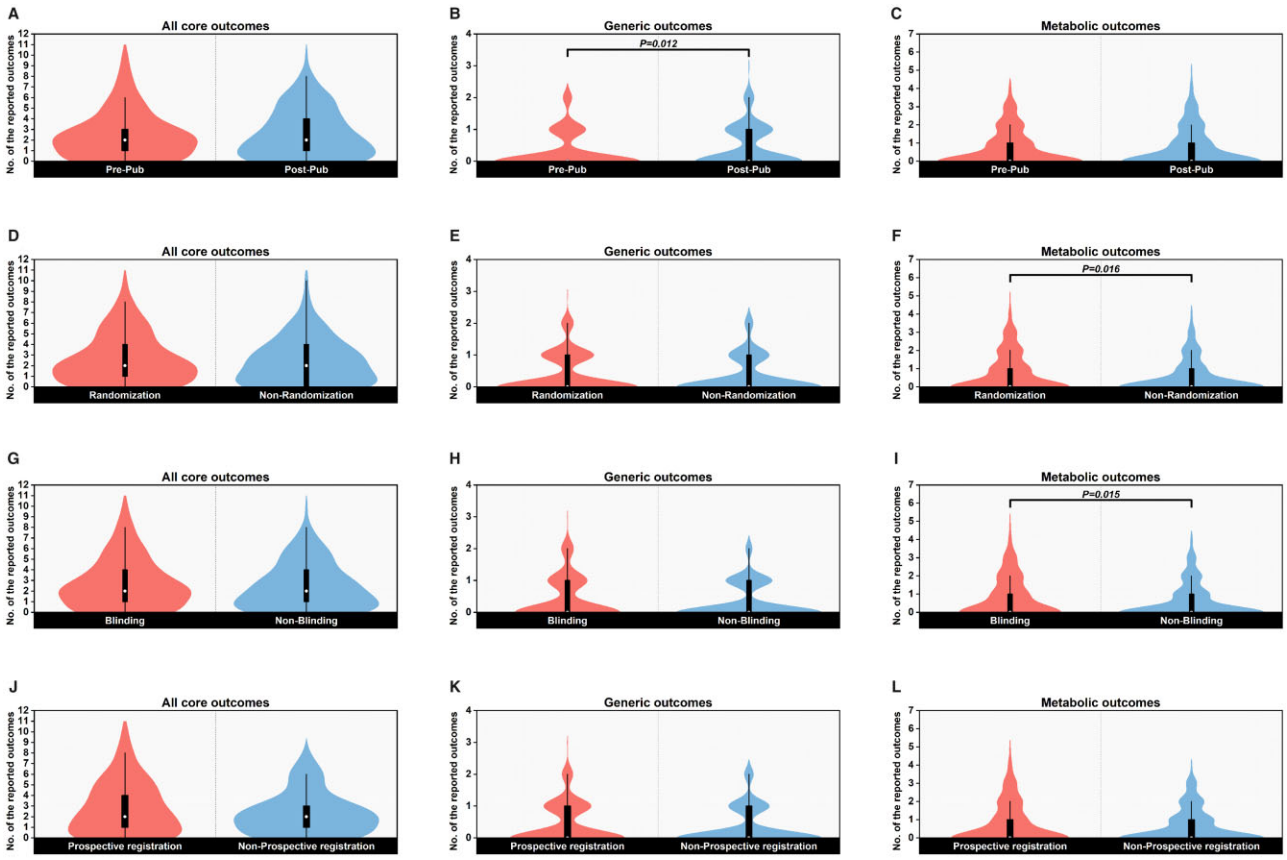


Figure 2. Distribution of the total number of reported core outcomes in all domains, the generic domain and the metabolic domain. The range of whiskers in the boxplot is ±1.5*IQR. There were statistically significant differences in the generic domain between the Post- versus Pre-Pub group ($P = 0.012$), in the metabolic domain between the randomization versus non-randomization group ($P = 0.016$), and in the metabolic domain between the blinding versus non-blinding group ($P = 0.015$). The effect of differences in the number of trials between the groups on the shape of distribution has been eliminated. Pre-Pub, pre-publication; Post-Pub, post-publication; IQR, interquartile range.

The oncological and long-term domains were not reported. The uptake percentage of the generic domain in the Pre-Pub and Post-Pub groups was 24.1% and 35.8%, respectively, reaching statistical significance ($P = 0.011$). Significant increases were observed in the metabolic domain between the randomization versus non-randomization group ($P = 0.017$) and between the blinding versus non-blinding group ($P = 0.023$) (Supplementary Table S1).

Nonetheless, multivariable logistic regression analysis (adjusted odds ratio = 1.78 (95% confidence interval: 1.14–2.78), $P = 0.011$) and the generalized linear model ($\beta = 0.41$, $P = 0.018$) only supported the statistical significance in the generic domain between the Post- versus Pre-Pub group (Table 3; Supplementary Table S2). Moreover, the practice of prospective registration may affect the total number of reported core outcomes in all domains ($\beta = 0.25$, $P < 0.001$).

The number and percentage of trials classified into each category are presented in Fig. 3. Among the four categories, the highest number was recorded for trials reporting only one domain (n = 128, 32.4%). Moreover, 19 (4.8%) trials and 1 (0.3%) trial reported four and five domains, respectively. The percentage of trials reporting three or more domains increased from 17.13% to 22.5%, indicating an improvement in the consistency of outcome reporting. However, there was no statistically significant difference between the two groups ($P = 0.578$).

As shown in Fig. 4, most domains were positively correlated with each other before and after the publication of the COS on PCOS. The generic and metabolic domains showed the strongest correlation in the Pre-Pub and Post-Pub groups ($r = 0.345$, $P < 0.001$ and $r = 0.406$, $P < 0.001$, respectively). This suggested that these domains were more likely to be simultaneously selected by researchers as the outcomes of clinical trials than other domains. However, we also found that the pregnancy domain was negatively correlated with the metabolic domain before and after the publication of this COS ($P = 0.003$ and 0.034, respectively). But these correlations were weak.

Sensitivity analysis

For sensitivity analysis, we altered the grading standard for the uptake of the COS on PCOS. There was no statistically significant evidence that the overall results of this study were substantially changed by the two adjudication standards and the time of trial registration (Supplementary Tables S3 and S4 and Figs S5, S6, S7, and S8).

Discussion

In the present study, the reproductive domain and reproductive hormonal profile (63.0% and 38.7%, respectively) were recorded as having the highest uptake. After the publication of the COS on PCOS, except for the generic domain and its included body mass index ($P = 0.011$ and 0.039, respectively), the number of trials reporting other outcomes or domains did not significantly change. The total number of reported core outcomes in the generic domain met statistical significance ($P = 0.012$). Additionally, multivariable analyses supported the above finding in the generic domain. Studies conducting prospective registration may have

Table 3. Uptake of the COS on PCOS in the clinical trial registrations before versus after publication (domain level) and multivariable logistic regression analysis for trials reporting each domain or not.

Domains	Pre-Pub n = 191 n (%)	Post-Pub n = 204 n (%)	Total n = 395 n (%)	P-value*	aOR (95% CI)#			
					Post- versus Pre-Pub	Randomization	Blinding	Prospective registration
Generic outcomes	46 (24.1)	73 (35.8)	119 (30.1)	0.011	1.78 (1.14–2.78)	1.34 (0.77–2.34)	1.30 (0.80–2.13)	1.51 (0.97–2.35)
Metabolic outcomes	72 (37.7)	75 (36.8)	147 (37.2)	0.848	1.00 (0.66–1.51)	1.55 (0.91–2.64)	1.37 (0.86–2.17)	1.21 (0.79–1.83)
Reproductive outcomes	123 (64.4)	126 (61.8)	249 (63.0)	0.588	0.87 (0.57–1.31)	1.07 (0.64–1.79)	0.81 (0.51–1.29)	1.29 (0.85–1.95)
Pregnancy outcomes	22 (11.5)	20 (9.8)	42 (10.6)	0.581	0.80 (0.42–1.52)	0.75 (0.35–1.62)	0.89 (0.43–1.84)	1.24 (0.64–2.37)
Psychological outcomes	18 (9.4)	19 (9.3)	37 (9.4)	0.970	1.01 (0.51–2.00)	1.17 (0.48–2.84)	1.21 (0.56–2.59)	1.00 (0.51–1.99)

* χ^2 test was used. The oncology and long-term domains were not reported.
\# Multivariable logistic regression model was used by adjusting for factors involving Post- versus Pre-Pub, randomization, blinding, and prospective registration.
PCOS, polycystic ovary syndrome; COS, core outcome set; Pre-Pub, pre-publication; Post-Pub, post-publication; aOR, adjusted odds ratio; CI, confidence interval.

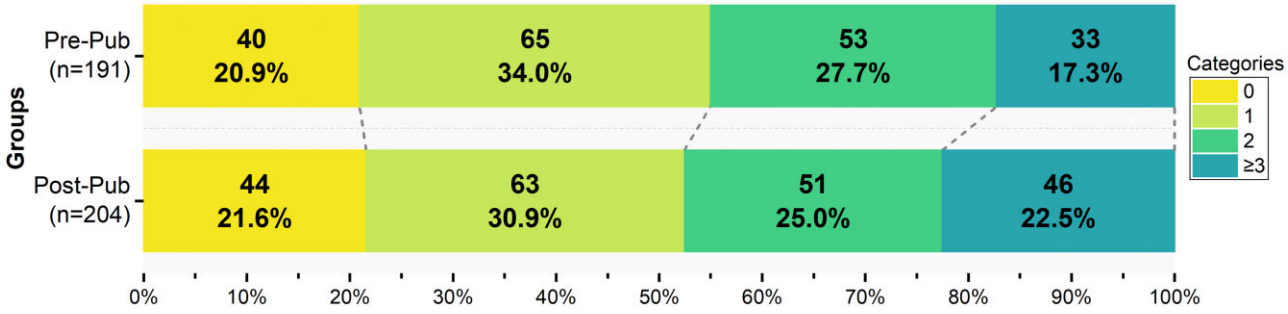


Figure 3. Number of trials for totals of reported core domains before versus after the publication of the COS on PCOS. The categories have been adjusted. Since few trials reported four or more domains, they were merged with the category reporting three domains. There was no statistically significant difference between the two groups ($\chi^2 = 1.974$, $P = 0.578$). PCOS, polycystic ovary syndrome; COS, core outcome set; Pre-Pub, pre-publication; Post-Pub, post-publication.

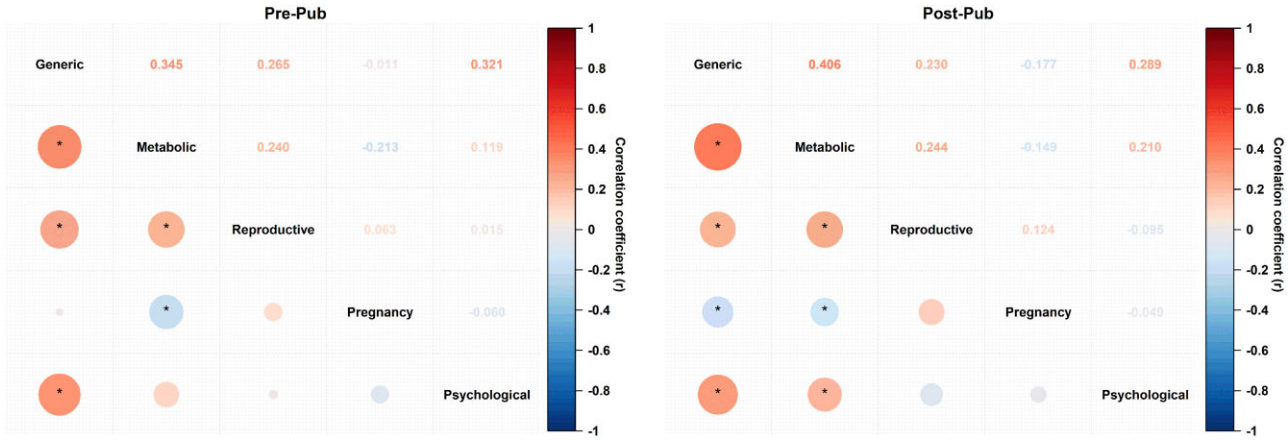


Figure 4. Pairwise correlation between the five domains before versus after the publication of the COS on PCOS. '*' represented that there was a statistical significance in the correlation between the pairwise of five domains by Spearman correlation (P-values ≤0.05). PCOS, polycystic ovary syndrome; COS, core outcome set; Pre-Pub, pre-publication; Post-Pub, post-publication.

more comprehensive research designs that took into account this COS when selecting study outcomes. In addition, the generic and metabolic domains showed a relatively strong correlation. In contrast, the pregnancy and metabolic domains were negatively correlated.

The results of this study showed that the uptake of some core domains and outcomes increased to a certain extent after the publication of the COS on PCOS. A comparison of Table 3 and Supplementary Table S4 shows that the percentage of derivable outcomes reported in the generic and reproductive domains decreased in the Post-Pub group compared with the Pre-Pub group (by 8.9% versus 6.8% and by 13.1% versus 7.8%, respectively). These results indicated that there are fewer ambiguous outcomes in these domains after the publication of this COS, and

that this COS improved consistency in reporting of these outcomes. However, the percentage of total uptake of each outcome or domain was low. In addition, no significant improvement was observed in terms of the consistency. Several similar studies also found low uptake rates for other COSs (Araújo *et al.*, 2015; Smith *et al.*, 2015, 2019a,b; Boric *et al.*, 2019; Farag *et al.*, 2019; Krsticevic *et al.*, 2020). Therefore, developers of COSs should further publicize these domains and outcomes to improve uptake.

The following reasons may be responsible for the unsatisfactory uptake of COSs. Firstly, researchers may have been unaware of the need to seek and use COSs at the stage of study design, or had a limited understanding of COSs. Therefore, more aggressive promotion of COSs may be needed (Kirkham *et al.*, 2013; Krsticevic *et al.*, 2020). Secondly, the COS on PCOS did not include recommendations for effective measurement tools, time frames and guidelines for defining outcomes; these aspects should be further investigated and updated (Palominos *et al.*, 2012; Farag *et al.*, 2019). In addition, the number of domains and outcomes recommended by this COS was high, as PCOS has a long disease cycle and involves a wide range of fields. However, no intervention could affect all domains of the PCOS. A survey showed that the inclusion of six domains in a COS is excessive (Boric *et al.*, 2018). Patients have perceived the completion of these measures of outcomes as a burden, which led to a reluctance to implement them (Mulla *et al.*, 2015). Finally, some studies emphasized the lack of representation of stakeholders in the development of COSs (Smith *et al.*, 2019a; Hughes *et al.*, 2021). In short, longer observation may be necessary to identify the effects of this COS (Al Wattar *et al.*, 2021). In this study, the main difficulty encountered during the assessment of consistency was the lack of clear definition of outcomes in this COS and clinical trial registries, which complicated the matching of the outcomes. For example, it was difficult to determine whether insulin resistance was consistent with fasting blood glucose or fasting insulin in trial registries. Therefore, it is preferable to have a detailed definition of outcomes in accordance with standards recognized by academia.

In this study, psychological outcomes were rarely reported, and oncological and long-term outcomes were not reported (Dokras *et al.*, 2018; Al Wattar *et al.*, 2021; Kiconco *et al.*, 2022a). The following reasons may have hindered researchers from selecting these outcomes and conducting the evidence synthesis. The standardized definition of eating disorders remained inconclusive; there is a paucity of scales specifically designed to assess depression, anxiety and eating disorders in PCOS populations (Al Wattar *et al.*, 2021), and longitudinal studies are confronted with challenges in conducting long-term follow-up (Gibson-Helm *et al.*, 2017; Kiconco *et al.*, 2022a). Furthermore, taking into consideration the aforementioned outcomes, it was proved to result in a significant increase in healthcare-related economic burdens (Riestenberg *et al.*, 2022). Hence these core outcomes remained highly significant and warranted attention. In addition to the international efforts made toward common PCOS features, Azziz *et al.* has continued to investigate long-term complications of PCOS, including psychiatric and neurological disorders and malignancy, and particularly explored the mechanisms by which the intrauterine environment and genetics impacted long-term offspring outcomes (Goodarzi *et al.*, 2011; Azziz *et al.*, 2016, 2019; Azziz, 2018; Joham *et al.*, 2022). The academic community should continuously recommend the rarely evaluated outcomes (Teede *et al.*, 2011; Dokras *et al.*, 2018), facilitate the development of long-term prospective longitudinal studies (Kiconco *et al.*, 2022a,b), and attract the attention of experts from diverse fields, including but not limited to gynecologists and endocrinologists. Developing the targeted outcome assessment scales and standardized outcome definitions may have been a significant promoting effect.

The correlation analysis showed that most of the domains were positively correlated with each other before and after the publication of the COS on PCOS. For example, researchers who reported the metabolic domain were more likely to focus on the generic domain. The pregnancy domain was negatively correlated with the metabolic domain, implying that researchers who paid attention to the pregnancy domain were less likely to focus on the metabolic domain. Furthermore, a certain competitive relationship may exist between these two domains. During pregnancy, some outcomes in the metabolic domain (e.g. hypertension and diabetes) were included in the pregnancy domain, which ultimately led to a negative correlation. However, PCOS is complex and shows discrepant subtypes and clinical manifestations (Teede *et al.*, 2018; Palomba *et al.*, 2021; Dapas and Dunaif, 2022). Therefore, concurrently focusing on the pregnancy domain and reporting the outcomes in metabolic domain may help in providing more comprehensive evidence or clues for the in-depth understanding of the mechanism of PCOS and more accurate treatment for PCOS.

The strength of this study is that we were able to obtain representative registration data from various clinical trial registries through the WHO ICRTP. A previous study showed that the uptake rate obtained for outcomes in trial registries can reasonably reflect the practical uptake of COSs in research publications (Kirkham *et al.*, 2017). Moreover, the method based on registration information was also preferable to citation analysis (Barnes *et al.*, 2017); it shortens the period required for conducting and publishing clinical trials, and offers first-hand information regarding the uptake of COSs.

Limitations

Firstly, some results in this study may have been influenced by the subjective judgment of the evaluators. However, the opinions of clinical experts in related fields were used to assess the consistency of outcomes, and independent assessments were used to enhance the objectivity of the assessment. In case, it was difficult to assess consistency with the COS on PCOS, the outcomes were categorized as derivable outcomes, and a sensitivity analysis was conducted. Secondly, this study evaluated registration information only years after the publication of this COS. This period may be insufficient to estimate the impact of this COS on researchers. Nonetheless, improvements in the uptake of certain outcomes and domains were observed, indicating that this COS plays a role in the consistency of outcome reporting.

Conclusion

Based on systematic assessment, the results of this study showed an increasing trend in the consistency between outcomes in PCOS trial registry entries and the COS on PCOS. However, the overall uptake of this COS was not ideal. The pertinent domains of PCOS deserve a more in-depth exploration by researchers. Moreover, developers of the COS on PCOS should clarify the specific definition of core outcomes, actively update this COS and promote its uptake in clinical trials. These measures may improve the comparability of research outcomes on PCOS to provide more meaningful clinical evidence for PCOS in the future.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article are available from the corresponding author.

Acknowledgements

The authors acknowledge and thank the developers of the core outcome set on polycystic ovary syndrome for their efforts in PCOS research. We are also grateful to the WHO International Clinical Trials Registry Platform for pooling clinical trials registration information to provide representative data for this study.

Authors' roles

L.Z. conceived the original idea and study design. W.Q.L. and G.L.L. participated in the assessment of consistency, analysis and interpretation of data, manuscript drafting, and critical discussion and revision for important intellectual content. L.Z., H.B.C., and H.N.W. contributed to the resolving the disagreements, revising the manuscript, and critically discussing and approving the final version of the paper.

Funding

This work was funded by Beijing Municipal Health Science and Technology Achievements and Appropriate Technology Promotion Project (BHTPP2022069), and the special fund of Beijing Key Clinical Specialty Construction Project.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Al Wattar BH, Bueno A, Martin MG, Ibáñez NC, Harasani K, Garad R, Franks S, Balen A, Bhide P, Piltonen T *et al.* Harmonizing research outcomes for polycystic ovary syndrome (HARP), a marathon not a sprint: current challenges and future research need. *Hum Reprod* 2021;**36**:523–528.

Al Wattar BH, Teede H, Garad R, Franks S, Balen A, Bhide P, Piltonen T, Romualdi D, Laven J, Thondan M *et al.* Harmonising research outcomes for polycystic ovary syndrome: an international multistakeholder core outcome set. *Hum Reprod* 2020;**35**:404–412.

Araújo F, Cordeiro I, Ramiro S, Falzon L, Branco JC, Buchbinder R. Outcomes assessed in trials of gout and accordance with OMERACT-proposed domains: a systematic literature review. *Rheumatology (Oxford)* 2015;**54**:981–993.

Azziz R. Polycystic ovary syndrome. *Obstet Gynecol* 2018;**132**:321–336.

Azziz R, Carmina E, Chen Z, Dunaif A, Laven JSE, Legro RS, Lizneva D, Natterson-Horowtiz B, Teede HJ, Yildiz BO. Polycystic ovary syndrome. *Nat Rev Dis Primers* 2016;**2**:16057.

Azziz R, Kintziger K, Li R, Laven J, Morin-Papunen L, Merkin SS, Teede H, Yildiz BO. Recommendations for epidemiologic and phenotypic research in polycystic ovary syndrome: an androgen excess and PCOS society resource. *Hum Reprod* 2019;**34**:2254–2265.

Barnes KL, Kirkham JJ, Clarke M, Williamson PR. Citation analysis did not provide a reliable assessment of core outcome set uptake. *J Clin Epidemiol* 2017;**86**:153–159.

Boric K, Boric M, Dosenovic S, Jelicic Kadic A, Batinic M, Cavar M, Jeric M, Puljak L. Authors' lack of awareness and use of core outcome set on postoperative pain in children is hindering comparative effectiveness research. *J Comp Eff Res* 2018;**7**:463–470.

Boric K, Jelicic Kadic A, Boric M, Zarandi-Nowroozi M, Jakus D, Cavar M, Dosenovic S, Jeric M, Batinic M, Vukovic I *et al.* Outcome domains and pain outcome measures in randomized controlled trials of interventions for postoperative pain in children and adolescents. *Eur J Pain* 2019;**23**:389–396.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;**374**:86–89.

Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev* 2016;**5**:11.

Dapas M, Dunaif A. Deconstructing a syndrome: genomic insights into PCOS causal mechanisms and classification. *Endocr Rev* 2022;**43**:927–965.

Dokras A, Stener-Victorin E, Yildiz BO, Li R, Ottey S, Shah D, Epperson N, Teede H. Androgen Excess- Polycystic Ovary Syndrome Society: position statement on depression, anxiety, quality of life, and eating disorders in polycystic ovary syndrome. *Fertil Steril* 2018;**109**:888–899.

Dwan K, Gamble C, Williamson PR, Kirkham JJ; Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PLoS One* 2013;**8**:e66844.

Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinol* 2018;**14**:270–284.

Farag AM, Albuquerque R, Ariyawardana A, Chmieliauskaite M, Forssell H, Nasri-Heir C, Klasser GD, Sardella A, Mignogna MD, Ingram M *et al.* World Workshop in Oral Medicine VII: reporting of IMMPACT-recommended outcome domains in randomized controlled trials of burning mouth syndrome: a systematic review. *Oral Dis* 2019;**25**(Suppl 1):122–140.

Gargon E, Williamson PR, Blazeby JM, Kirkham JJ. Improvement was needed in the standards of development for cancer core outcome sets. *J Clin Epidemiol* 2019;**112**:36–44.

Gibson-Helm M, Teede H, Dunaif A, Dokras A. Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2017;**102**:604–612.

Goodarzi MO, Dumesic DA, Chazenbalk G, Azziz R. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol* 2011;**7**:219–231.

Goren K, Monsour A, Stallwood E, Offringa M, Butcher NJ. Pediatric core outcome sets had deficiencies and lacked child and family input: a methodological review. *J Clin Epidemiol* 2023;**155**:13–21.

Hughes KL, Clarke M, Williamson PR. A systematic review finds core outcome set uptake varies widely across different areas of health. *J Clin Epidemiol* 2021;**129**:114–123.

Joham AE, Norman RJ, Stener-Victorin E, Legro RS, Franks S, Moran LJ, Boyle J, Teede HJ. Polycystic ovary syndrome. *Lancet Diabetes Endocrinol* 2022;**10**:668–680.

Kiconco S, Tay CT, Rassie KL, Azziz R, Teede HJ, Joham AE. Where are we in understanding the natural history of polycystic ovary syndrome? A systematic review of longitudinal cohort studies. *Hum Reprod* 2022a;**37**:1255–1273.

Kiconco S, Tay CT, Rassie KL, Azziz R, Teede HJ, Joham AE. Natural history of polycystic ovary syndrome: a systematic review of cardiometabolic outcomes from longitudinal cohort studies. *Clin Endocrinol (Oxf)* 2022b;**96**:475–498.

Kirkham JJ, Boers M, Tugwell P, Clarke M, Williamson PR. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials* 2013;**14**:324.

Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using

ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. *BMJ* 2017;**357**:j2262.

Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;**340**:c365.

Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Moher D, Schmitt J, Tugwell P *et al.* Core outcome set—STAndards for reporting: the COS-STAR statement. *PLoS Med* 2016;**13**:e1002148.

Krsticevic M, Dosenovic S, Dimcea DA-M, Jedrzejewska D, Marques Lameirão AC, Almeida ES, Jelicic Kadic A, Jeric Kegalj M, Boric K, Puljak L. Outcome domains, outcome measures, and characteristics of randomized controlled trials testing nonsurgical interventions for osteoarthritis. *J Rheumatol* 2020;**47**:126–131.

Mulla SM, Maqbool A, Sivananthan L, Lopes LC, Schandelmaier S, Kamaleldin M, Hsu S, Riva JJ, Vandvik PO, Tsoi L *et al.* Reporting of IMMPACT-recommended core outcome domains among trials assessing opioids for chronic non-cancer pain. *Pain* 2015;**156**:1615–1619.

Palomba S, Piltonen TT, Giudice LC. Endometrial function in women with polycystic ovary syndrome: a comprehensive review. *Hum Reprod Update* 2021;**27**:584–618.

Palominos PE, Gaujoux-Viala C, Fautrel B, Dougados M, Gossec L. Clinical outcomes in psoriatic arthritis: a systematic literature review. *Arthritis Care Res (Hoboken)* 2012;**64**:397–406.

Riestenberg C, Jagasia A, Markovic D, Buyalos RP, Azziz R. Health care-related economic burden of polycystic ovary syndrome in the United States: pregnancy-related and long-term health consequences. *J Clin Endocrinol Metab* 2022;**107**:575–585.

Smith TO, Collier T, Sheehan KJ, Sherrington C. The uptake of the hip fracture core outcome set: analysis of 20 years of hip fracture trials. *Age Ageing* 2019a;**48**:595–598.

Smith TO, Mansfield M, Hawker GA, Hunter DJ, March LM, Boers M, Shea BJ, Christensen R, Guillemin F, Terwee CB *et al.* Uptake of the OMERACT-OARSI hip and knee osteoarthritis core outcome set: review of randomized controlled trials from 1997 to 2017. *J Rheumatol* 2019b;**46**:976–980.

Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. *J Clin Epidemiol* 2015;**68**:237–245.

Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;**33**:1602–1618.

Teede HJ, Misso ML, Deeks AA, Moran LJ, Stuckey BGA, Wong JLA, Norman RJ, Costello MF; Guideline Development Groups. Assessment and management of polycystic ovary syndrome: summary of an evidence-based guideline. *Med J Aust* 2011;**195**:S65–S112.

Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;**317**:1181–1184.

Topjian AA, Scholefield BR, Pinto NP, Fink EL, Buysse CMP, Haywood K, Maconochie I, Nadkarni VM, de Caen A, Escalante-Kanashiro R *et al.* P-COSCA (pediatric core outcome set for cardiac arrest) in children: an advisory statement from the international liaison committee on resuscitation. *Circulation* 2020;**142**:e246–e261.

WHO Working Group on the Clinical Characterisation and Management of COVID-19 Infection. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis* 2020;**20**:e192–e197.

Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;**13**:132.

Williamson PR, Ávila Oliveira R, de Clarke M, Gorst SL, Hughes K, Kirkham JJ, Li T, Saldanha IJ, Schmitt J. Assessing the relevance and uptake of core outcome sets (an agreed minimum collection of outcomes to measure in research studies) in Cochrane systematic reviews: a review. *BMJ Open* 2020;**10**:e036562.

Science of the Total Environment 902 (2023) 166088

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The short-term effect of ozone on pregnancy loss modified by temperature: Findings from a nationwide epidemiological study in the contiguous United States

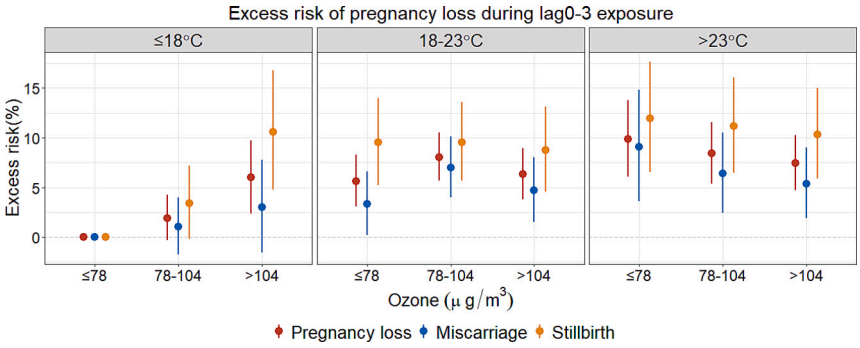
Mingkun Tong [a], Meng Wang [b,c,d], Pengfei Li [a,e,f], Jicheng Gong [g,h], Tong Zhu [g,h], Tao Xue [a,e,h,*]

[a] Institute of Reproductive and Child Health, National Health Commission Key Laboratory of Reproductive Health and Department of Epidemiology and Biostatistics, Ministry of Education Key Laboratory of Epidemiology of Major Diseases (PKU), School of Public Health, Peking University Health Science Center, Beijing, China
[b] Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY 14214, United States
[c] Research and Education in Energy, Environment and Water Institute, University at Buffalo, Buffalo, NY 14214, United States
[d] Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98115, United States
[e] Advanced Institute of Information Technology, Peking University, Hangzhou, China
[f] National Institute of Health Data Science, Peking University, Beijing, China
[g] College of Environmental Sciences and Engineering, Peking University, Beijing, China
[h] State Environmental Protection Key Laboratory of Atmospheric Exposure and Health Risk Management and Center for Environment and Health, Peking University, Beijing, China

HIGHLIGHTS

- Short-term exposure to ozone or high temperature increases the risk for pregnancy loss.
- Temperature antagonizes the adverse effect of ozone on pregnancy loss.
- The risk for ozone-related pregnancy loss can be considerable in high-latitude and high-altitude regions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Lidia Minguez Alarcon

Keywords:
Pregnancy loss
Ozone
Temperature
Joint exposure
Effect modification

ABSTRACT

Background: Pregnancy loss, a major health issue that affects human sustainability, has been linked to short-term exposure to ground-surface ozone (O_3). However, the association is inconsistent, possibly because of the co-occurrence of O_3 and heat episodes, as increased temperature is a risk factor for pregnancy loss. To explain this inconsistency, the effect of O_3 on pregnancy loss needs to be examined jointly with that of high temperature. *Methods:* A total of 247,305 pregnancy losses during the warm season were extracted from fetal death certificates from the 386 counties in contiguous United States from 1989 to 2005. We assessed environmental exposure based on the daily maximum 8 h average of O_3 from Air Quality System monitors and the 24 h average temperature from the North American Regional Reanalysis product. We conducted a bidirectional, time-stratified

* Corresponding author at: School of Public Health, Peking University Health Science Center, Beijing, China.
E-mail address: txue@hsc.pku.edu.cn (T. Xue).

https://doi.org/10.1016/j.scitotenv.2023.166088
Received 15 May 2023; Received in revised form 18 July 2023; Accepted 4 August 2023
Available online 6 August 2023

case-crossover study of the association between pregnancy loss and exposures to O_3 and temperature and their multiplicative interaction. The main time window for the exposure assessment was the day of case occurrence and the preceding 3 days. To estimate the association, we used conditional logistic regression with adjustment for relative humidity, height of the planetary boundary layer, and holidays. Sensitivity analyses were performed on the lagged structure, nonlinearity, and between-subpopulation heterogeneity of the estimated joint effect.
Results: The joint effect was first estimated by the regression against categorical exposure by tertile. Compared to the low-low exposure group ($O_3 \leq 78$ μg/m^3 and temperature ≤ 18 °C), the odds of pregnancy loss was significantly higher by 6.0 % (95 % confidence interval [CI] 2.4–9.7 %), 9.8 % (6.1–13.8 %), and 7.5 % (4.7–10.3 %) in the high-low (>104 μg/m^3 and ≤18 °C), low-high (≤78 μg/m^3 and >23 °C), and high-high (>104 μg/m^3 and >23 °C) groups. The model of linear exposure and the multiplicative interaction yielded similar results. Each increment of 10 μg/m^3 in O_3 and 1 °C in temperature was associated with a 3.0 % (2.0 %–4.0 %) and 3.9 % (3.5 %–4.3 %), respectively, increase in the odds of pregnancy loss. A decrease in odds of 0.2 % (0.1 %–0.2 %) was associated with the temperature × O_3 interaction. The finding of an antagonistic interaction between temperature and O_3 was confirmed by models parametrizing the joint exposure as alternative nonlinear terms (i.e., a two-dimensional spline term or a varying-coefficient term) and was robust to a variety of exposure lags and stratifications. Therefore, the marginal effect of O_3 was estimated to vary by climate zone. A significant association between O_3 and pregnancy loss was observed in the northern, but not southern, United States.
Conclusion: Joint exposure to O_3 and high temperature can increase the risk for pregnancy loss. The adverse effect of O_3 is potentially modified by ambient temperature. In high-latitude cities, controlling for O_3 pollution could protect maternal health.

1. Introduction

Pregnancy loss (also known as fetal death), including miscarriage and stillbirth, harms maternal health and adversely affects human sustainability. Previous epidemiology evidences suggested the adverse effect of air pollution on pregnancy loss, with the pronounced and consistent findings on gestational exposure of fine particulate matters globally (Xie et al., 2021; Zhang et al., 2021). Over the past decade, global levels of ozone exposure have been slowly but steadily increasing (Health Effects Institute, 2020). In the United States (US) at 2016, among all air pollution events, the majority was caused by ozone (O_3) exceeding the national ambient air quality standard (Zhang et al., 2019). A limited number of studies have linked ground-surface O_3 exposure during pregnancy or a specific trimester to the risk for pregnancy loss (Rammah et al., 2019). Ground-surface O_3 concentrations fluctuate seasonally and diurnally and typically peak on warm days. Therefore, the long-term average might over-smooth peak values of O_3 exposure, thereby underestimating the effects of episodes of high O_3 pollution. However, few studies have evaluated the acute effect of short-term O_3 exposure on pregnancy loss, and their results vary. For instance, a meta-analysis of four independent studies on the association between stillbirth and short-term O_3 exposure reported an excess risk (ER) of 0.2 % (95 % confidence interval [CI] 0.1–0.4 %) per 10 μg/m^3 (Zhang et al., 2021). The magnitude of the estimated association varied considerably among studies. The point estimate of the ER ranged from 0 % (−10.2 %, 10.5 %) in Harris County, Texas, to 3.0 % (−0.3 %, 6.5 %) in the Consortium on Safe Labor database of 19 hospitals across the United States. The discrepancies in the evidence for the acute effect of O_3 on pregnancy loss may be due to differences in study populations, epidemiological designs, statistical models, or adjusted covariates. Therefore, a large nationwide epidemiological study is warranted.

The effects of heat waves or increased temperature on pregnancy loss have been evaluated. Chersich et al. pooled eight independent studies and reported that the risk for stillbirth increased by 5 % (1 %, 8 %) for each 1 °C increase in temperature (Chersich et al., 2020). During extreme heat events, ambient temperature and intense sunlight can increase the ambient O_3. The co-occurrence of a heat wave and photochemical air pollution leads to joint exposure to O_3 and high temperature, both of which contribute to pregnancy loss. Given the complexities of the relationship between O_3 and temperature and its effect on health, it is critical to explore the joint effect of co-exposure. High temperature can confound the effect of O_3 exposure. For instance, a time-stratified case-crossover study in California reported a significant unadjusted association between O_3 and stillbirth, but adjusting for apparent temperature attenuated the association, leading to a loss of significance (Sarovar et al., 2020). Moreover, the effects of O_3 and temperature may not be additive, resulting in a synergic or antagonistic interaction. Given the increase in O_3 levels driven by global warming (Ban et al., 2022), studying this interaction would enable accurate assessment of the impact of climate change on pregnancy loss. Some studies have reported a synergic interaction between O_3 and high temperature (Filleul et al., 2006; Ren et al., 2009; Shi et al., 2020), whereas others have shown an antagonistic interaction (Ren et al., 2008; Ren et al., 2009) or no interaction (Jhun et al., 2014; Schwarz et al., 2021). However, the interactive effect of O_3 and temperature on pregnancy loss has not been investigated. In addition, analyzing this interaction is equivalent to examining how the effect of O_3 is modified by temperature or climate zone, a variable determined by multiple meteorological conditions, including temperature. Therefore, the geographic variation in associations between O_3 and stillbirth might be a result of the O_3 × temperature interaction.

To evaluate the association between pregnancy loss and short-term O_3 exposure, we conducted a nationwide case-crossover study. We collected all available records of pregnancy loss among singleton deliveries in the contiguous United States from 1989 to 2005. To explore geographic variation in the association, we focused on the joint effect of O_3 and high temperature, and so we investigated pregnancy loss only during the warm season.

2. Methods

2.1. Study population

Data for the contiguous United States were obtained from the birth certificate system of the National Center for Health Statistics, a publicly available data set that contains demographic information on fetuses, mothers, and fathers and obstetric information. In the birth certificate system, pregnancy loss (or fetal death) is defined as death prior to the complete expulsion or extraction from the mother of a product of human conception. Records included all pregnancy losses, irrespective of the duration of the pregnancy (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm), but excluded pregnancies ended by induced termination. For reasons of confidentiality, geographic codes have not been made publicly available since 2005. Records with last menstrual period dates between January 1, 1989, and December 31, 2004 were included in the analyses (Tong et al., 2022). We included only pregnancy losses that occurred during the warm season. We excluded multiple pregnancies, cases that lacked a date for the pregnancy loss, records with missing

geographic codes, and records that could not be matched with a valid exposure variable. Ultimately, 247,305 pregnancy losses from 386 counties across the contiguous United States were analyzed. The outcome of interest was pregnancy loss, including miscarriage (fetal death before the 20th gestational week) and stillbirth (fetal death at or after the 20th gestational week).

2.2. Exposure assessment

Daily maximum 8 h average (MDA8) concentrations of O_3 from all monitors in the contiguous United States from 1989 to 2005 were obtained from the US Environmental Protection Agency (EPA) Air Quality System (AQS) database (https://www.epa.gov/outdoor-air-quality-data). The data were processed according to uniform EPA guidelines, including those on quality assurance and quality control. In total, 1992 sites reported O_3 measurements from 1989 to 2005, albeit with different spatiotemporal coverages. To address this, we interpolated missing values in AQS monitoring observations using the ordinary kriging method by day. A complete daily time series of O_3 was prepared for each station, and the station-level data were aggregated into county-level time series for the 387 counties.

Daily mean temperature from 1989 to 2005 was obtained from the North American Regional Reanalysis (NARR) product (https://psl.noaa.gov/data/gridded/data.narr.html), which provides meteorological variables in a grid of 32 × 32 km cells across North America. County-specific daily mean temperatures were calculated as the weighted averages of values from all overlaid grid cells. We focused on exposure during the warm season, as defined by the county-specific temperature time series. For each county and year, we calculated the maximum 6-month (i.e., 182 days) moving average temperature and used the corresponding duration as the warm season. To control for potential confounders, we extracted relative humidity (RH) and the height of the planetary boundary layer (PBL) from the NARR product and processed them as described previously. Our study utilized PBL as an indicator of short-term variations in $PM_{2.5}$ exposure, which is majorly determined by meteorological conditions and their interactions with air pollution emissions. The PBL is a critical parameter that affects near-surface air quality, and has been utilized as an instrumental variable to estimate the effect of $PM_{2.5}$ on many other outcomes, such as mortality (Schwartz et al., 2017; Schwartz et al., 2018).

2.3. Statistical analyses

We designed a case-crossover study of the association between pregnancy loss and co-exposure to O_3 and temperature. For each pregnancy loss, the hazard day was defined as the date of death. We used a bidirectional method to select control days, which were defined as 7 days before and after the hazard day. The bidirectional design eliminated potential confounding effects of the day of the week, seasonality, and long-term trends and balanced baseline risks, which vary with gestational age. The bidirectional, time-stratified design allowed each case of pregnancy loss to serve as its own control and enabled us to draw conclusions based on within-individual comparisons of exposure and outcome. Therefore, all known and unknown confounders, which were unlikely to vary in the short term, were controlled for by the study design.

Conditional logistic regression was used to evaluate the association. In the main model, to estimate the joint effect, we first divided the continuous variable of O_3 or temperature into three groups according to the tertiles (i.e., low, moderate, and high exposure levels). The first, second, and third tertile groups of O_3 and temperature were ≤78 μg/m^3, 78–104 μg/m^3, >104 μg/m^3, and ≤18 °C, 18–23 °C, >23 °C, respectively. Then, we created a nine-level categorical variable by combining the groups. The low-O_3 and low-temperature group (i.e., the low-low group) was defined as the reference group. The effects of main interest were based on the difference between the high-exposure groups (i.e., the high-low, low-high, and high-high groups) and the reference group. The logistic regression against categorical co-exposure was parametrized as.

$$\mathrm{Logit}\left(y_{i,j}\right) = \boldsymbol{x}_{i,j}\boldsymbol{\beta} + f\left(\mathrm{RH}_{i,j}\right) + b_1\mathrm{PBL}_{i,j} + b_2\mathrm{H}_{i,j} + \theta(i), \quad (1)$$

where i and j denote the indices for subject and time, respectively; y is a binary outcome variable (1 = hazard; 0 = control); $\boldsymbol{x}$ denotes an eight-dimensional vector for categorical co-exposure; $\theta(i)$ is stratification by subject index; and $\boldsymbol{\beta}$ and b are regression coefficients. The model was adjusted for several potential time-varying confounders: (1) PBL as a surrogate for ambient particulate matter, (2) a binary variable for holiday (H), and (3) a nonlinear spline term for RH ($df = 3$). The association was evaluated by ER for the comparison of an exposure group to the reference group, such as $\mathrm{ER}_{\text{high-high}} = \exp.(\beta_{\text{high-high}})-1$. The interaction was further evaluated by the relative excess risk due to interaction (RERI) (Richardson and Kaufman, 2009), which was specified as follows:

$$\mathrm{RERI} = \mathrm{ER}_{\text{high-high}} - \mathrm{ER}_{\text{low-high}} - \mathrm{ER}_{\text{high-low}} \quad (2)$$

In sensitivity analyses of the main model, we first explored how a lag in exposure affected the estimates. We estimated the effects of co-exposure at lags of 0, 1, 2, or 3 days in separate models. For representativeness, we used the average O_3 and temperature values for lags of 0–3 days as co-exposure in the main model. Next, we examined between-individual heterogeneity in the estimated joint effect by stratifying the data set by each of 13 individual-level variables. Those variables were maternal age (≤20, 20–30, 30–40, or >40 years old), maternal ethnicity (white or nonwhite), education (≤12 years or > 12 years), prenatal care (yes or no), born in hospital (yes or no), history of pregnancy termination (none, 1, or >1), first pregnancy, parity (nulliparous or multiparous), tobacco use (yes or no), alcohol use (yes or no), chronic diabetes (yes or no), chronic hypertension (yes or no), and subtype of pregnancy loss (stillbirth or miscarriage).

We also explored the joint effect using alternative models. First, we replaced categorical co-exposure with linear terms with a multiplicative interaction, which led to the following equation:

$$\mathrm{Logit}\left(y_{i,j}\right) = \mathrm{O}_{i,j}\beta_1 + \mathrm{T}_{i,j}\beta_2 + \left(\mathrm{O}_{i,j} \times \mathrm{T}_{i,j}\right)\beta_3 + f\left(\mathrm{RH}_{i,j}\right) + b_1\mathrm{PBL}_{i,j} + b_2\mathrm{H}_{i,j} + \theta(i) \quad (3)$$

where O and T are exposure to O_3 and temperature, respectively, and β_3 is the coefficient of the interaction effect. Second, we modeled co-exposure using a two-dimensional nonlinear term instead of the categorical variable. The model was specified as

$$\mathrm{Logit}\left(y_{i,j}\right) = t\left(\mathrm{O}_{i,j}, \mathrm{T}_{i,j}\right) + f\left(\mathrm{RH}_{i,j}\right) + b_1\mathrm{PBL}_{i,j} + b_2\mathrm{H}_{i,j} + \theta(i), \quad (4)$$

where $t()$ denotes a full tensor product of a cubic spline term for O_3 ($df = 3$) and another cubic spline term for temperature ($df = 3$). Finally, to assess how temperature modifies the association between O_3 and pregnancy loss, we developed a varying-coefficient model, which parametrized the effect of O_3 as a nonlinear function of temperature. The model was specified as.

$$\mathrm{Logit}\left(y_{i,j}\right) = \beta\left(\mathrm{T}_{i,j}\right) \times \mathrm{O}_{i,j} + g\left(\mathrm{T}_{i,j}\right) + f\left(\mathrm{RH}_{i,j}\right) + b_1\mathrm{PBL}_{i,j} + b_2\mathrm{H}_{i,j} + \theta(i), \quad (5)$$

where $g()$ denotes a natural cubic spline function ($df = 3$) to control for the nonlinear effect of temperature, and $\beta()$ is the varying-coefficient spline function ($df = 3$) to estimate the effect of a per-unit increment in O_3 according to temperature. Using Eq. (5), we calculated the effects of each 10 μg/m^3 increment in O_3 according to county-specific average temperatures in the warm season. The estimated effects were used to explore geographic variation in the association between O_3 and pregnancy loss. Statistical analyses were performed with R (v4.2.0). Statistical tests were two-sided, and $P < 0.05$ was considered indicative of statistical significance.

3. Results

3.1. Characteristics of the cases

Of the 247,305 cases of pregnancy loss analyzed in this case-crossover study, 56.3 % (139,237) and 43.7 % (108,068) were miscarriages and stillbirths, respectively, and 67.7 % (167,537) were reported by white mothers. The cases were distributed across the contiguous United States, with 50.2 % (124,157), 25.6 % (63,223), 10.7 % (26,368), 8.8 % (21,732), and 4.8 % (11,825) in the Northeast, Southeast, West, Midwest, and Southwest, respectively (Fig. 1). The reporting rate of pregnancy loss, in particular miscarriage, was geographically unbalanced. For instance, almost all miscarriage cases (98 %) were reported in the Northeast and Southeast. The mean maternal age was 28.8 (standard deviation 6.8) years. The characteristics of the cases and mothers are listed in Table S1.

Table 1 lists MDA8 O_3, temperature, RH, and PBL. There were small differences in exposure levels and meteorological factors between the hazard period and the reference period. For the cases, the mean O_3 exposure was highest in the West (101.8 μg/m^3), followed by the Southeast (101.2 μg/m^3), Southwest (95.3 μg/m^3), Midwest (89.6 μg/m^3), and Northeast (88.1 μg/m^3). In general, temperature decreased from high-latitude to low-latitude regions. There was a significant positive correlation between O_3 and temperature ($r = 0.493$, $p < 0.001$), which indicated that O_3 episodes typically occurred during heat waves (Fig. S1).

Table 1
Description of the ozone and meteorological conditions (mean ± SD) during lag0–3 in five regions.

Regions		O_3 (μg/m^3)	Temperature (°C)	Related humidity (%)	Height of planetary boundary layer (m)
Whole (*N* = 247,305)	Hazard	93.4 ± 29.9	20.2 ± 5.3	73.1 ± 13.1	832.1 ± 282.3
	Reference	93.2 ± 30.0	20.1 ± 5.5	73.1 ± 13.1	835.3 ± 281.4
Southeast (*N* = 63,223)	Hazard	101.2 ± 29.5	23.2 ± 4.4	75.4 ± 6.7	797.3 ± 172.0
	Reference	101.0 ± 29.7	23.0 ± 4.6	75.3 ± 6.7	797.9 ± 171.2
Southwest (*N* = 11,825)	Hazard	95.3 ± 28.7	26.9 ± 3.9	60.3 ± 20.6	1078.8 ± 370.9
	Reference	94.8 ± 28.4	26.7 ± 4.2	60.4 ± 20.5	1078.1 ± 370.4
West (*N* = 26,368)	Hazard	101.8 ± 31.3	20.1 ± 5.1	49.2 ± 17.4	1091.8 ± 505.0
	Reference	101.4 ± 31.4	20.0 ± 5.2	49.5 ± 17.5	1092.4 ± 503.9
Northeast (*N* = 124,157)	Hazard	88.1 ± 29.2	18.3 ± 4.7	78.0 ± 6.6	777.4 ± 216.3
	Reference	88.0 ± 29.3	18.2 ± 4.9	77.9 ± 6.6	783.2 ± 216.6
Midwest (*N* = 21,732)	Hazard	89.6 ± 25.7	18.9 ± 5.2	74.4 ± 7.0	796.8 ± 179.4
	Reference	89.3 ± 26.0	18.8 ± 5.4	74.5 ± 7.0	797.6 ± 178.2

3.2. Association between pregnancy loss and co-exposure

We examined the joint effect of O_3 and high temperature using categorical models (Figs. 2 and S3, Table S2). In general, the risk for pregnancy loss increased for each increment in O_3 or temperature. The estimates of the joint effect of co-exposure showed similar patterns irrespective of the lag of co-exposure. The joint effect was similar for both subtypes of pregnancy loss, although stillbirth was more strongly associated with co-exposure than was miscarriage (Fig. 2b). The main model focused on the association between pregnancy loss and average co-exposure for lags of 0–3 days. Given a constant level of O_3 exposure (low, moderate, or high), the odds of pregnancy loss increased at higher temperatures. For exposure to low temperature (≤18 °C), the odds increased with each increment in O_3 exposure. By contrast, the odds were unchanged or slightly decreased for exposure to high temperature (>23 °C). Compared with the low-low co-exposure group (i.e., the reference group: $O_3 \leq 78$ μg/m^3 and temperature ≤ 18 °C), the odds of pregnancy loss were significantly higher by 6.0 % (95 % CI: 1.1–11.2 %), 9.8 % (4.7–15.3 %), and 7.5 % (3.7–11.4 %) for the high-low (>104 μg/m^3 and ≤18 °C), low-high (≤78 μg/m^3 and >23 °C), and high-high (>104 μg/m^3 and >23 °C) groups, respectively. Therefore, a marginal increment in exposure to high temperature or O_3 was positively

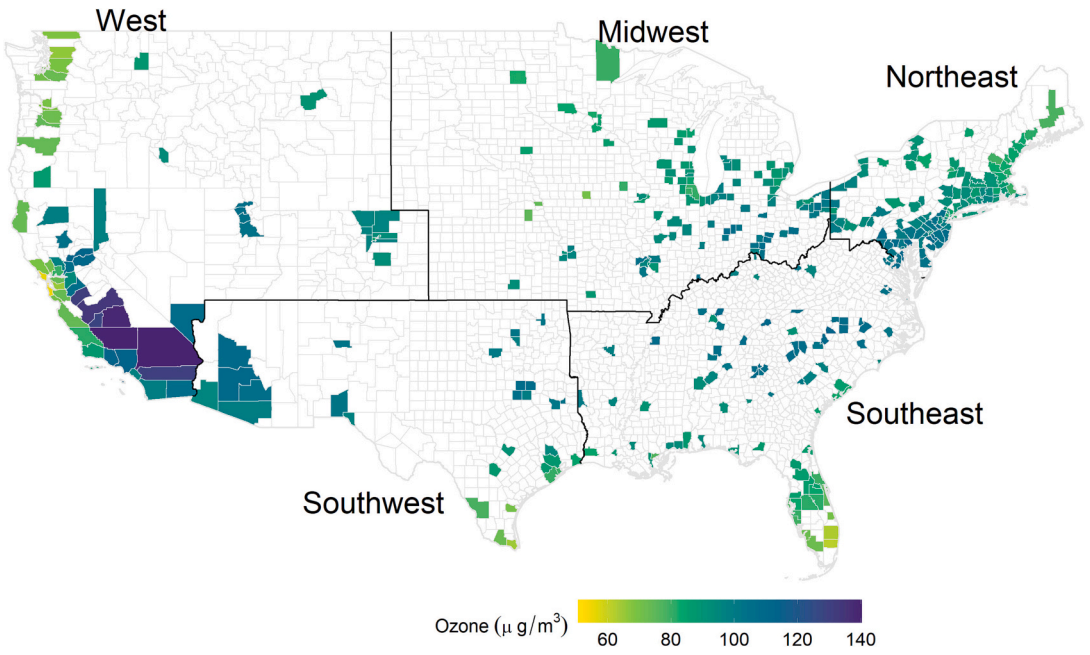


Fig. 1. Average of warm-season MDA8 ozone concentration (μg/m^3) during 1989–2004 across the 386 counties in contiguous United States. MDA8: daily maximum 8-h average; Warm season: the maximum of 6-months moving-average temperature.

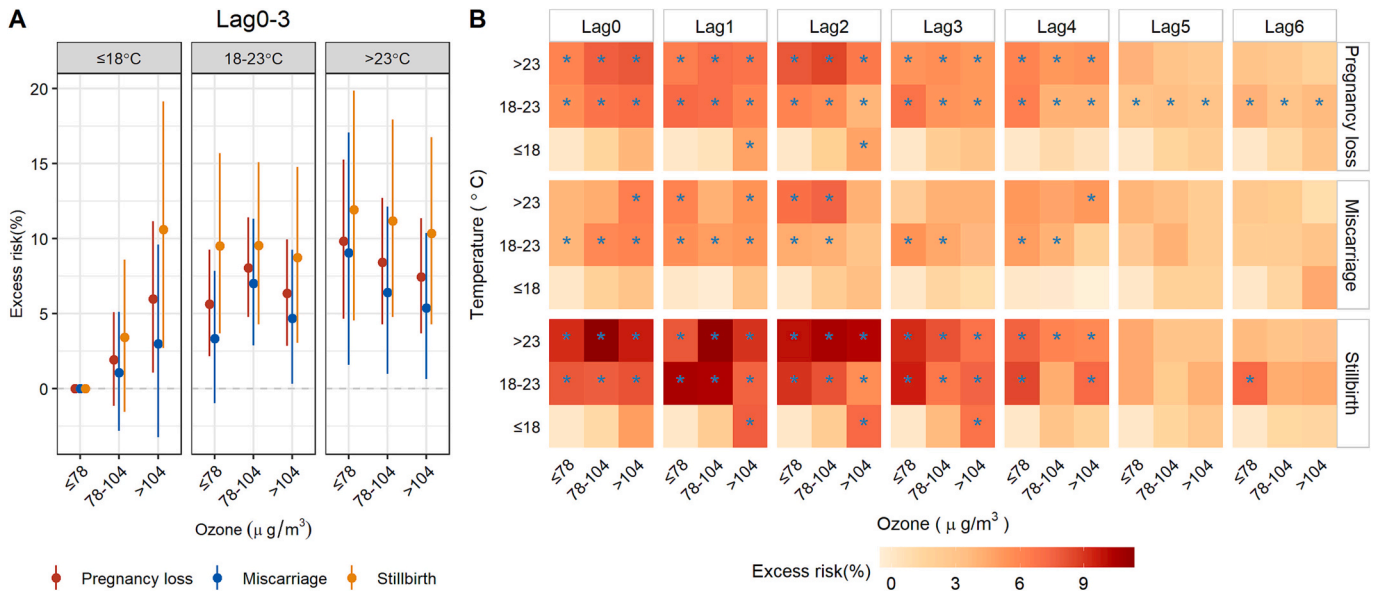


Fig. 2. The association between O_3-temperature co-exposures and risk of pregnancy loss, miscarriage or stillbirth. (A) The excess risks and corresponding 95 % confidence intervals for the averaged exposures of lag 0–3. (B) The excess risks by different lags. O_3 and temperature were categorized into three groups according to their own tertiles, and the reference exposure level was temperature $\leq$ 18 °C and $O_3 \leq$ 78 μg/m^3. *, Bonferroni-corrected $P < 0.05$.

associated with an increased risk for pregnancy loss. However, there was a potentially antagonistic interaction between O_3 and temperature, which led to a RERI of −0.084 (−0.132 to −0.036).

We estimated the joint effect (high-high vs. low-low), marginal effect of O_3 (high-low vs. low-low), marginal effect of temperature (low-high vs. low-low), and RERI by subpopulation (Fig. S3 and Table S3). Generally speaking, the effects observed specifically for the subpopulations were consistent with results of the main models. We observed significant effects of O_3, temperature, or co-exposure in nearly all subpopulations, except for some subgroups with small sizes of samples. Additionally, the interaction effects were also estimated as statistically significant for almost all subpopulations, suggesting a robust pattern in the joint effect of temperature and O_3.

Next, we quantified the joint effect using linear single-exposure, double-exposure, and interaction models (Table 2). O_3 or temperature was significantly associated with pregnancy loss in the single-exposure model. In the double-exposure model, the estimated effect of O_3 was markedly attenuated, changing the association from positive to negative. Adding the interaction term into the double-exposure model reversed the direction of the estimated association, which suggested that the main effect of O_3 was harmful. The negative association between pregnancy loss and the multiplicative interaction term was inconsistent with the RERI estimated by the categorical model, which suggested an antagonistic interaction. We developed linear models for miscarriage and stillbirth (Table 2) and for different geographic regions (Fig. S4).

Based on the complexities embedded in the joint effect, we developed a nonlinear model that linked pregnancy loss with co-exposure, parametrized by a two-dimensional nonlinear function of O_3 and temperature. As shown in Fig. S5, the odds ratio (OR) of pregnancy loss increased with each increment in O_3 exposure when temperature was maintained at <15 °C. However, the pattern changed at temperatures >15 °C. The risk for pregnancy loss was highest for co-exposure to $O_3 <$ 120 μg/m^3 and a temperature of 20–25 °C. The contour lines in Fig. S5 represent different co-exposures that had equal ORs for pregnancy loss. The contours are not parallel, which suggests complex interactions between O_3 and temperature.

3.3. *Geographic variation in the association between pregnancy loss and O_3 due to temperature*

The estimated effect of O_3 on pregnancy loss according to temperature is shown in Fig. 3. In consistency with the results above, we found positive association between O_3 and pregnancy loss at low temperatures (about 15 °C and below). When temperate rising, the association was not significant, or even became slightly negative. These results should not be interpreted as indicating a protective effect of O_3 exposure. In our opinion, the adverse effect of O_3 exposure can be masked or suppressed by the adverse effect of high temperature. The temperature-varying nonlinear curves illustrate how temperature modifies the effect of O_3. The modification pattern enables identification of the spatiotemporal

Table 2
The association between pregnancy loss and co-exposure to ozone and temperature, estimated by linear models with or without a multiplicative interaction.

Models[a]	Exposure	Excess risk (95 % confidence interval)		
		Pregnancy loss	Miscarriage	Stillbirth
Single-exposure model	O_3 (10 μg/m^3)	0.45 (0.19,0.71)	0.35 (0.01,0.68)	0.58 (0.17,1.00)
	Temperature (1 °C)	2.05 (1.85,2.25)	2.06 (1.80,2.33)	2.01 (1.71,2.31)
Double-exposures model	O_3 (10 μg/m^3)	−1.14 (−1.44,−0.85)	−1.47 (−1.86,−1.08)	−0.73 (−1.19,−0.28)
	Temperature (1 °C)	2.49 (2.26,2.72)	2.70 (2.38,3.02)	2.23 (1.90,2.57)
Interaction model	O_3 (10 μg/m^3)	3.04 (1.98,4.11)	2.76 (1.38,4.16)	3.69 (2.02,5.38)
	Temperature (1 °C)	3.92 (3.50,4.33)	4.19 (3.63,4.76)	3.70 (3.08,4.33)
	O_3 × Temperature	−0.18 (−0.22,−0.14)	−0.19 (−0.24,−0.13)	−0.18 (−0.25,−0.12)

[a] The single exposure model included only ozone (O_3) or temperature; The double-exposures model included ozone and temperature simultaneously; The interaction model further included a term of the product of ozone and temperature; All models were adjusted for relative humidity, planetary boundary layer height, and holiday. The exposures were averaged during the 0–3 lagged days.

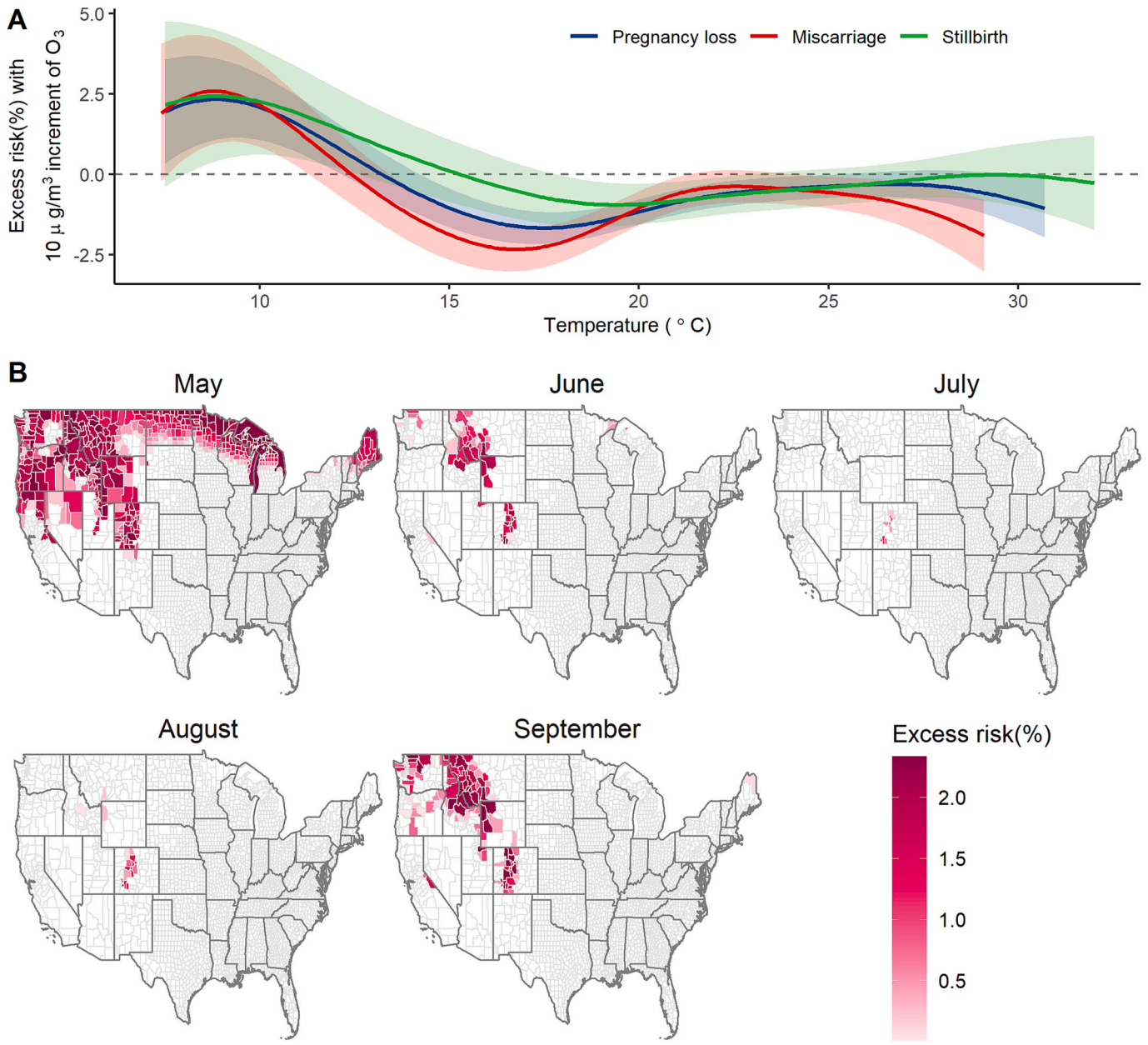


Fig. 3. Temperature-varying (A) or geographically- and temporally-varying association (B) between O_3 exposure and pregnancy loss. The excess risk for per 10 μg/m^3 increment in O_3 exposure is estimated as a function varying with different levels of temperature (A), which leads to a geographic variation in the county-specific estimates (B).

domain where O_3 is the major contributor to the adverse effect on pregnancy loss. For instance, based on the curve, geographic and temporal variation in the association between O_3 and pregnancy loss could be estimated based on the county-specific monthly average temperatures during the warm season (e.g., from May to September) in 2004 (Fig. 3b and Fig. S6). We found the association between O_3 and pregnancy loss varied geographically and temporally. Generally, on the temporal scale, O_3-related pregnancy loss was obvious in May and September. Moreover, the effects were more particularly pronounced in high-latitude or high-altitude areas, such as the Rocky Mountain region. Similar spatial-temporal variation pattern was observed in the adverse effect of O_3 on miscarriage and stillbirth (Fig. S6).

4. Discussion

To the best of our knowledge, this is the first study of the effect of short-term co-exposure to O_3 and high temperature on pregnancy loss. Each increment in exposure to O_3 or high temperature was associated with a higher risk for pregnancy loss. We also found an antagonistic effect of joint exposure to O_3 and high temperature. Although the mechanisms underlying the O_3 × temperature interaction are unclear, the antagonistic effect suggests that the risk created by O_3 exposure can be considerable in high-latitude and high-altitude regions.

Previous studies have ignored the joint effect of O_3 and high temperature on pregnancy loss or stillbirth. Most have explored the association between O_3 and stillbirth, treating temperature as a confounder to be adjusted for in the regression models. The mixed results suggest neglected effects, such as an interaction effect. For example, Dastoorpoor et al. conducted a time-series study from 2008 to 2018 in Iran and found no significant association between short-term O_3 exposure and stillbirth or spontaneous abortion after adjusting for temperature (Dastoorpoor et al., 2021). Rammah et al. conducted a case-crossover

analysis using birth and fetal death records from Texas from 2008 to 2013 and detected no significant effect of O_3 after adjusting for apparent temperature (Rammah et al., 2019). Sarovar et al. performed a time-stratified case-crossover study of fetal death certificate data from 1999 to 2009 in California and found that the effect of O_3 exposure was attenuated from significant to nonsignificant after adjustment for apparent temperature (Sarovar et al., 2020). By contrast, Mendola et al. examined delivery records from a retrospective cohort in 12 clinical centers across the United States from 2002 to 2008 and reported that acute O_3 exposure 2, 3, and 5–7 days prior to delivery was associated with stillbirth after they adjusted for temperature (Mendola et al., 2017). In this study, a comparison of single-exposure and double-exposure models showed that the effect of O_3 on pregnancy loss was markedly attenuated by adjustment for temperature.

Although no previous study on the interaction between O_3 and temperature has focused on pregnancy loss, there is indirect evidence for preterm birth (PTB) and low birthweight (LBW). A study in Australia on PTB and LBW reported that the effect of exposure to O_3 during the entire pregnancy was enhanced at low temperatures (Chen et al., 2018a). However, studies in California and China showed that a heat wave or high temperature enhances the estimated effect of O_3 on PTB, LBW, and large for gestational age (Chen et al., 2023; Sun et al., 2020). Several studies have investigated the interactive effect of O_3 and temperature on mortality. A meta-analysis by Ito et al. showed that the effect of O_3 on mortality was negatively associated with mean temperature (Ito et al., 2005). Studies conducted in Shanghai and Hong Kong reported that the effect of O_3 on mortality and cardiovascular morbidity was enhanced by extreme cold weather (Cheng and Kan, 2012; Qiu et al., 2013). These results are consistent with our findings. However, Chen et al. analyzed the air temperature–stratified association between O_3 exposure and total mortality in eight European urban areas and found a synergistic effect of high temperature and O_3 (Chen et al., 2018b). In a study of 95 US cities, increasing temperature enhanced the effect of O_3-related cardiovascular mortality, in particular in northern cities (Ren et al., 2009). Therefore, the effect of O_3 on mortality or adverse birth outcomes is modified by temperature in a complex manner, and previous findings are inclusive.

The attenuated association between O_3 and pregnancy loss at high temperatures is plausible for the following reasons. First, the adverse effects of O_3 and high temperature on pregnancy loss may share biological pathways. For example, O_3 and high temperature induce oxidative stress and systemic inflammation, which leads to endothelial dysfunction and the development of a pro-thrombotic vascular state (Balmes et al., 2019; Samuels et al., 2022; Sharkhuu et al., 2011). O_3 and heat exposure also reduce placental blood flow and impair the energy efficiency of the placenta, restricting fetal growth (Miller et al., 2017; Samuels et al., 2022). Therefore, the joint effect of co-exposure can be saturated, resulting in a weak effect of a marginal increment in O_3 exposure in hot weather. Second, O_3 exposure has been utilized as medical treatment to decreases blood pressure and prevent the progression of hypertensive disease (Akcılar et al., 2015), which may protect against hypertension and cardiovascular diseases caused by exposure to high temperature (Alpérovitch et al., 2009; Yin and Wang, 2017). Third, the relationship between O_3 and temperature is determined by nonlinear atmospheric physical and chemical processes. This complex relationship may have been simplified by our epidemiological models, possibly leading to an underestimation of the uncertainty in our findings.

Global climate change will increase the frequency, intensity, and duration of extreme weather events (e.g., heat waves) as well as the O_3 concentration (Zhang et al., 2018). Exploring the joint effect of O_3 and high temperature has important implications for public health. Till now, this joint effect has not been thoroughly investigated due to complicated relationship between O_3 and temperature in the epidemiology study. Considering the formation process of O_3, it is plausible that O_3 acts as an intermediate factor between temperature and health outcomes (e.g., mortality) (Reid et al., 2012). Given their correlations, temperature can also play a role as a confounder on the effect of O_3. Previous studies mainly focused on the confounding effect, and controlled for it by adjusting for temperature in regression models. Our study further estimated the effect modification of temperature which hasn't been explored yet. Additionally, the joint effect has not been concluded consistently. For example, high temperature could magnify the effect of air pollution on children's health outcomes (e.g., asthma) (Lu et al., 2022). However, our finding that temperature antagonizes O_3-related pregnancy loss implies that the overall effect of combined exposure may be overestimated if additive interaction is employed. Therefore, further studies should investigate co-exposure to temperature and O_3 pollution using advanced methods to examine modifying, confounding, and mediating effects to accurately estimate the disease burden of climate change.

Our study provides important evidence with regard to pregnancy loss, an important but previously neglected public health issue. The estimated number of stillbirths after 28 gestational weeks was 2.0 million (90 % uncertainty range [UR]: 1.9–2.2) globally in 2019, for a rate of 13.9 per 1000 total births (90 % UR: 13.5–15.4) (Hug et al., 2021). Mothers who experience pregnancy loss might feel abandoned or unsupported by their family, lack confidence during subsequent pregnancies, avoid social contact, or suffer depressive episodes, all of which impose substantial costs on women and their families, government, and society (Heazell et al., 2016; Kavanaugh and Hershberger, 2005). In response to our findings, policymakers should implement regulations to mitigate environmental issues, which would improve fetal health and promote social equity.

The strengths of this study are its coverage of a large and diverse population in the contiguous United States and its use of a case-crossover design, which enabled us to control for unobservable individual characteristics that did not change during the study period (as well as other risk factors for stillbirth, such as maternal age at conception, smoking behavior, and history of termination). Yet this study also has several limitations. First, because of the lack of residential addresses, O_3 exposure and temperature were assigned based on county, which precluded our measuring exposure at a lower level of aggregation or total exposure across all daily activities and possibly resulted in misclassification of exposure. In addition, when the temperature is high, pregnant women are more likely to remain indoors, so the ambient O_3 concentration may not accurately reflect their O_3 exposure. This could partly explain why the risk for O_3-related pregnancy loss is lower at high temperatures. Second, O_3 concentrations after the kriging imputation were used to evaluate county-level exposure, which might have introduced exposure bias, but the fact that reanalyzes of associations with original monitoring data produced consistent results suggests that the influence of this bias would have been minimal (Fig. S7). Third, because we lacked information on its causes, we treated pregnancy loss as a single outcome. However, different causes of death may be associated with different risk factors. The use of an aggregate single outcome instead of cause-specific fetal deaths might have led to an underestimation of the strength of the association. The similar results for stillbirth and miscarriage suggest the robustness of the findings. Fourth, we documented the time at which pregnancy loss was observed but not the time of actual embryonic or fetal demise. Because we could not ascertain the exact time of stillbirth, the effect of exposure lag should be interpreted carefully.

5. Conclusion

We evaluated the association between pregnancy loss and short-term co-exposure to O_3 and temperature. Short-term O_3 exposure and short-term exposure to high temperature during the warm season increased the risk for pregnancy loss. Temperature modified the relationship between O_3 and pregnancy loss, and the marginal effect of O_3 varied geographically, being mainly observed in the northern United States. In high-latitude cities, controlling for O_3 pollution would protect maternal

health.

Funding

National Natural Science Foundation of China (42293324 and 42175182), Ministry of Science and Technology of the People's Republic of China (2022YFC3703000), National Institute of Environmental Health Sciences of the United States (ES031986), and Energy Foundation (G-2208-34045, G-2107-33169, and R-2109-33379).

CRediT authorship contribution statement

Mingkun Tong: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Meng Wang:** Methodology, Resources, Supervision, Validation. **Pengfei Li:** Methodology, Supervision, Validation. **Jicheng Gong:** Supervision, Validation. **Tong Zhu:** Supervision, Validation. **Tao Xue:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data underlying this article are based on publicly available datasets, and data sources have been provided in the article

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.166088.

References

Akcılar, R., Akçer, S., Şimşek, H., Akcılar, A., Bayat, Z., Genç, O., 2015. The effect of ozone on blood pressure in DOCA-salt-induced hypertensive rats. Int. J. Clin. Exp. Med. 8, 12783–12791.

Alpérovitch, A., Lacombe, J.-M., Hanon, O., Dartigues, J.-F., Ritchie, K., Ducimetière, P., et al., 2009. Relationship between blood pressure and outdoor temperature in a large sample of elderly individuals: the Three-City study. Arch. Intern. Med. 169, 75–80.

Balmes, J.R., Arjomandi, M., Bromberg, P.A., Costantini, M.G., Dagincourt, N., Hazucha, M.J., et al., 2019. Ozone effects on blood biomarkers of systemic inflammation, oxidative stress, endothelial function, and thrombosis: the multicenter ozone study in oldEr subjects (MOSES). PLoS One 14, e0222601.

Ban, J., Lu, K., Wang, Q., Li, T., 2022. Climate change will amplify the inequitable exposure to compound heatwave and ozone pollution. One Earth 5, 677–686.

Chen, G., Guo, Y., Abramson, M.J., Williams, G., Li, S., 2018a. Exposure to low concentrations of air pollutants and adverse birth outcomes in Brisbane, Australia, 2003–2013. Sci. Total Environ. 622-623, 721–726.

Chen, K., Wolf, K., Breitner, S., Gasparrini, A., Stafoggia, M., Samoli, E., et al., 2018b. Two-way effect modifications of air pollution and air temperature on total natural and cardiovascular mortality in eight European urban areas. Environ. Int. 116, 186–196.

Chen, J., Guo, L., Liu, H., Jin, L., Meng, W., Fang, J., et al., 2023. Modification effects of ambient temperature on associations of ambient ozone exposure before and during pregnancy with adverse birth outcomes: a multicity study in China. Environ. Int. 172, 107791.

Cheng, Y., Kan, H., 2012. Effect of the interaction between outdoor air pollution and extreme temperature on daily mortality in Shanghai, China. J. Epidemiol. 22, 28–36.

Chersich, M.F., Pham, M.D., Areal, A., Haghighi, M.M., Manyuchi, A., Swift, C.P., et al., 2020. Associations between high temperatures in pregnancy and risk of preterm birth, low birth weight, and stillbirths: systematic review and meta-analysis. BMJ 371, m3811.

Dastoorpoor, M., Khanjani, N., Moradgholi, A., Sarizadeh, R., Cheraghi, M., Estebsari, F., 2021. Prenatal exposure to ambient air pollution and adverse pregnancy outcomes in Ahvaz, Iran: a generalized additive model. Int. Arch. Occup. Environ. Health 94, 309–324.

Filleul, L., Cassadou, S., Médina, S., Fabres, P., Lefranc, A., Eilstein, D., et al., 2006. The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003. Environ. Health Perspect. 114, 1344–1347.

Health Effects Institute, 2020. State of Global Air 2020. Special Report. Health Effects Institute, Boston, MA.

Heazell, A.E.P., Siassakos, D., Blencowe, H., Burden, C., Bhutta, Z.A., Cacciatore, J., et al., 2016. Stillbirths: economic and psychosocial consequences. Lancet 387, 604–616.

Hug, L., You, D., Blencowe, H., Mishra, A., Wang, Z., Fix, M.J., et al., 2021. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. Lancet 398, 772–785.

Ito, K., De Leon, S.F., Lippmann, M., 2005. Associations between ozone and daily mortality: analysis and meta-analysis. Epidemiology 16, 446–457.

Jhun, I., Fann, N., Zanobetti, A., Hubbell, B., 2014. Effect modification of ozone-related mortality risks by temperature in 97 US cities. Environ. Int. 73, 128–134.

Kavanaugh, K., Hershberger, P., 2005. Perinatal loss in low-income African American parents. J Obstet Gynecol Neonatal Nurs 34, 595–605.

Lu, C., Zhang, Y., Li, B., Zhao, Z., Huang, C., Zhang, X., et al., 2022. Interaction effect of prenatal and postnatal exposure to ambient air pollution and temperature on childhood asthma. Environ. Int. 167, 107456.

Mendola, P., Ha, S., Pollack, A.Z., Zhu, Y., Seeni, I., Kim, S.S., et al., 2017. Chronic and acute ozone exposure in the week prior to delivery is associated with the risk of stillbirth. Int. J. Environ. Res. Public Health 14, 731.

Miller, C.N., Dye, J.A., Ledbetter, A.D., Schladweiler, M.C., Richards, J.H., Snow, S.J., et al., 2017. Uterine artery flow and offspring growth in long-Evans rats following maternal exposure to ozone during implantation. Environ. Health Perspect. 125, 127005.

Qiu, H., Yu, I.T., Wang, X., Tian, L., Tse, L.A., Wong, T.W., 2013. Cool and dry weather enhances the effects of air pollution on emergency IHD hospital admissions. Int. J. Cardiol. 168, 500–505.

Rammah, A., Whitworth, K.W., Han, I., Chan, W., Symanski, E., 2019. Time-varying exposure to ozone and risk of stillbirth in a nonattainment urban region. Am. J. Epidemiol. 188, 1288–1295.

Reid, C.E., Snowden, J.M., Kontgis, C., Tager, I.B., 2012. The role of ambient ozone in epidemiologic studies of heat-related mortality. Environ. Health Perspect. 120, 1627–1630.

Ren, C., Williams, G.M., Mengersen, K., Morawska, L., Tong, S., 2008. Does temperature modify short-term effects of ozone on total mortality in 60 large eastern US communities? — An assessment using the NMMAPS data. Environ. Int. 34, 451–458.

Ren, C., Williams, G.M., Mengersen, K., Morawska, L., Tong, S., 2009. Temperature enhanced effects of ozone on cardiovascular mortality in 95 large US communities, 1987–2000: assessment using the NMMAPS data. Arch. Environ. Occup. Health 64, 177–184.

Richardson, D.B., Kaufman, J.S., 2009. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am. J. Epidemiol. 169, 756–760.

Samuels, L., Nakstad, B., Roos, N., Bonell, A., Chersich, M., Havenith, G., et al., 2022. Physiological mechanisms of the impact of heat during pregnancy and the clinical implications: review of the evidence from an expert group meeting. Int. J. Biometeorol. 66, 1505–1513.

Sarovar, V., Malig, B.J., Basu, R., 2020. A case-crossover study of short-term air pollution exposure and the risk of stillbirth in California, 1999–2009. Environ. Res. 191, 110103.

Schwartz, J., Bind, M.A., Koutrakis, P., 2017. Estimating causal effects of local air pollution on daily deaths: effect of low levels. Environ. Health Perspect. 125, 23–29.

Schwartz, J., Fong, K., Zanobetti, A., 2018. A national multicity analysis of the causal effect of local pollution, NO2, and PM2.5 on mortality. Environ. Health Perspect. 126, 87004.

Schwarz, L., Hansen, K., Alari, A., Ilango, S.D., Bernal, N., Basu, R., et al., 2021. Spatial variation in the joint effect of extreme heat events and ozone on respiratory hospitalizations in California. Proc. Natl. Acad. Sci. 118, e2023078118.

Sharkhuu, T., Doerfler, D.L., Copeland, C., Luebke, R.W., Gilmour, M.I., 2011. Effect of maternal exposure to ozone on reproductive outcome and immune, inflammatory, and allergic responses in the offspring. J. Immunotoxicol. 8, 183–194.

Shi, W., Sun, Q., Du, P., Tang, S., Chen, C., Sun, Z., et al., 2020. Modification effects of temperature on the ozone-mortality relationship: a nationwide multicounty study in China. Environ. Sci. Technol. 54, 2859–2868.

Sun, Y., Ilango, S.D., Schwarz, L., Wang, Q., Chen, J.-C., Lawrence, J.M., et al., 2020. Examining the joint effects of heatwaves, air pollution, and green space on the risk of preterm birth in California. Environ. Res. Lett. 15, 104099.

Tong, M., Li, P., Wang, M., Sun, Y., Han, Y., Liu, H., et al., 2022. Time-varying association between fetal death and gestational exposure to ambient fine particles: a nationwide epidemiological study of 49 million fetuses in the contiguous US from 1989 to 2004. Int. J. Epidemiol. 51, 1984–1999.

Xie, G., Sun, L., Yang, W., Wang, R., Shang, L., Yang, L., et al., 2021. Maternal exposure to PM2.5 was linked to elevated risk of stillbirth. Chemosphere 283, 131169.

Yin, Q., Wang, J., 2017. The association between consecutive days' heat wave and cardiovascular disease mortality in Beijing, China. BMC Public Health 17, 1–9.

Zhang, J., Gao, Y., Luo, K., Leung, L.R., Zhang, Y., Wang, K., et al., 2018. Impacts of compound extreme weather events on ozone in the present and future. Atmos. Chem. Phys. 18, 9861–9877.

Zhang, J., Wei, Y., Fang, Z., 2019. Ozone pollution: a major health hazard worldwide. Front. Immunol. 10.

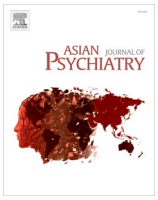
Zhang, H., Zhang, X., Wang, Q., Xu, Y., Feng, Y., Yu, Z., et al., 2021. Ambient air pollution and stillbirth: an updated systematic review and meta-analysis of epidemiological studies. Environ. Pollut. 278, 116752.

Asian Journal of Psychiatry 88 (2023) 103737

Contents lists available at ScienceDirect

Asian Journal of Psychiatry

journal homepage: www.elsevier.com/locate/ajp



Letter to the Editor

Triggering factors of major depressive disorder among adolescents in China

To the editor:

There is a rapidly increasing prevalence of major depressive disorder (MDD) among adolescents in the world (Shorey et al., 2022) and adolescent depression may lead to impaired social functioning, self-harm, and suicidal behavior (Clayborne et al., 2019; Sandoval-Ato et al., 2018). Adolescence, a formative and turbulent phase in which rapid physical, social, emotional, and cognitive changes occur, is a particularly vulnerable period of psychological disorders in one's life (Thapar et al., 2012). Prevention of depression among adolescents is an important cost-effective strategy for MDD intervention (Bodden et al., 2018), considering the unsatisfactory curative effect (Bschor and Kilarski, 2016) and high relapse rate (Rao and Chen, 2022) of depressive disorder.

To help develop depression prevention strategies among adolescents, we conducted a qualitative study to explore risk factors of depression among Chinese adolescents with MDD, especially focusing on the direct and indirect impacts of family factors. Adolescents with MDD aged 12–18 years were invited to participate in the study after providing informed consent. The sample size was determined according to the " thematic saturation" principle, and a total of 30 adolescents with MDD were consecutively recruited at the department of child and adolescent psychiatry in Beijing Anding Hospital, Capital Medical University. Three main themes were identified as family risk factors in the onset of depression, including academic pressure interacting with family's unreasonable response, trauma in family relationships, and lack of parental support and help for school problems (Table 1).

The first main theme was academic pressure interacting with family unreasonable response. Chinese adolescents faced the academic pressure rooted in the university entrance examination system in China, which was strengthened by families with unreasonable responses. As a university education especially in prestigious universities remarkably increases employment opportunities and the chances of success in life, parents with sky-high hopes for their children have been exerting more and more pressure on adolescents. Exam preparation began at an early age even in primary school due to intense competition. The high expectations which can't be achieved and the academic pressure made adolescents a sense of powerlessness and self-loathing. In addition, to achieve such academic goals, most parents were overly authoritarian and forced adolescents to spend almost all their time studying, especially under-performers (Luebbe et al., 2018). Chinese parents tend to use less praise and they are apt to use more authoritarian parenting strategies by controlling their children (Wang et al., 2021). Moreover, adolescents usually faced negative comments and severe criticism from their parents due to unsatisfactory academic performances. The parents frequently complained their kids were not doing as brilliantly as "next-door kids". These negative comments and accusations hurt the child's self-esteem which subsequently internalized low self-esteem and pessimism (Zhou et al., 2020). Some parents even use verbal and physical violence to punish their children for their poor academic performance.

We shed light on the links between trauma in family relationships and adolescent depression. Lacking emotional support, ineffective communication, and physical and verbal violence resulted in unsatisfactory parent-child relationships. Along with China's urbanization and modernization, lots of parents in rural and underdeveloped regions moved to cities for job opportunities. The absence of parents led to decreased care and disrupted parent-child relationships which subsequently increased the risk of psychological problems. They felt lonely and lack of warmth/love because of insufficient companionship and psychological support. Some parents showed emotional anger, even physical and verbal violence, which erupted into parent-child conflict. Some adolescents were entangled in family conflict which was characterized by long-term discordant relationships of parents, endless quarrels even violence. Destructive marital relationships impaired the emotional security of the children, which would inevitably result in mental health problems (Li et al., 2020). The adolescents with siblings complained that parental favoritism led to disappointment, resentment, and adverse emotional experiences. A study also found that parental favoritism was related to anxiety and depression, and it was speculated that adolescents were very sensitive to perceived parental preferences (Tong et al., 2022).

The third main theme was lacking parental support and help for school problems, especially school bullying. A study has shown the correlation of school problems with adolescent depression (Ye et al., 2023). Some adolescents reported that they had been bullied on campus and treated unfairly, but their parents misjudged school bullying as small conflicts between students and ignored it without any support. Unfortunately, some parents even criticized and blamed their children. Without parental support and help and effective solutions to these problems, adolescents suppress emotional needs instead of seeking help until they become seriously ill.

Chinese teenagers are facing the high risk of depression during their grown-up, and family factors exert direct or indirect influence on it. Academic pressure interacting with family unreasonable response, trauma in family relationships, and lack of parental support and help for school problems were identified as three main themes in family risk factors for the onset of depression. As we know, there are few prevention policies or strategies for adolescents' depression at the national level. Therefore, to reduce the high incidence of MDD among adolescents, the family should be integrated into the prevention strategy in China. The parents should realize the academic pressure of the adolescents and change their traditional notion "a boy becomes a dragon and a girl become a phoenix in the future". And they also should be educated about the skills of communication, conflict management, and emotional relief.

https://doi.org/10.1016/j.ajp.2023.103737
Received 17 July 2023; Received in revised form 4 August 2023; Accepted 10 August 2023
Available online 11 August 2023

Table 1
Main Themes and Subthemes in triggering factors of major depressive disorder.

Main themes	N (%)	Explanation
Academic pressure interacting with family unreasonable response	20 (66.7 %)	Parental high expectations on academic performance that are difficult to achieve. Authoritarian parenting style. Negative comments and accusations, even violent punishment toward the adolescents for poor academic performance.
Trauma in family relationships	19 (63.3 %)	The unsatisfactory relationship between parents and adolescents characterized by lack of emotional support/reciprocity and /or physical and verbal violence. Entangled in conflict of parents. Parental favoritism in multiple-child families.
Lack of supporting for school problems	16 (53.3 %)	Insufficient supporting or negligence for school problems faced by their adolescents, e.g., school bullying, social awkwardness, injustice from teachers, and high-pressure teaching and management.

Ethics statements

The study protocol was reviewed and approved by both the institutional review board (IRB) of Peking University Health Science Center (IRB00001052-22086) and Beijing Anding Hospital, Capital Medical University (2022-70). All Patients and their parents provided written informed consent prior to enrollment in the study and the manuscript does not include personal identity information.

Author contributions

Haibo Wang, Xu Chen, and Hongling Chu contributed to the conception and design of the study. Wenjie Yun, Kun Liu, Gaoyang Xu, Mengqi Liu and Qianqian Zhu contributed to the acquisition, analysis, or interpretation of data for the work, while Wenjie Yun wrote the first draft of the manuscript. Wang Haibo, Chu Hongling, and Chen Xu were responsible for the conception, design, and revision of the paper, while Wang Haibo was responsible for the quality control and proofreading of the paper. All authors contributed to the manuscript revision, and read, and approved the submitted version.

Funding

The study was funded by the Sci-Tech Innovation 2030 - Major Project of Brain Science and Brain-inspired Intelligence Technology (2021ZD0200600).

Declaration of Competing Interest

All authors have reported no conflicts of interest.

Acknowledgments

We are grateful to all the young people sharing their experiences in the study. Special thanks to Dr. Wu Yangfeng for his supervision and advice, and also to psychiatrists at the Department of Child and Adolescent Psychiatry, Beijing Anding Hospital, Capital Medical University for their valuable support in recruiting young people.

References

Bodden, D., Stikkelbroek, Y., Dirksen, C., 2018. Societal burden of adolescent depression, an overview and cost-of-illness study. J. Affect. Disord. 241, 256–262.
Bschor, T., Kilarski, L.L., 2016. Are antidepressants effective? A debate on their efficacy for the treatment of major depression in adults. Expert Rev. Neurother. 16 (4), 367–374.
Clayborne, Z.M., Varin, M., Colman, I., 2019. Systematic review and meta-analysis: adolescent depression and long-term psychosocial outcomes. J. Am. Acad. Child Adolesc. Psychiatry 58 (1), 72–79.
Li, C., Jiang, S., Fan, X., Zhang, Q., 2020. Exploring the impact of marital relationship on the mental health of children: does parent–child relationship matter? J. Health Psychol. 25 (10–11), 1669–1680.
Luebbe, A.M., Tu, C., Fredrick, J.W., 2018. Socialization goals, parental psychological control, and youth anxiety in Chinese students: moderated indirect effects based on school type. J. Youth Adolesc. 47, 413–429.
Rao, U., Chen, L.-A., 2022. Characteristics, correlates, and outcomes of childhood and adolescent depressive disorders. Dialog-. Clin. Neurosci.
Sandoval-Ato, R., Vilela-Estrada, M.A., Mejia, C.R., Caballero-Alvarado, J., 2018. Suicide risk associated with bullying and depression in high school. Rev. Chil. De. Pedia 89 (2), 208–215.
Shorey, S., Ng, E.D., Wong, C.H.J., 2022. Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. Br. J. Clin. Psychol. 61 (2), 287–305. https://doi.org/10.1111/bjc.12333.
Thapar, A., Collishaw, S., Pine, D.S., Thapar, A.K., 2012. Depression in adolescence. Lancet 379 (9820), 1056–1067. https://doi.org/10.1016/S0140-6736(11)60871-4.
Tong, J., Zhang, T., Chen, F., Wang, Q., Zhao, X., Hu, M., 2022. Prevalence and contributing factors of childhood trauma, anxiety, and depression among adolescents from two-child families in China. Front. Psychiatry 13, 782087. https://doi.org/10.3389/fpsyt.2022.782087.
Wang, C., Shao, X., Do, K.A., Lu, H.K., O'Neal, C.R., Zhang, Y., 2021. Using participatory culture-specific consultation with Asian American communities: identifying challenges and solutions for Asian American immigrant families. J. Educ. Psychol. Consult. 31 (1), 17–38.
Ye, Z., Wu, D., He, X., Ma, Q., Peng, J., Mao, G., Feng, L., Tong, Y., 2023. Meta-analysis of the relationship between bullying and depressive symptoms in children and adolescents. BMC Psychiatry 23 (1), 1–13.
Zhou, Z., Shek, D.T.L., Zhu, X., Dou, D., 2020. Positive youth development and adolescent depression: a longitudinal study based on mainland Chinese high school students. Int. J. Environ. Res. Public Health 17 (12), 4457. https://doi.org/10.3390/ijerph17124457.

Wenjie Yun[a,b]
[a] *Peking University Clinical Research Institute, Peking University First Hospital, Xueyuan Rd 38#, Haidian Dist, Beijing 100191, China*
[b] *Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China*

Kun Liu[a]
[a] *Department of Computer Science, Peking University, Yiheyuan Rd 5#, Haidian Dist, Beijing 100191, China*

Gaoyang Xu[a], Mengqi Liu[a]
[a] *National Clinical Research Center for Mental Disorders, Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Beijing Institute for Brain Disorders, Capital Medical University, Deshengmenwai 5#, Xicheng Dist, Beijing 100088, China*

Qianqian Zhu[a]
[a] *Department of Psychiatry, Wuhan Mental Health Center, Tongji Medical College of Huazhong University of Science and Technology, Jianshe Rd 920#, Jiangan Dist, Wuhan, China*

Hongling Chu[a]
[a] *Clinical Epidemiology Research Center, Peking University Third Hospital, Huayuan Rd 49#, Haidian Dist, Beijing 100083, China*

Xu Chen[a]
[a] *National Clinical Research Center for Mental Disorders, Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Beijing Institute for Brain Disorders, Capital Medical University, Deshengmenwai 5#, Xicheng Dist, Beijing 100088, China*

Haibo Wang[a,b,*]
[a] *Peking University Clinical Research Institute, Peking University First Hospital, Xueyuan Rd 38#, Haidian Dist, Beijing 100191, China*
[b] *Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China*

* Corresponding author at: Peking University Clinical Research Institute, Peking University First Hospital, No.38, Xueyuan Rd, Haidian District, Beijing 100191, China.

E-mail address: hbwang2005@163.com (H. Wang).

Journal of Affective Disorders 341 (2023) 211–218

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

The moderating role of psychological resilience in the relationship between falls, anxiety and depressive symptoms

Jingjing Wang [a,b], Shaojie Li [a,b], Yang Hu [a,b], Longbing Ren [a,b], Rui Yang [a,b], Yuling Jiang [a,b], Mingzhi Yu [a,b], Zhouwei Liu [a,b], Yifei Wu [a,b], Ziqi Dong [a,b], Chi Zhang [c], Wentian Dong [d,e,*], Yao Yao [b,f,**]

[a] School of Public Health, Peking University, Beijing, China
[b] China Center for Health Development Studies, Peking University, Beijing, China
[c] The Key Laboratory of Geriatrics, Beijing Institute of Geriatrics, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing Hospital, National Center of Gerontology of National Health Commission, Beijing, China
[d] Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health, Peking University, Beijing, China
[e] National Clinical Research Center for Mental Disorders, Peking University Sixth Hospital, Beijing, China
[f] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China

ARTICLE INFO

Keywords:
Falls
Severe falls
Anxiety symptoms
Depressive symptoms
Psychological resilience
CLHLS

ABSTRACT

Background: There is a greater likelihood of anxiety and depression among older adults who suffer falls. This study examined the relationships of falls and severe falls with anxiety and depressive symptoms, and the moderating role of psychological resilience on these associations.

Methods: Our study recruited participants from the 2018 Chinese Longitudinal Healthy Longevity Survey (CLHLS), a nationally representative cohort study. A total of 11,857 participants included in the analysis. We used a linear regression model to investigate the relationship between falls/severe falls and anxiety/depressive symptoms, adjusting for a range of potential covariates and a bootstrapping sample test to examine the potential moderating role of psychological resilience in these relationships.

Results: Older adults who suffered the falls have higher anxiety/depressive symptoms ($\beta = 0.28$ [0.23, 0.32] for anxiety symptoms, $p < 0.001$; $\beta = 0.21$ [0.16, 0.25] for depressive symptoms, $p < 0.001$), and those who suffered the severe falls have higher anxiety/depressive symptoms ($\beta = 0.30$ [0.24, 0.37] for anxiety symptoms, $p < 0.001$; $\beta = 0.21$ [0.15, 0.27] for depressive symptoms, $p < 0.001$), in the fully adjusted model. The relationship between falls/severe falls and anxiety/depressive symptoms was mitigated in participants with higher levels of psychological resilience.

Limitations: The present study is based on cross-sectional data, which limits the ability to infer causal relationships.

Conclusions: Falls/severe falls were positively associated with anxiety and depression, and that psychological resilience could moderate this association. Our findings suggest that psychological resilience may be an effective target for intervention and prevention of fall-related symptoms of anxiety and depression.

1. Introduction

Age-related anxiety and depression are common problems among older adults in both developing and developed countries. Anxiety and Depression are at the forefront of a wide range of co-morbid psychological, neurological, and substance abuse disorders, as well as pain in older adults (Haller et al., 2014; Wittchen, 2002; Charlson et al., 2016). China has the largest older population across the globe. According to two meta-analyses conducted in China, anxiety and depression symptoms are prevalent in the older population at 22.1 % and 23.6 %, respectively (Su et al., 2011; Li et al., 2014). For the prevention and control of anxiety and depression, it is imperative to identify modifiable

* Co-correspondence to: W. Dong, Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health, Peking University, National Clinical Research Center for Mental Disorders, Peking University Sixth Hospital, Beijing, China.
** Correspondence to: Y. Yao, Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
E-mail addresses: dongwentian@bjmu.edu.cn (W. Dong), yao.yao@bjmu.edu.cn (Y. Yao).

https://doi.org/10.1016/j.jad.2023.08.060
Received 23 April 2023; Received in revised form 30 July 2023; Accepted 11 August 2023
Available online 12 August 2023

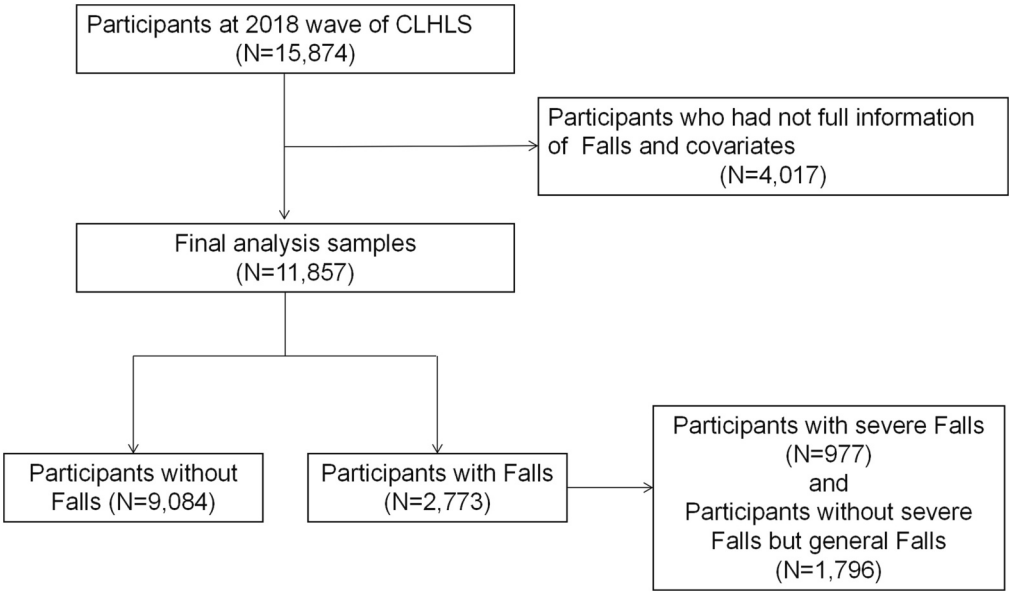


Fig. 1. Flow chart of the study.

risk factors and implement targeted interventions.

Previous studies have examined the link between falls-related psychological concerns and anxiety/depression in older people from developed countries (Hull et al., 2013; Painter et al., 2012). Several underlying mechanisms of the relationship between falls and depression have proposed, including the mediating effects of social factors (Lohman et al., 2023; Zhang et al., 2021), frailty (Kim and Cho, 2022), antidepressant use (Lohman et al., 2021), cognitive and motor slowing (Cohen et al., 2021) and functional limitation (Yang et al., 2023), and the moderating effects of social factors (e.g., marital status, living arrangement, family and friend network) (Rhee et al., 2021;Yang et al., 2023). In contrast, little attention has been given to the underlying mechanisms of the relationship between falls and anxiety. As far as we are aware, only one study has examined the mediating effects of functional ability and social participation on falls and anxiety (Yue et al., 2022). In theory, the pathway between falls and anxiety may involve unique variances that are shared with depression, such as reduced attention or muscle strength, but further empirical investigation is necessary (Hallford et al., 2017).

To date, little attention has been given to the symptoms of anxiety and depression arising from falls, and its psychological mechanism was yet to investigate among Chinese older adults. Studying the mechanism underlying the relationship between falls and severe falls and anxiety and depression symptoms will contribute to the literature and practice concerning the prevention and treatment of anxiety and depression caused by falls. In this context, older adults in a positive mental state may be better protected against falls/severe falls that trigger anxiety or depression. A previous study looked at the psychological dimension primarily from a negative perspective, for example, the fear of falling (Gambaro et al., 2022). No study has, however, examined the mechanism of anxiety/depressive symptoms caused by falls/severe falls from the perspective of positive mental health. According to the interaction model of psychological resilience, regulatory factors can modify the impact of stress or danger on the development or adaptation of psychosocial functions (Masten and Reed, 2002). When faced with negative experiences, psychological resilience allows one to adapt and cope with. Previous empirical studies have shown that psychological resilience moderates anxiety and depression in older adults (Miller et al., 2023).

As an effort to fill the above gaps, we conducted a large-scale national study in China (Chinese Longitudinal Healthy Longevity Survey, CLHLS) to examine the relationship between falls/severe falls and anxiety/depressive symptoms. Moreover, we investigated whether psychological resilience could moderate the association between falls/severe falls and anxiety/depression symptoms in this population.

2. Methods

2.1. Participants

The study recruited 15,874 participants from the 2018 wave of the CLHLS. The CLHLS selected half of all counties and cities in 22 of the 31 provinces, covering approximately 85 % of the total population of those provinces. CLHLS used a targeted random sample design to ensure sample representativeness. Data collection was conducted through face-to-face interviews. A detailed sampling description has been reported in our previous work (Zeng et al., 2017). CLHLS was approved by the Ethics Committee of Peking University (IRB00001052–13074). As the information regarding falls/severe falls and anxiety/depression in older adults was only assessed in the CLHLS 2018, we conducted a cross-sectional study to investigate whether falls/severe falls are associated with anxiety/depression symptoms among older Chinese individuals. During the study, 4017 participants were excluded for not meeting the following inclusion criteria: (a) age 60 or older, and (b) no missing data for the independent variable (falls/severe falls) and other covariates. The final sample consisted of 11,857 individuals. Flowchart for the inclusion and exclusion of participants can be found in Fig. 1.

2.2. Measurements

2.2.1. Falls/severe falls

Falling was assessed by asking the question: “Have you fallen in the past year?”. The answer choices were “Yes” or “No”. The respondents who answered “Yes” were considered to have fallen in the past year; among them Severe falls were further assessed by asking the following question: “Was it severe enough to require hospital treatment?” There was a choice of “Yes/No”. The respondents who answered “Yes” were considered to have suffered a severe fall in the past year.

2.2.2. Anxiety/depressive symptoms

We assessed anxiety symptoms using the 7-item Generalized Anxiety Disorder (GAD-7) scale (Spitzer et al., 2006), which has been widely used to measure anxiety symptoms in older Chinese (Li et al., 2021)

Table 1
Demographics characteristics and the distribution of anxiety and depressive symptoms.

Variables	Respondents		Anxiety symptoms				Depressive symptoms			
	n	%	M	SD	t/F	P-value	M	SD	t/F	P-value
Sex					10.45	<0.001			12.28	<0.001
Female	6742	56.9 %	1.68	2.98			8.05	4.37		
Male	5115	43.1 %	1.16	2.36			7.08	4.18		
Marital status					−2.77	0.006			−14.51	<0.001
Married	4900	41.3 %	1.37	2.65			6.95	4.26		
Divorced/Widowed/Never married	6957	58.7 %	1.51	2.80			8.11	4.29		
Current residence					19.85	<0.001			25.28	<0.001
City	2862	24.1 %	1.19	2.32			7.13	4.40		
Town	3917	33.0 %	1.61	2.96			7.83	4.34		
Rural	5078	42.9 %	1.50	2.77			7.75	4.22		
Age					1.53	0.204			62.20	<0.001
60–70	1642	13.8 %	1.48	2.78			6.63	4.33		
71–80	2844	24.0 %	1.55	2.85			7.21	4.37		
81–90	2909	24.5 %	1.43	2.73			7.79	4.48		
91 above	4462	37.6 %	1.41	2.66			8.15	4.08		
Education year					47.45	<0.001			171.75	<0.001
0	5899	49.8 %	1.69	2.95			8.32	4.27		
1–6	3767	31.8 %	1.32	2.60			7.16	4.21		
6 above	2191	18.5 %	1.08	2.29			6.55	4.29		
Family economic status					94.46	<0.001			108.59	<0.001
Very poor	167	1.4 %	4.16	5.30			11.89	5.98		
General	11,378	96.0 %	1.44	2.69			7.61	4.26		
Very rich	312	2.6 %	0.71	1.68			5.90	3.95		
Co-residence					10.71	<0.001			59.29	<0.001
With household member(s)	9579	80.8 %	1.40	2.68			7.42	4.23		
Alone	1851	15.6 %	1.69	2.99			8.39	4.59		
In an institution	427	3.6 %	1.72	2.90			8.92	4.52		
Chronic diseases					4.94	<0.001			5.46	0.007
Yes	8042	67.8 %	1.54	2.79			7.77	4.42		
No	3815	32.2 %	1.28	2.62			7.32	4.07		
Smoking					−4.98	<0.001			−7.73	<0.001
Yes	1752	14.8 %	1.19	2.33			6.93	4.08		
No	10,105	85.2 %	1.50	2.80			7.75	4.34		
Drinking alcohol					−5.89	<0.001			−11.49	<0.001
Yes	1665	14.0 %	1.13	2.43			6.53	4.19		
No	10,192	86.0 %	1.51	2.78			7.81	4.31		
Leisure activity					−7.58	<0.001			−19.29	<0.001
Yes	9693	81.7 %	1.35	2.56			7.27	4.24		
No	2164	18.3 %	1.94	3.38			9.24	4.31		
Childhood hunger					28.20	<0.001			29.94	<0.001
Yes	7962	67.2 %	1.59	2.87			7.81	4.35		
No	3019	25.5 %	1.18	2.47			7.10	4.37		
Missing	876	7.3 %	1.21	2.28			7.78	3.63		
Activities of daily living					−6.63	<0.001			−19.18	<0.001
Without assistance	8675	73.2 %	1.35	2.63			7.17	4.23		
Assistance	3182	26.8 %	1.75	3.01			8.87	4.31		
Falls					105.5	<0.001			94.00	<0.001
No falls	9084	76.6 %	1.26	2.50			7.34	4.15		
Falls	1796	15.1 %	1.97	3.21			8.39	4.62		
Severe falls	977	8.2 %	2.34	3.53			8.92	4.80		

Participants aged 65 years and older were asked to recall a series of feelings from the past two weeks using seven items with four options each: 0 = never, 1 = for several days, 2 = more than half of the days, and 3 = almost every day. Scores ranged from 0 to 21. A higher score indicates more severe anxiety symptoms. The internal consistency reliability of the anxiety scale was satisfactory (Cronbach's alpha coefficient = 0.920).

Depressive symptoms were examined using the 10-item the Center for Epidemiologic Studies Depression Scale (CES-D-10) (Andresen et al., 1994), which has been validated and applied in Chinese older adults (Cheng and Chan, 2005). The participants were asked to recall experiences from the past week. Each item had four options: 0 = "rarely", 1 = "some days" (1–2 days), 2 = "occasionally" (3–4 days), and 3 = "most of the time" (5–7 days). The total score ranged from 0 to 30. Higher scores indicated more severe depressive symptoms. Its internal consistency reliability was acceptable (Cronbach's alpha coefficient = 0.749).

2.2.3. Psychological resilience

In accordance to previous study, resilience was measured using following five items (Yang and Wen, 2015): "Do you always look on the bright side of things?", "Can you make your own decisions concerning your personal affairs?", "Do you find that the older you get, the more useless you are, and have trouble doing anything?", "Do you often feel fearful or anxious?", and "Do you often feel lonely and isolated?". The answers ranged from 1 (always) to 5 (never), and the total score ranged from 5 to 25. Higher scores indicated a higher level of psychological resilience. In terms of internal consistency reliability, the resilience scale has an acceptable value of 0.625 (Cronbach's alpha coefficient), which is above the acceptable value of 0.6 (Nunnally, 1978).

2.2.4. Covariates

Our analysis controlled for a set of covariates, including demographic, socioeconomic characteristics, lifestyles, and health status (Yue et al., 2022; Yang and Wen, 2015). Demographic characteristics included the following: sex (male, female), age (60–70, 71–80, 81–90,

Table 2
The Association between falls/severe falls and anxiety/depression symptoms among Chinese elderly.

Variables	Anxiety symptoms			Depression symptoms		
	β	95%CI	*P*-values	β	95%CI	P-values
Falls						
Model 1[a]	0.30	(0.26,0.35)	<0.001	0.29	(0.25,0.33)	<0.001
Model 2[b]	0.30	(0.26,0.34)	<0.001	0.24	(0.20,0.29)	<0.001
Model 3[c]	0.29	(0.25,0.33)	<0.001	0.23	(0.19,0.27)	<0.001
Model 4[d]	0.28	(0.23,0.32)	<0.001	0.21	(0.16,0.25)	<0.001
Severe falls						
Model 1[a]	0.35	(0.28,0.41)	<0.001	0.33	(0.26,0.39)	<0.001
Model 2[b]	0.34	(0.27,0.40)	<0.001	0.27	(0.20,0.33)	<0.001
Model 3[c]	0.33	(0.26,0.39)	<0.001	0.25	(0.19,0.32)	<0.001
Model 4[d]	0.30	(0.24,0.37)	<0.001	0.21	(0.15,0.27)	<0.001

Note:
[a] Unadjusted model.
[b] Adjusted for model 1 and Demographics variables (age, sex, education level, marital status and current residence).
[c] Adjusted for model 2 and Family socioeconomic variables (family economic status, co-residence with family members, leisure activity, childhood hunger).
[d] Adjusted for model 3 and health status (number of chronic diseases, smoking and drinking, activities of daily living).

90+), educational level (0 year, 0–6 years, above 6 years), marital status (married, divorced/widowed/never married) and current residence (city, town, rural). Socioeconomic characteristics included family economic status (very poor, general, very rich), and co-residence with family members (with household members, alone, in an institution). Leisure activity included following activities: tai ji, square dance, visiting and interacting with friends, other outdoor activity, garden work, reading newspapers and books, playing cards or Mahjong, watching television or listening to radio, social activities (Yes = 6 above/activity; No = 6 score/without activity). Childhood starvation was recorded as yes or no. Health status included the number of chronic diseases, smoking (yes, no), drinking (yes, no) and activities of daily living (ADL) (there is a total score calculated for the dependence of following variables: bathing, dressing, toileting, indoor transferring, continence, and eating; (Yes = 6 above/assistance;No = 6 score/without assistance)).

2.3. Statistic analysis

The demographic characteristics and the distribution of anxiety and depression symptoms were presented either as means (standard deviation [SD]) or as frequencies with percentages. An independent sample *t*-test and a one-way ANOVA were conducted using IBM SPSS 27.0. We used linear regression models to investigate the association between falls/severe falls and anxiety/depressive symptoms. Model 1 was unadjusted; Model 2 was adjusted for age, gender, educational level, marital status, and current residence; Model 3 further controlled for family economic status, co-residence with family members, leisure activities, and childhood hunger based on Model 2; and Model 4 further controlled for chronic diseases, smoking and drinking, and ADL based on Model 3. We conducted subgroup analyses for age, sex, educational level, and resilience level in order to examine the heterogeneity of the association between falls/severe falls and anxiety/depressive symptoms. A sensitivity analysis was also conducted to assess the robustness of the above relationship. For example, we excluded participants with these scores from Model 4, as those with severe cognitive impairment may be more likely to recall incidents relating to falls.

SPSS PROCESS macro with a bootstrapping test was used to estimate the moderating effect of psychological resilience on the association between falls/severe falls and anxiety/depressive symptoms (Hayes, 2018). Age, sex, education level, marital status, current residence,

Table 3
Subgroup analyses of associations between falls and anxiety/depression.

Subgroups	Anxiety	P-interaction	Depression	P-interaction
Associations between falls and anxiety/depression				
By sex		0.231		0.098
Men	0.23 (0.17,0.29) ***		0.24 (0.17,0.30) ***	
Women	0.29 (0.24,0.35) ***		0.18 (0.13,0.24) ***	
By age		<0.001		<0.001
60–80	0.37 (0.29,0.45) ***		0.30 (0.23,0.38) ***	
81 above	0.23 (0.18,0.28) ***		0.16 (0.11,0.21) ***	
Education year		0.188		0.015
0	0.30 (0.24,0.36) ***		0.18 (0.12,0.23) ***	
1–6	0.25 (0.18,0.33) ***		0.21 (0.14,0.28) ***	
6 above	0.20 (0.12,0.29) ***		0.28 (0.18,0.39) ***	
Resilience level		<0.001		<0.001
Low	0.41 (0.30,0.51) ***		0.23 (0.16,0.30) ***	
High	0.11 (0.07,0.16) ***		0.08 (0.02,0.13)**	
Associations between severe falls and anxiety/depression				
By sex		0.710		0.480
Men	0.27 (0.17,0.36) ***		0.16 (0.05,0.27)**	
Women	0.31 (0.22,0.39) ***		0.23 (0.15,0.31) ***	
By age		<0.001		0.030
60–80	0.53 (0.40,0.66) ***		0.30 (0.18,0.43) ***	
81 above	0.21 (0.13,0.29) ***		0.17 (0.10,0.25) ***	
Education year		0.904		0.895
0	0.29 (0.20,0.38) ***		0.21 (0.13,0.30) ***	
1–6	0.33 (0.22,0.44) ***		0.24 (0.12,0.35) ***	
6 above	0.25 (0.10,0.40) **		0.12 (−0.06,0.29)	
Resilience level		<0.001		<0.001
Low	0.43 (0.28,0.58) ***		0.22 (0.12,0.32) ***	
High	0.15 (0.08,0.23) ***		0.09 (−0.00,0.19)	

Associations of Falls/Severe falls with Anxiety/Depression symptoms among subgroups.
Model 4 controlling for Demographics variables (age, sex, education level, marital status and current residence); Family socioeconomic variables (family

economic status, co-residence with family members, Leisure activity, childhood hunger); health status (number of chronic diseases, smoking and drinking, activities of daily living).
*P < 0.05; **P < 0.01; ***P < 0.001.

family economic status, co-residence with family members, leisure activity, childhood hunger, number of chronic diseases, smoking, drinking and ADL were considered covariates in this moderation analysis. Regarding missing data of key continuous variables (anxiety/depressive symptoms and psychological resilience), multiple imputations by the Markov Chain Monte Carlo (MCMC) method were used to impute 20 times with the aim of mitigating the potential bias (Cleophas and Zwinderman, 2018). SPSS (27.0 version, IBM Co.) was used to perform all analyses and $P < 0.05$ indicates statistically significant.

3. Results

3.1. Descriptive statistics

A summary of the demographic characteristics and distribution of anxiety and depressive symptoms scores can be found in Table 1. The study included a total of 11,857 older individuals, with 43.1 % were males. People aged 60–70, 71–80, and 81–90 constituted 13.8 %, 24 %, and 24.5 % of the population, respectively. Among older adults, 49.8 % were illiterate (0 schooling year) and 31.8 % had 1–6 years of schooling. A total of 76.6 % of the participants had not experienced falls, with a mean anxiety score of 1.26 and a mean depression score of 7.34. The percentage of normal falls in older adults was 15.1 %, with a mean anxiety score of 1.97 and a mean depression score of 8.39. A percentage of older adults who had experienced severe falls was 8.2 %, with a mean score of 2.34 for anxiety symptoms and a mean score of 8.92 for depression symptoms.

3.2. Regression model results

Table 2 shows both unadjusted and adjusted regression models that examined the relationship between falls/severe falls and anxiety/depressive symptoms. In the unadjusted model (Model 1), the relationship between falls and anxiety/depressive symptoms was statistically significant (β = 0.30 [0.26, 0.35] for anxiety symptoms, $p < 0.001$; β = 0.29 [0.25, 0.33] for depressive symptoms, p < 0.001), and the relationship between severe falls and anxiety/depressive symptoms was also statistically significant (β = 0.35 [0.28, 0.41] for anxiety symptoms, p < 0.001; β = 0.33 [0.26, 0.39] for depressive symptoms, p < 0.001).

Using adjusted models (from Models 2 to 4), where age, gender, education level, marital status, current residence, family economic status, co-residence with family members, leisure activities, childhood hunger, chronic diseases, smoking and drinking, and activities of daily living were controlled for, the increase in control covariates resulted in a gradual decrease in the correlation between falls and anxiety/depressive symptoms; this result was still statistically significant. In Model 4,the relationship between falls and anxiety/depressive symptoms was significant (β = 0.28 [0.23, 0.32] for anxiety symptoms, $p < 0.001$; β = 0.21 [0.16, 0.25] for depressive symptoms, p < 0.001), as was the relationship between severe falls and anxiety/depressive symptoms (β = 0.30 [0.24, 0.37] for anxiety symptoms, p < 0.001; β = 0.21 [0.15, 0.27] for

Table 4
Moderating role of psychological resilience to falls – anxiety/depression symptoms.

	Variables	Coefficient	SE	t value	Bootstrap 95 % CI	
					Lower	Upper
Anxiety symptoms(Y)						
Independent variable	Falls(X)	0.18	0.02	9.35	0.14	0.22
Moderator variable	Psychological resilience(W)	−0.43	0.01	−44.04	−0.45	−0.40
Interaction	X × W	−0.18	0.03	−9.62	−0.24	−0.12
Constant		0.70	0.12	7.29	0.48	0.93
Depression symptoms(Y)						
Independent variable	Falls(X)	0.08	0.01	5.52	0.05	0.10
Moderator variable	Psychological resilience(W)	−0.74	0.01	−105.84	−0.76	−0.72
Interaction	X × W	−0.09	0.02	−6.64	−0.12	−0.06
Constant		0.44	0.07	6.30	0.29	0.58

Note:
Control Variable: Demographics variables (age, sex, education level, marital status and current residence);Family socioeconomic variables (family economic status, co-residence with family members , Leisure activity, childhood hunger); health status (number of chronic diseases, smoking and drinking, activities of daily living).

Table 5
Moderating role of psychological resilience to severe falls – anxiety/depression symptoms.

	Variables	Coefficient	SE	t value	Bootstrap 95 % CI	
					Lower	Upper
Anxiety symptoms(Y)						
Independent variable	Severe Falls(X)	0.17	0.03	5.45	0.10	0.23
Moderator variable	Psychological resilience(W)	−0.45	0.01	−51.41	−0.48	−0.43
Interaction	X × W	−0.22	0.05	−7.69	−0.31	−0.12
Constant		0.72	0.12	7.46	0.50	0.95
Depression symptoms(Y)						
Independent variable	Severe Falls(X)	0.06	0.02	2.53	0.01	0.10
Moderator variable	Psychological resilience(W)	−0.76	0.01	−118.41	−0.77	−0.74
Interaction	X × W	−0.09	0.02	−4.61	−0.14	−0.05
Constant		0.45	0.08	6.44	0.30	0.59

Note:
Control Variable: Demographics variables (age, sex, education level, marital status and current residence);Family socioeconomic variables (family economic status, co-residence with family members , Leisure activity, childhood hunger); health status (number of chronic diseases, smoking and drinking, activities of daily living).

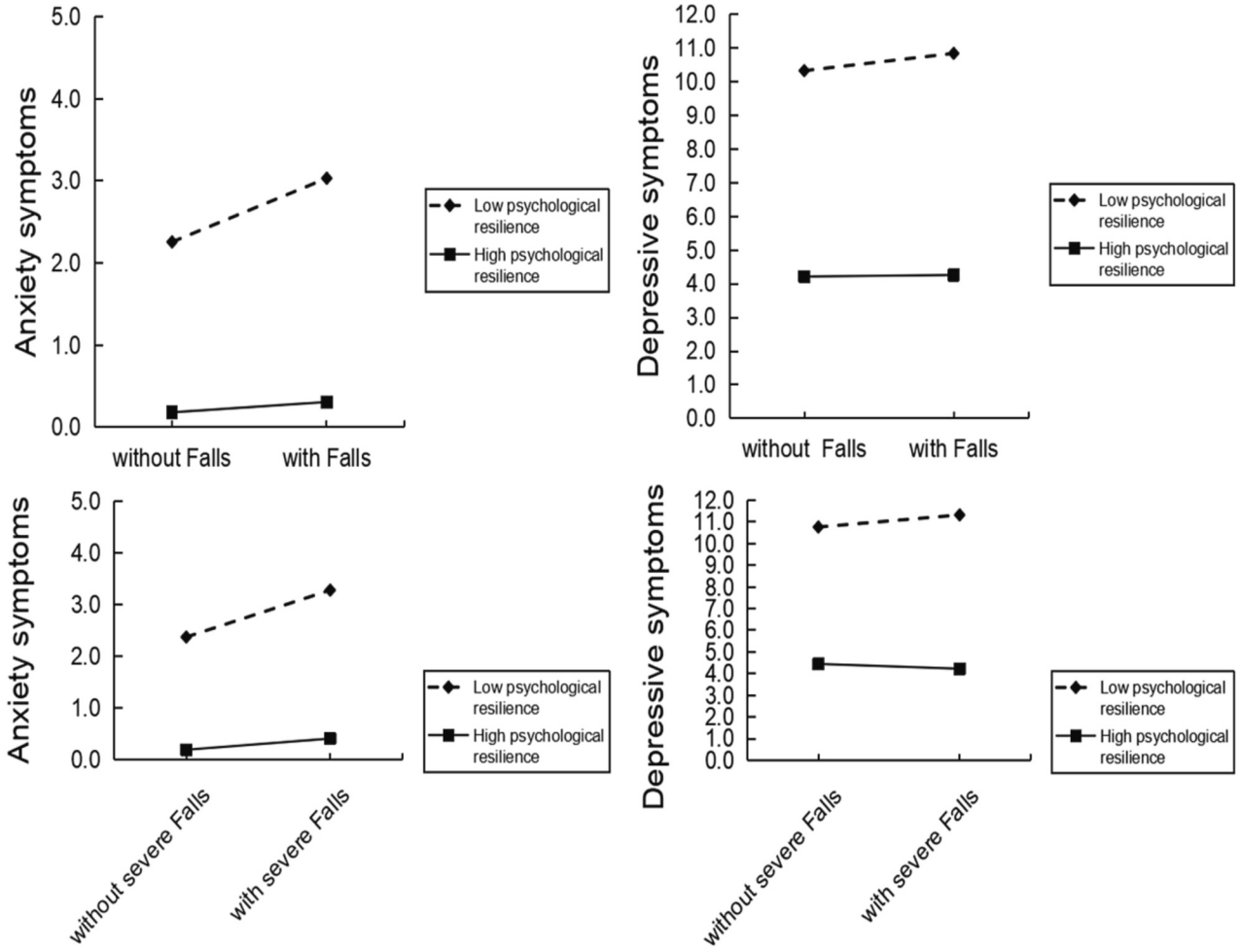


Fig. 2. The interaction effect of Falls/severe falls and psychological resilience on anxiety/depressive symptoms.

depressive symptoms, $p < 0.001$).

3.3. Subgroup and sensitivity analyses

The results of the subgroup analyses of the relationship between falls/severe falls and anxiety/depressive symptoms are shown in Table 3. When examining the relationship between falls/severe falls and anxiety symptoms, the association was robust in the following subgroup analyses: sex, age, education, and resilience level. We found, however, that the relationship between falls/severe falls and depressive symptoms was heterogeneous across education levels and levels of resilience. Accordingly, the relationship between severe falls and depressive symptoms did not exist in the group with more than six years of education and in the group with high levels of resilience.

Table A6 (Supplementary materials) indicates that the associations between falls/severe falls and anxiety/depressive symptoms remained significant after excluding participants with severe cognitive impairment and MMSE scores <21. In both the unadjusted and adjusted models, consistent results were shown.

3.4. Moderating analyses

According to Tables 4 and 5 and Fig. 2, psychological resilience moderated the association between falls/severe falls and anxiety/depressive symptoms. The bootstrapping results showed a significant interaction effect for falls and anxiety symptoms ($\beta = -0.18$ [−0.24, −0.12]), severe falls and anxiety symptoms ($\beta = -0.22$ [−0.31, −0.12]), falls and depressive symptoms ($\beta = -0.09$ [−0.12, −0.06]), and severe falls and depressive symptoms ($\beta = -0.09$ [−0.14, −0.05]). The relationship between falls/severe falls and anxiety/depressive symptoms was mitigated in participants with higher levels of psychological resilience.

4. Discussion

Using a nationally representative sample of data from the CLHLS, this study examined the relationship between falls/severe falls and anxiety/depressive symptoms and the moderating effects of psychological resilience. As a result of our research, we found that falls/severe falls were positively associated with anxiety and depression, and that psychological resilience could moderate this association. The results of our study indicate a promising direction for future research regarding the prevention of fall-related symptoms of anxiety and depression.

First, our study showed that participants who had falls/severe falls had more serious anxiety/depressive symptoms, which is consistent with previous studies (Yue et al., 2022; Grenier et al., 2014; Hoffman et al., 2017). A possible explanation could be that anxiety symptoms are correlated with four fall-related psychological concerns (fear of falling, fall-related self-efficacy, balance confidence, and outcome expectancy), whereas depressive symptoms are associated with avoidance of activity because of the fear of falling (Hull et al., 2013). In addition, according to the subgroup analyses, the relationship between severe falls and anxiety was robust regardless of age, sex, educational and resilience level, while depressive symptoms in this association were not significant in the group with more than six years of education. This may be because people with middle school education or higher have higher levels of

health literacy (Liu et al., 2015), which can help participants to learn more about falls prevention knowledge.

Second, our study revealed the moderating role of psychological resilience in the relationship between falls/severe falls and anxiety/depressive symptoms. To our knowledge, this is the first study to explore the moderating role of psychological resilience on the relationships between falls/severe falls and anxiety/depressive symptoms. A relevant qualitative study showed that participating in physical activity after falling enhances psychological resilience as a protective factor (Martin and Kasser, 2021). Moreover, many previous studies have investigated that the protective effect of resilience on anxiety/depressive symptoms in different groups such as veterans (Pietrzak et al., 2009), adolescent survivors of the Wenchuan earthquake (Ying et al., 2014), and children (Ding et al., 2017). Consistent with Masten and Reed's (2002) interaction model of psychological resilience, moderating factors alter the influence of risk factors on psycho-social functioning. In this study, high psychological resilience decreased the relationship between falls/severe falls and anxiety/depressive symptoms. In the low psychological resilience group, falls and severe falls **were positively associated with** anxiety and depressive symptoms. These results highlight the need to pay attention to interventions aimed at psychological resilience to promote its protective role against the effects of falls/severe falls on anxiety/depressive symptoms among older Chinese adults.

Our study has several strengths including the sample representativeness and novel findings of moderating role of psychological resilience in the association of falls and symptoms of anxiety and depression. There are some limitations to our study. First, this study used a cross-sectional design, which makes it difficult to interpret causal relationships. Longitudinal data should be applied in future studies to further elucidate the causal direction. Second, in line with previous studies, this study used a self-report form which does not remove the recall bias. Further experimental study with more accurate documentation of falls and symptoms of anxiety and depression was warranted.

5. Conclusion

This study found that falls and severe falls were significantly associated with higher risks of anxiety and depression symptoms in Chinese older adults. The effects of falls/severe falls on anxiety/depressive symptoms were moderated by psychological resilience. An increased correlation between falls/severe falls and anxiety/depressive symptoms was observed among participants with lower levels of psychological resilience. In light of this, future psychiatric intervention strategies on fall-related symptoms of anxiety and depression could target the protective role of psychological resilience.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2023.08.060.

CRediT authorship contribution statement

Jingjing Wang, Wentian Dong, and Yao Yao designed the study. Data collection was led by Jingjing Wang and Yao Yao, analysed by Jingjing Wang and Yao Yao. The manuscript was written by Jingjing Wang and Yao Yao and reviewed/revised by all authors.

Funding

This research was supported by the National Natural Science Foundation of China (72374013).

Data statement

The data presented in this study are available upon request from the corresponding author (Dr. Yao Yao, yao.yao@bjmu.edu.cn).

Declaration of competing interest

None.

Acknowledgements

We are grateful to the CLHLS participants for providing the data for this research. The CLHLS was supported by funds from the U.S. National Institute on Aging, National Institutes of Health, the Duke/Duke-NUS Collaboration Pilot Project, the National Natural Science Foundation of China, the China Social Science Foundation, and the United Nations Fund for Population Activities (UNFPA). The CLHLS was managed by the Center for Healthy Aging and Development Studies, Peking University. We also thank the support from the Healthy Aging Consortium of the China Cohort Consortium (see http://chinacohort.bjmu.edu.cn/).

References

Andresen, E.M., Malmgren, J.A., Carter, W.B., Patrick, D.L., 1994. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am. J. Prev. Med. 10 (2), 77–84. https://doi.org/10.1016/S0749-3797(18)30622-6.

Charlson, F.J., Baxter, A.J., Cheng, H.G., Shidhaye, R., Whiteford, H.A., 2016. The burden of mental, neurological, and substance use disorders in China and India: a systematic analysis of community representative epidemiological studies. Lancet 388 (10042), 376–389. https://doi.org/10.1016/S0140-6736(16)30590-6.

Cheng, S.T., Chan, A.C., 2005. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. Int. J. Geriatr. Psychiatry 20 (5), 465–470. https://doi.org/10.1002/gps.1314.

Cleophas, T.J., Zwinderman, A.H., 2018. Bayesian statistics: Markov chain Monte Carlo sampling. In: Modern Bayesian Statistics in Clinical Research. Springer International Publishing, pp. 119–130. https://doi.org/10.1007/978-3-319-92747-3_12.

Cohen, J.N., Seng, E., Foley, F.W., 2021. Cognitive and motor slowing mediate the relationship between depression and falls in multiple sclerosis patients. Mult. Scler. Relat. Disord. 50, 102808 https://doi.org/10.1016/j.msard.2021.102808.

Ding, H., Han, J., Zhang, M., Wang, K., Gong, J., Yang, S., 2017. Moderating and mediating effects of resilience between childhood trauma and depressive symptoms in chinese children. J. Affect. Disord. 211, 130–135. https://doi.org/10.1016/j.jad.2016.12.056.

Gambaro, E., Gramaglia, C., Azzolina, D., Campani, D., Molin, A.D., Zeppegno, P., 2022. The complex associations between late life depression, fear of falling and risk of falls. A systematic review and meta-analysis. Ageing Res. Rev. 73, 101532 https://doi.org/10.1016/j.arr.2021.101532.

Grenier, S., Payette, M., Langlois, F., Vu, T.T.M., Bherer, L., 2014. Depressive symptoms are independently associated with recurrent falls in community-dwelling older adults. Int. Psychogeriatr. 26 (9), 1511–1519. https://doi.org/10.1017/S104161021400074X.

Haller, H., Cramer, H., Lauche, R., Gass, F., Dobos, G.J., 2014. The prevalence and burden of subthreshold generalized anxiety disorder: a systematic review. BMC Psychiatry 14 (1), 128. https://doi.org/10.1186/1471-244X-14-128.

Hallford, D.J., Nicholson, G., Sanders, K., McCabe, M.P., 2017. The association between anxiety and falls: a meta-analysis. J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. 72 (5), 729–741. https://doi.org/10.1093/geronb/gbv160.

Hayes, A.F., 2018. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, Second ed. Guilford Press.

Hoffman, G.J., Hays, R.D., Wallace, S.P., Shapiro, M.F., Ettner, S.L., 2017. Depressive symptomatology and fall risk among community-dwelling older adults. Soc. Sci. Med. 1982 (178), 206–213. https://doi.org/10.1016/j.socscimed.2017.02.020.

Hull, S.L., Kneebone, I.I., Farquharson, L., 2013. Anxiety, depression, and fall-related psychological concerns in community-dwelling older people. Am. J. Geriatr. Psychiatry 21 (12), 1287–1291. https://doi.org/10.1016/j.jagp.2013.01.038.

Kim, Y.H., Cho, C.M., 2022. The mediating effect of frailty in the relationship between depression and falls among older people living alone in Korea. Iran. J. Public Health 51 (3), 596–605. https://doi.org/10.18502/ijph.v51i3.8936.

Li, D., Zhang, D., Shao, J., Qi, X., Tian, L., 2014. A meta-analysis of the prevalence of depressive symptoms in Chinese older adults. Arch. Gerontol. Geriatr. 58, 1–9. https://doi.org/10.1016/j.archger.2013.07.016.

Li, W., Sun, H., Xu, W., Ma, W., Yuan, X., Wu, H., Kou, C., 2021. Leisure activity and cognitive function among Chinese old adults: the multiple mediation effect of anxiety and loneliness. J. Affect. Disord. 294, 137–142. https://doi.org/10.1016/j.jad.2021.07.051 doi.

Liu, Y.B., Liu, L., Li, Y.F., Chen, Y.L., 2015. Relationship between health literacy, health-related behaviors and health status: a survey of elderly Chinese. Int. J. Environ. Res. Public Health 12 (8), 9714–9725. https://doi.org/10.3390/ijerph120809714.

Lohman, M.C., Fairchild, A.J., Merchant, A.T., 2021. Antidepressant use partially mediates the association between depression and risk of falls and fall injuries among older adults. J. Gerontol. A Biol. Sci. Med. Sci. 76 (9), e171–e178. https://doi.org/10.1093/gerona/glaa253.

Lohman, M.C., Fallahi, A., Mishio Bawa, E., Wei, J., Merchant, A.T., 2023. Social mediators of the association between depression and falls among older adults. J. Aging Health 35 (7–8), 593–603. https://doi.org/10.1177/08982643231152276.

Martin, S., Kasser, S.L., 2021. The role of resilience: physical activity continuation after falling in adults with multiple sclerosis. Disabil. Health J. 14 (2), 101046. https://doi.org/10.1016/j.dhjo.2020.101046.

Masten, A.S., Reed, M.G.J., 2002. Resilience in development. In: Snyder, C.R., Lopez, S.J. (Eds.), Handbook of Positive Psychology. Oxford University Press, New York, pp. 74–88.

Miller, L.R., Divers, R., Reed, C., Pugh, E., Calamia, M., 2023. Resilience as a moderator of depression and anxiety: a bidimensional approach to predictors of subjective cognition in older adults. Aging Ment. Health 27 (1), 29–34. https://doi.org/10.1080/13607863.2021.2013432.

Nunnally, J.D., 1978. Psychometric Theory, 2nd ed. McGraw-Hill, New York.

Painter, J.A., Allison, L., Dhingra, P., Daughtery, J., Cogdill, K., Trujillo, L.G., 2012. Fear of falling and its relationship with anxiety, depression, and activity engagement among community-dwelling older adults. Am. J. Occup. Ther. 66 (2), 169–176. https://doi.org/10.5014/ajot.2012.002535.

Pietrzak, R.H., Johnson, D.C., Goldstein, M.B., Malley, J.C., Rivers, A.J., Morgan, C.A., Southwick, S.M., 2009. Psychosocial buffers of traumatic stress, depressive symptoms, and psychosocial difficulties in veterans of operations enduring freedom and iraqi freedom: the role of resilience, unit support, and postdeployment social support. J. Affect. Disord. 120 (1), 188–192. https://doi.org/10.1016/j.jad.2009.04.015.

Rhee, M.K., Jang, Y., Kim, S.Y., Chang, S., 2021. The moderating role of social factors in the relationship between an incident of fall and depressive symptoms: a study with a national sample of older adults in South Korea. Aging Ment. Health 25 (6), 1086–1093. https://doi.org/10.1080/13607863.2020.1758911.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092 doi.

Su, L., Cai, Y., Shi, S., Wang, L., 2011. A meta analysis of prevalence in anxiety disorders of elderly people in China. J. Clin. Psychiatry 21, 87–90 (CNKI:SUN:LCJS.0.2011-02-010).

Wittchen, H., 2002. Generalized anxiety disorder: prevalence, burden, and cost to society. Depress. Anxiety 16 (4), 162–171. https://doi.org/10.1002/da.10065.

Yang, Y., Wen, M., 2015. Psychological resilience and the onset of activity of daily living disability among older adults in China: a nationwide longitudinal analysis. J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. 70 (3), 470–480. https://doi.org/10.1093/geronb/gbu068.

Yang, R., Wang, H., Tracy, E.L., Jo, Y.J., Sward, K.A., Edelman, L.S., Demiris, G., 2023. What is the relationship between falls, functional limitations, and depressive symptoms among Chinese older adults? The role of living alone. Maturitas 168, 78–83. https://doi.org/10.1016/j.maturitas.2022.10.007.

Ying, L., Wu, X., Lin, C., Jiang, L., 2014. Traumatic severity and trait resilience as predictors of posttraumatic stress disorder and depressive symptoms among adolescent survivors of the wenchuan earthquake: E89401. PLoS One 9 (2). https://doi.org/10.1371/journal.pone.0089401.

Yue, Z., Liang, H., Gao, X., Qin, X., Li, H., Xiang, N., Liu, E., 2022. The association between falls and anxiety among elderly chinese individuals: the mediating roles of functional ability and social participation. J. Affect. Disord. 301, 300–306. https://doi.org/10.1016/j.jad.2022.01.070.

Zeng, Y., Feng, Q., Hesketh, T., Christensen, K., Vaupel, J.W., 2017. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest old in China: a cohort study. Lancet 389 (10079), 1619–1629. https://doi.org/10.1016/S0140-6736(17)30548-2 doi.

Zhang, Y., Zhang, L., Zhang, X., Sun, J., Wang, D., Chen, G., 2021. Fall injuries and depressive symptoms among older adults and the mediating effects of social participation - China, 2011-2018. China CDC Week. 3 (40), 837–841. https://doi.org/10.46234/ccdcw2021.207.

Original Paper

Association of Catastrophic Health Expenditure With the Risk of Depression in Chinese Adults: Population-Based Cohort Study

Yaping Wang[1], PhD; Wannian Liang[2,3], PhD; Min Liu[1,4], PhD; Jue Liu[1,4,5,6], PhD

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

[2]Vanke School of Public Health, Tsinghua University, Beijing, China

[3]Institute for Healthy China, Tsinghua University, Beijing, China

[4]Key Laboratory of Epidemiology of Major Diseases, Peking University, Ministry of Education, Beijing, China

[5]Institute for Global Health and Development, Peking University, Beijing, China

[6]Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health, Peking University, Beijing, China

Corresponding Author:
Min Liu, PhD
Department of Epidemiology and Biostatistics
School of Public Health
Peking University
No. 38 Xueyuan Road
Haidian District
Beijing, 100191
China
Phone: 86 10 8 2805146
Fax: 86 10 8 2805146
Email: liumin@bjmu.edu.cn

Abstract

Background: Depression is one of the most common mental illnesses, and it may have a lasting effect on one's whole life. As a form of financial hardship, catastrophic health expenditure (CHE) may be associated with depression. However, current evidence about the relationship between CHE and the risk of depression is insufficient.

Objective: This study aimed to explore the relationship between CHE and the risk of depression among Chinese adults.

Methods: In this study, we used 3 waves of the China Family Panel Studies (CFPS) from 2012, 2016, and 2018. The CFPS are a nationally representative study covering 25 of 31 provinces in Chinese mainland and representing nearly 94.5% of the total population. We selected eligible household heads as participants, divided them into 2 groups by CHE events at baseline (exposed group: with CHE; unexposed group: without CHE), and followed them up. Households with CHE were defined as having out-of-pocket medical expenditures exceeding 40% of the total household nonfood expenditure, and people with depression were identified by the 8-item Centre for Epidemiological Studies Depression Scale (CES-D). We first described the baseline characteristics and used logistical regression to estimate their effects on CHE events. Then, we used Cox proportional hazard models to estimate adjusted hazard ratios and 95% CIs of depression among participants with CHE compared with those without CHE. Finally, we analyzed the subgroup difference in the association between CHE and depression.

Results: Of a total of 13,315 households, 9629 were eligible for analysis. Among them, 6824 (70.9%) were men. The mean age was 50.15 (SD 12.84) years. Only 987 (10.3%) participants had no medical insurance. The prevalence of CHE at baseline was 12.9% (1393/9629). Participants with a higher family economic level (adjusted odds ratio [aOR] 1.15, 95% CI 1.02-1.31) and with the highest socioeconomic development level (aOR 1.18, 95% CI 1.04-1.34) had a higher prevalence of CHE than reference groups. During a median of 71 (IQR 69-72) person-months of follow-up, the depression incidence of participants with CHE (1.41 per 1000 person-months) was higher than those without CHE (0.73 per 1000 person-months). Multivariable models revealed that the adjusted hazard ratio for the incidence of depression in participants with CHE was 1.33 (95% CI 1.08-1.64), and this association appeared to be greater in participants without outpatient services (for interaction, *P*=.048).

Conclusions: CHE was significantly associated with increased risk of depression among Chinese adults. Concentrated work should be done to monitor CHE, and more efforts to ensure financial protection need to be made to prevent depression, especially for people with high health care needs.

(JMIR Public Health Surveill 2023;9:e42469) doi: 10.2196/42469

KEYWORDS

catastrophic health expenditure; depression; universal health coverage; economic burden; socioeconomic status

Introduction

There is a strong bidirectional linkage between health and poverty. To cut off this linkage, the United Nation's Sustainable Development Goals include target 3.8, which aims to achieve universal health coverage (UHC) by 2030 [1]. UHC means that all people can receive the health services they need without experiencing financial hardship [1]. According to the World Health Organization, globally in 2017, almost 1.4 billion people experienced financial hardship due to out-of-pocket (OOP) health payments, among whom nearly 1 billion people were pushed into extreme poverty [2]. Therefore, there is still a long way to go to achieve UHC goals by 2030 considering the additional impact of COVID-19.

One of the 2 essential parts of UHC is providing financial protection for people to pay for health services [3], and the government of China has made great progress on this front. For example, as early as 2009, China reformed a series of health care reforms including three national basic medical insurance programs: (1) Urban Employee Basic Medical Insurance (UEBMI) designed for employed urban residents; (2) the Urban Resident Basic Medical Insurance (URBMI) covering the unemployed, retired, older adults, students, and children in urban areas; and (3) the New Rural Cooperative Medical Scheme (NRCMS) for rural residents [4]. Currently, China's national basic insurance programs cover over 1.35 billion people, about 97% of the total population [5]. Additionally, to eliminate poverty, the Decision on Winning the Battle Against Poverty policy proposed by China in 2015 achieved substantial results in education, primary medical care, and the basic living needs of people living in poverty. In fact, in 2020, the government of China announced that all low-income counties in China had been lifted out of poverty [6]. Benefitting from policies and measures on financial protection of health, the incidence of catastrophic health expenditure (CHE) in China declined from 14.7% in 2010 to 8.7% in 2018 [7]. However, there are still a few people encountering financial hardship, and the impact on mental and physical health caused by CHE is still worth studying.

Depression is a prevalent mental illness globally that contributes to the global burden of diseases as the leading mental cause of mortality for all ages [8]. Once depression appears, it can have a lasting and profound influence on one's whole life. The direct outcome of depression is a poor quality of life [9]. A study conducted on patients with schizophrenia indicated that depression has a strong negative effect on all 8 domains of subjective quality of life [10]. Furthermore, a meta-analysis showed that compared to controls, patients with depression had significant moderate cognitive deficits in executive function, memory, and attention (Cohen *d* effect sizes ranging from −0.34 to −0.65) [11]. Moreover, depression is a negative factor in cardiovascular disease (CVD) incidence, severity, and outcomes [12]. Rajan et al [13] found that depression was associated with CVD incidence (hazard ratio [HR] 1.14, 95% CI 1.05-1.24) and myocardial infarction (HR 1.14, 95% CI 1.05-1.24). Meng et al [14] also found that depression was associated with a higher risk of CVD mortality (HR 1.22, 95% CI 1.04-1.44). Therefore, the prevention and management of depression is a crucial and urgent public health issue.

A series of social, psychological, and biological factors and their complex interactions can play a role in depression occurrence [15,16]. Some study results posit that lower social support, a lower socioeconomic position, and economic difficulties are associated with a higher risk of depression [17,18], and the association is stronger between financial hardship and depression than other socioeconomic variables [19]. CHE, in theory, may have an impact on the mental health of family members because of reduced necessary expenditures. Nevertheless, current studies mostly focus on the likelihood of CHE events among people with depression, not the impact of CHE on depression occurrence [20,21]. In this study, we used 3 waves (2012, 2016, and 2018) of nationally representative data from the China Family Panel Studies (CFPS) to analyze the association of CHE with the risk of depression.

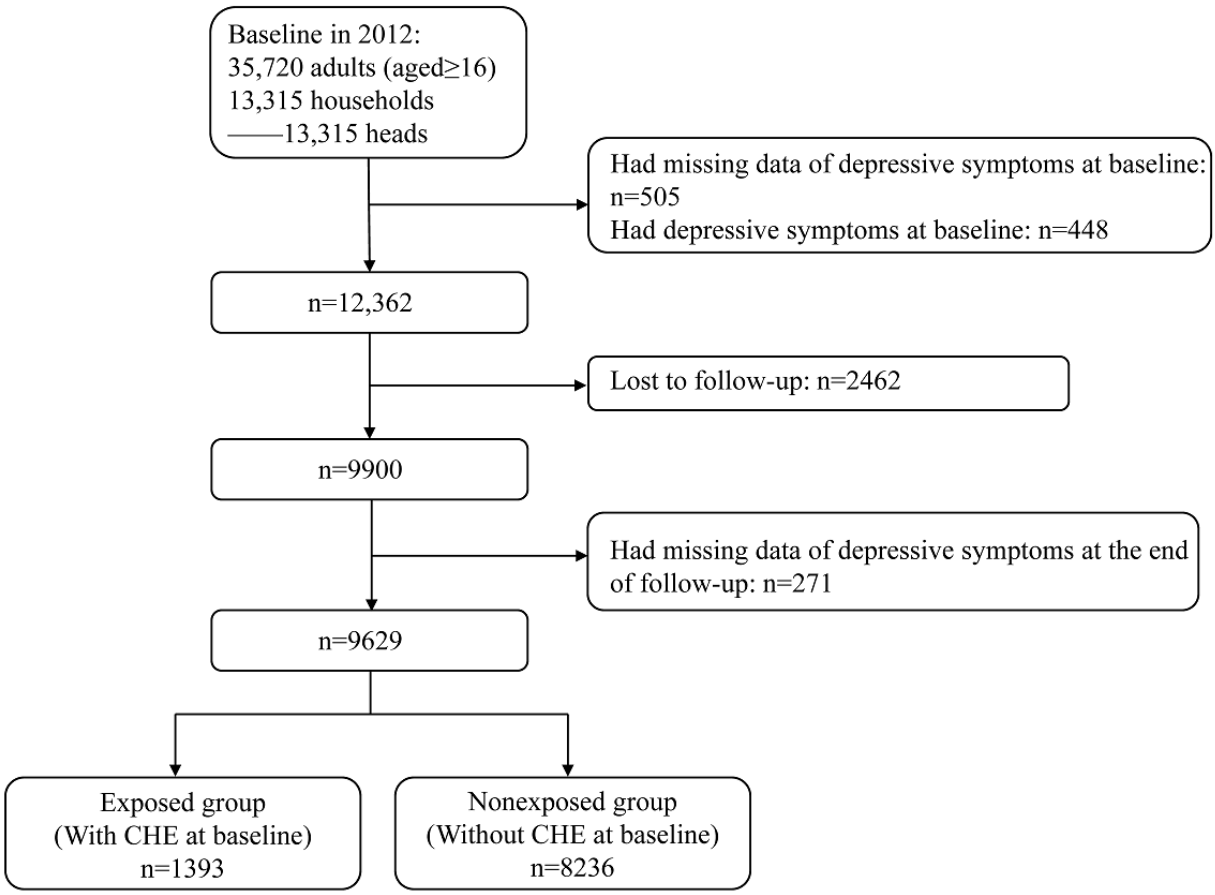
Methods

Study Design and Participants

Data in this study were obtained from CFPS, which is almost a nationally representative longitudinal study covering 25 of 31 provinces/municipalities in Chinese mainland (not including Xinjiang, Tibet, Inner Mongolia, Ningxia autonomous region, and Qinghai and Hainan provinces), representing nearly 94.5% of the total population in the Chinese mainland [22]. The CFPS were implemented by the Institute of Social Science Survey of Peking University to collect individual-, household-, and community-level data every 2 years. A baseline survey was conducted in 2010, and follow-up data from 2012, 2014, 2016, and 2018 were available for download from the official CFPS website [23].

As the information on depression in some of the CFPS waves was deficient, we used data from 2012, 2016, and 2018. The survey in 2012 included 35,720 adults (aged ≥16 years) and 13,315 households with valid interview responses. Individuals who had missing data at baseline (n=505), had depression at baseline (n=448), were lost to follow-up (n=2462), and had missing data of depression at the end of follow-up (n=271) were excluded. Finally, a total of 9629 households (household heads) were included in this study (Figure 1).

Figure 1. Flowchart of participants selected from China Family Panel Studies (CFPS). CHE: catastrophic health expenditure.



Ethics Approval

The CFPS, which involved human participants, were approved by the Biomedical Ethics Review Committee of Peking University (IRB00001052-14010). All written informed consent was provided by participants aged over 15 years or their parents (for those aged 15 years and under). The participants' personal information and privacy were strictly protected by the CFPS according to the rules set by Peking University's Biomedical Ethics Review Committee.

Measurements of Catastrophic Health Expenditure

The measurement of catastrophic health expenditure (CHE) was based on households, where a household member was defined by marriage, blood, or adoptive relationship, as well as an ongoing economic tie [24]. The head of the household was identified as the key decision-maker when the household faced important matters and decisions. Household OOP health payments were measured as the medical expenditure of all family members excluding reimbursed expenses in the past 12 months. The yearly household food expenditure was estimated as the monthly meal expenses multiplied by 12, and the total household expenditure in the past 12 months was calculated as the sum of monthly daily expenditures (food, daily used commodities and necessities, transportation, etc) multiplied by 12 plus yearly special expenditures (electricity, medical care, clothing, etc).

Households who experienced CHE were defined as those with OOP medical expenditures exceeding 40% of the household's capacity to pay (calculated as total household expenditure minus household food expenditure) [25]. CHE prevalence at baseline was measured as the percentage of the number of household heads incurring CHE to total participants. The formula was:

$$CHE\ prevalence = (\frac{1}{N}\sum_{i}^{N} CHE_i) \times 100\%$$

where N is the total number of participants, and CHE_i is 1 when the i_{th} household incurred CHE and 0 otherwise.

Measurements of Depression

Depression was measured by the 8-item Center for Epidemiological Studies Depression Scale (CES-D), which is a shortened version of the original 20-item CES-D including 8 depressive symptoms (Table S1 in Multimedia Appendix 1). CES-D was not designed as a diagnostic tool but is widely used to identify individuals at high risk of depression in general populations and various subpopulations [26]. The score of each item in the 20-item CES-D ranges from 0 to 3 according to the frequency response of depressive symptoms in the past week, with 0 indicating rarely or none of the time (<1 day), 1 indicating some or a little of the time (1-2 days), 2 indicating occasionally or a moderate amount of time (3-4 days), and 3 indicating most or all of the time (5-7 days) [26]. The commonly used 8-item

CES-D asked respondents whether they experienced any of the 8 depressive symptoms "most of time during the last week" [27], which was scored according to the answer (yes=1; no=0). The 8-item CES-D used in CFPS included the same depressive symptoms but was scored by frequency response. As "most or all of the time" in frequency response is equal to the "yes" answer to "most of time last week," we transferred the 4-score level of frequency response to the yes/no response [28]. Hence, the score of 8-item CES-D ranged from 0-8, and a total score ≥3 was deemed to indicate depression [28]. As the 8-item CES-D cannot diagnose depressive disorders, people with an 8-item CES-D score ≥3 in our study were identified as having depressive symptoms [29].

Covariates

Covariates in this study included (1) demographic characteristics: gender (male, female), age group (16-39, 40-49, 50-49, ≥60 years), marital status (married/partnered, other), education (no or some formal education, primary school, middle school, high school and above), and insurance (without any insurance, UEBMI, URBMI, NRCMS, other); (2) health-related characteristics: self-reported health (good, medium, poor), chronic diseases (yes, no), BMI (normal, lower, overweight, obese), outpatient services (yes, no), inpatient services (yes, no), current smoking (yes, no), and drinking (yes, no); and (3) socioeconomic characteristics: residence (urban, rural), family economic level (lowest, lower, higher, highest), family size (1-2, 3-4, ≥5), and socioeconomic development level (lowest, lower, higher, highest).

The family economic level was classified by the quartiles of household annual income, which was converted into 2012 US dollars based on purchasing power parities published by the World Bank [30] (lowest: <$4775.56, lower: $4775.56 to <$9831.46, higher: $9831.46 to <$17173.60, highest: ≥$17173.60). Numerous studies have reported that nighttime light intensity is highly associated with socioeconomic development [31,32]. Therefore, the socioeconomic development level in this study was identified by the quartiles of provinces' mean nighttime light intensity in 2012 (lowest: <0.089, lower: 0.089-0.31, higher: 0.32-0.68, highest: ≥0.69), which was obtained from the Harvard Dataverse [33].

Statistical Analyses

Baseline characteristics of the participants were presented as mean (SD) for continuous variables or frequencies and percentages for categorical variables. The Pearson χ^2 test was used to compare distributions of CHE according to different baseline characteristics. A multivariate logistical regression model was used to analyze the determinants of CHE.

We calculated the incidence rates (number of events divided by accumulated person-month) and used the univariate and multivariate Cox proportional hazard models to estimate the HRs and 95% CIs of depression among participants with CHE compared with those without CHE. First, the model adjusted demographic characteristics in multivariate model 1, including gender, age group, marital status, education, and insurance. Next, the model adjusted health-related factors in model 2 based on model 1, including self-reported health, chronic diseases, BMI, outpatient services, inpatient services, current smoking, and current drinking. Finally, the model adjusted socioeconomic factors in model 3 based on model 2, including residence, family economic level, family size, and socioeconomic development level.

To examine the robustness of our findings, we conducted 3 sensitivity analyses. First, we defined households with CHE as those having OOP medical expenditure exceeding 25% of the total household expenditure according to the World Bank. [25] Second, we transferred the categorical variables of age group and family economic level into continuous variables and conducted the same analysis in the final model. Third, we changed the variable nighttime light intensity into the gross regional product (GRP, divided into 4 groups based on the quartiles of 2012 per capita GRP [34]) to indicate the socioeconomic development level in the final model.

Finally, the analysis was stratified by gender, age group, insurance, chronic diseases, self-reported health, outpatient services, inpatient services, residence, socioeconomic development level, and family economic level, and a multiplicative interaction term was included in the final model with the stratification variable removed for subgroup analysis.

All data were analyzed in R software (version 4.2.1; R Foundation for Statistical Computing). A 2-sided *P* value <.05 was considered to be significant.

Results

Baseline Characteristics

A total of 9629 participants were enrolled in this study, of which 6824 (70.9%) were male. The mean age was 50.15 (SD 12.84) years. Among the participants, 89.5% (8622/9629) were married or partnered. Only 987 (10.3%) participants had no medical insurance, and 1341 (13.9%), 603 (6.3%), and 6317 (65.6%) participants had UEBMI, URBMI, and NRCMS, respectively (Table 1).

Table 1. Distribution of CHE[a] by baseline characteristics.

Characteristics	Total (N=9629)	Nonexposed group[b] (n=8236), n (%)	Exposed group[c] (n=1393), n (%)	Chi-square (*df*)	*P* value
Demographic characteristics					
Gender				3.6 (1)	.06
Male	6824	5867 (86)	957 (14)		
Female	2805	2369 (84.5)	436 (15.5)		
Age group (years)				346.3 (3)	<.001
16-39	1931	1753 (90.8)	178 (9.2)		
40-49	3016	2750 (91.2)	266 (8.8)		
50-59	2320	1969 (84.9)	351 (15.1)		
≥60	2362	1764 (74.7)	598 (25.3)		
Marital status				9.1 (1)	.003
Married/partnered	8622	7407 (85.9)	1215 (14.1)		
Other	1007	829 (82.3)	178 (17.7)		
Education				97.3 (3)	<.001
No or some formal education	2506	2005 (80)	501 (20)		
Primary school	2322	1980 (85.3)	342 (14.7)		
Middle school	2789	2475 (88.7)	314 (11.3)		
High school and above	2012	1776 (88.3)	236 (11.7)		
Insurance				5.5 (4)	.24
None	987	856 (86.7)	131 (13.3)		
UEBMI[d]	1341	1169 (87.2)	172 (12.8)		
URBMI[e]	603	511 (84.7)	92 (15.3)		
NRCMS[f]	6317	5378 (85.1)	939 (14.9)		
Other	381	322 (84.5)	59 (15.5)		
Health-related characteristics					
Self-reported health				245.5 (2)	<.001
Good	5874	5241 (89.2)	633 (10.8)		
Medium	2001	1692 (84.6)	309 (15.4)		
Poor	1754	1303 (74.3)	451 (25.7)		
Currently smoking				12.6 (1)	<.001
No	5355	4519 (84.4)	836 (15.6)		
Yes	4274	3717 (87)	557 (13)		
Drinking				16.4 (1)	<.001
No	7282	6168 (84.7)	1114 (15.3)		
Yes	2347	2068 (88.1)	279 (11.9)		
Chronic diseases				104 (1)	<.001
No	8268	7195 (87)	1073 (13)		
Yes	1361	1041 (76.5)	320 (23.5)		
BMI				25.8 (3)	<.001
Normal	5596	4776 (85.3)	820 (14.7)		
Lower	711	569 (80)	142 (20)		

Characteristics	Total (N=9629)	Nonexposed group[b] (n=8236), n (%)	Exposed group[c] (n=1393), n (%)	Chi-square (*df*)	*P* value
Overweight	2658	2301 (86.6)	357 (13.4)		
Obese	664	590 (88.9)	74 (11.1)		
Outpatient services				101.8 (1)	<.001
No	7577	6624 (87.4)	953 (12.6)		
Yes	2052	1612 (78.6)	440 (21.4)		
Inpatient services				164.4 (1)	<.001
No	8747	7644 (87.4)	1103 (12.6)		
Yes	882	592 (67.1)	290 (32.9)		
Socioeconomic characteristics					
Residence				11.7 (1)	<.001
Urban	4315	3750 (86.9)	565 (13.1)		
Rural	5314	4486 (84.4)	828 (15.6)		
Family economic level				55.8 (3)	<.001
Lowest	2988	2439 (81.6)	549 (18.4)		
Lower	2336	2021 (86.5)	315 (13.5)		
Higher	2366	2084 (88.1)	282 (11.9)		
Highest	1939	1692 (87.3)	247 (12.7)		
Family size				151.5 (2)	<.001
1-2	2158	1677 (77.7)	481 (22.3)		
3-4	4302	3833 (89.1)	469 (10.9)		
≥5	3169	2726 (86)	443 (14)		
Socioeconomic development level				6.4 (3)	.09
Lowest	2491	2158 (86.6)	333 (13.4)		
Lower	1492	1258 (84.3)	234 (15.7)		
Higher	3726	3199 (85.9)	527 (14.1)		
Highest	1920	1621 (84.4)	299 (15.6)		

[a]CHE: catastrophic health expenditure.

[b]People without CHE.

[c]People with CHE.

[d]UEBMI: Urban Employee Basic Medical Insurance.

[e]URBMI: Urban Resident Basic Medical Insurance.

[f]NRCMS: New Rural Cooperative Medical Scheme.

CHE Prevalence

At baseline, the prevalence of CHE was 12.9% (1393/9629) among the participants. Except for gender, insurance, and socioeconomic development level, the distribution of baseline characteristics was significantly different between households with CHE and those without CHE (Table 1). The logistical regression analysis revealed that participants with poor (adjusted odds ratio [aOR] 1.64, 95% CI 1.39-1.94) and medium self-reported health (aOR 1.17, 95% CI 1-1.37), chronic diseases (aOR 1.30, 95% CI 1.11-1.52), outpatient services (aOR 1.16, 95% CI 1-1.35) and inpatient services (aOR 2.40, 95% CI 2.03-2.84), rural residence (aOR 1.21, 95% CI 1.05-1.39), family size ≥5 people (aOR 1.22, 95% CI 1.09-1.36), higher family economic level (aOR 1.15, 95% CI 1.02-1.31), and the highest socioeconomic development level (aOR 1.18, 95% CI 1.04-1.34) had a higher prevalence of CHE than reference groups (Table S2 in Multimedia Appendix 1).

Risk of Depression

During a median of 71 (interquartile range: 69-72) person-months of follow-up, 532 (5.53%) of 9629 participants developed depression, of which 403 (75.75%) cases had no CHE and 129 (24.25%) cases had CHE. The incidence rate of depression among participants without and with CHE was 0.73 and 1.41 per 1000 person-months, respectively (Table 2). In the unadjusted analysis, participants who had CHE at baseline were associated with a 99% higher risk of depression (crude HR 1.99,

95% CI 1.63-2.42; Table 2). All multivariable–adjusted analyses showed a significant association of CHE with the risk of depression (Table S3 in Multimedia Appendix 1). In the fully adjusted model, participants with CHE had a 33% increased risk of developing depression compared to those without CHE (adjusted HR [aHR] 1.33, 95% CI 1.08-1.64; Table 2).

In the sensitivity analyses, the association between CHE and risk of depression was stable when we (1) defined CHE as OOP medical expenditure exceeding 25% of the total household expenditure, (2) transferred the categorical variables age group and family economic level to continuous variables, and (3) changed the variable nightlight time intensity into GRP to indicate socioeconomic development level in the final model (Table S4 in Multimedia Appendix 1).

Table 2. Association of CHE[a] with risk of depression in the univariate and multivariate Cox proportional hazard models.

Characteristics	Outcome (N=532), n (%)	Incidence (per 1000 person-months)	Univariate model		Multivariate model[b]	
			cHR[c] (95% CI)	*P* value	aHR[d] (95% CI)	*P* value
CHE						
Without	403 (75.8)	0.73	Reference		Reference	
With	129 (24.2)	1.41	1.99 (1.63-2.42)	<.001	1.33 (1.08-1.64)	.008
Gender						
Male	318 (59.8)	0.70	Reference		Reference	
Female	214 (40.2)	1.15	1.69 (1.42-2.01)	<.001	1.51 (1.19-1.91)	.001
Age group (years)						
16-39	50 (9.4)	0.39	Reference		Reference	
40-49	144 (27.1)	0.71	2.39 (1.93-2.97)	<.001	1.61 (1.27-2.06)	<.001
50-59	154 (28.9)	0.99	0.84 (0.69-1.02)	.08	0.78 (0.64-0.96)	.02
≥60	184 (34.6)	1.20	1.03 (0.87-1.22)	.70	0.97 (0.82-1.15)	.74
Marital status						
Married/partnered	433 (81.4)	0.75	Reference		Reference	
Other	99 (18.6)	1.54	2.22 (1.78-2.76)	<.001	1.34 (1.05-1.70)	.02
Education						
No or some formal education	255 (47.9)	1.56	Reference		Reference	
Primary school	121 (22.7)	0.78	0.36 (0.29-0.44)	<.001	0.63 (0.50-0.80)	<.001
Middle school	100 (18.8)	0.53	1.27 (1.04-1.54)	.02	1.18 (0.96-1.44)	.12
High school and above	56 (10.5)	0.42	0.95 (0.79-1.15)	.61	1.01 (0.83-1.22)	.95
Insurance						
None	71 (13.3)	1.11	Reference		Reference	
UEBMI[e]	35 (6.6)	0.39	0.35 (0.23-0.52)	<.001	0.52 (0.34-0.79)	.002
URBMI[f]	23 (4.3)	0.57	0.51 (0.32-0.81)	.004	0.53 (0.33-0.85)	.009
NRCMS[g]	394 (74.1)	0.93	0.82 (0.64-1.06)	.13	0.68 (0.52-0.89)	.004
Other	9 (1.7)	0.35	0.30 (0.15-0.61)	.001	0.43 (0.21-0.88)	.02
Self-reported health						
Good	201 (37.8)	0.51	Reference		Reference	
Medium	110 (20.7)	0.83	1.61 (1.28-2.03)	<.001	1.32 (1.04-1.68)	.02
Poor	221 (41.5)	1.93	3.89 (3.21-4.71)	<.001	2.10 (1.67-2.64)	<.001
Currently smoking						
No	314 (59.0)	0.88	Reference		Reference	
Yes	218 (41.0)	0.77	0.86 (0.72-1.02)	.08	1.14 (0.92-1.42)	.24
Drinking						
No	435 (81.8)	0.90	Reference		Reference	
Yes	97 (18.2)	0.62	0.67 (0.54-0.84)	.001	0.88 (0.69-1.11)	.29
Chronic diseases						
No	419 (78.8)	0.76	Reference		Reference	
Yes	113 (21.2)	1.25	1.62 (1.31-1.99)	<.001	1.17 (0.94-1.46)	.15

Characteristics	Outcome (N=532), n (%)	Incidence (per 1000 person-months)	Univariate model		Multivariate model[b]	
			cHR[c] (95% CI)	*P* value	aHR[d] (95% CI)	*P* value
BMI						
Normal	311 (58.5)	0.84	Reference		Reference	
Lower	65 (12.2)	1.41	1.75 (1.34-2.28)	<.001	1.13 (0.86-1.48)	.40
Overweight	118 (22.2)	0.66	0.79 (0.64-0.98)	.03	0.92 (0.74-1.13)	.42
Obese	38 (7.1)	0.85	1.05 (0.75-1.47)	.78	1.13 (0.80-1.58)	.50
Outpatient services						
No	326 (61.3)	0.64	Reference		Reference	
Yes	206 (38.7)	1.52	2.36 (1.98-2.81)	<.001	1.41 (1.16-1.72)	.001
Inpatient services						
No	477 (89.7)	0.82	Reference		Reference	
Yes	55 (10.3)	0.95	1.18 (0.89-1.56)	.24	0.70 (0.52-0.94)	.02
Residence						
Urban	173 (32.5)	0.61	Reference		Reference	
Rural	359 (67.5)	1.01	1.65 (1.37-1.98)	<.001	1.23 (1-1.52)	.06
Family economic level						
Lowest	240 (45.1)	1.21	Reference		Reference	
Lower	109 (20.5)	0.70	0.55 (0.46-0.67)	<.001	0.75 (0.61-0.93)	.008
Higher	116 (21.8)	0.74	1.11 (0.92-1.34)	.28	1.03 (0.85-1.24)	.79
Highest	67 (12.6)	0.52	0.79 (0.66-0.96)	.02	0.83 (0.69-1)	.049
Family size						
1-2	165 (31)	1.17	Reference		Reference	
3-4	200 (37.6)	0.69	0.72 (0.62-0.84)	<.001	0.96 (0.81-1.15)	.69
≥5	167 (31.4)	0.79	1.34 (1.16-1.55)	<.001	1.02 (0.88-1.19)	.78
Socioeconomic development level						
Lowest	174 (32.7)	1.05	Reference		Reference	
Lower	92 (17.3)	0.93	0.69 (0.57-0.83)	.001	1.01 (0.78-1.32)	.92
Higher	186 (35.0)	0.75	0.92 (0.77-1.11)	.39	0.86 (0.69-1.06)	.16
Highest	80 (15.0)	0.63	1.01 (0.84-1.20)	.94	0.73 (0.55-0.97)	.03

[a]CHE: catastrophic health expenditure.

[b]Adjusted for demographic characteristics (gender, age group, education, marital status, and insurance), health-related characteristics (self-reported health, smoking status, drinking, chronic disease, BMI, and outpatient and inpatient services), and socioeconomic characteristics (residence, family economic level, family size, and socioeconomic development level).

[c]cHR: crude hazard ratio.

[d]aHR: adjusted hazard ratio.

[e]UEBMI: Urban Employee Basic Medical Insurance.

[f]URBMI: Urban Resident Basic Medical Insurance.

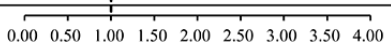
[g]NRCMS: New Rural Cooperative Medical Scheme.

Subgroup Analyses

In the subgroup analyses, the association of CHE with the risk of depression did not appear to be modified by most selected baseline characteristics, such as gender, age group, insurance, or chronic diseases (Figure 2). However, the association between CHE and the risk of depression was found to be somewhat greater among participants without outpatient services (for interaction, *P*=.048). The risk of depression was higher among participants with CHE who did not have outpatient services (aHR 1.57, 95% CI 1.10-2.25) than those who had outpatient services (aHR 1.06, 95% CI 0.76-1.48).

Figure 2. Association of catastrophic health expenditure (CHE) with risk of depression in univariate and multivariate Cox proportional hazard models. Superscripted a refers to people with CHE. Superscripted b refers to people without CHE. aHR: adjusted hazard ratio; NRCMS: New Rural Cooperative Medical Scheme; UEBMI: Urban Employee Basic Medical Insurance; URBMI: Urban Resident Basic Medical Insurance.

Subgroup	Total events (N, %)	Exposed group[a] (Events/Objects)	Nonexposed group[b] (Events/Objects)	aHR (95% CI)	*P* for interaction
All	532 (100.0)	129/1393	403/8236	1.33 (1.08-1.64)	
Gender					.75
Male	318 (59.8)	71/957	247/5867	1.32 (1.00-1.75)	
Female	214 (40.2)	58/436	156/2369	1.41 (1.02-1.95)	
Age group					.93
16-39	50 (9.4)	6/178	44/1753	1.44 (0.59-3.48)	
40-49	144 (27.1)	21/266	123/2750	1.34 (0.82-2.20)	
50-59	154 (28.9)	38/351	116/1969	1.54 (1.05-2.27)	
⩾60	184 (34.6)	64/598	120/1764	1.23 (0.89-1.69)	
Insurance					.29
None	71 (13.3)	12/131	59/856	1.10 (0.56-2.15)	
UEBMI	35 (6.6)	10/172	25/1169	3.13 (1.39-7.03)	
URBMI	23 (4.3)	8/92	15/511	1.77 (0.63-4.97)	
NRCMS	394 (74.1)	97/939	297/5378	1.27 (1.00-1.62)	
Other	9 (1.7)	2/59	7/322	0.78 (0.07-8.21)	
Chronic diseases					.70
No	419 (78.8)	90/1073	329/7195	1.28 (1.00-1.63)	
Yes	113 (21.2)	39/320	74/1041	1.44 (0.93-2.22)	
Outpatient services					.048
No	326 (61.3)	74/953	252/6624	1.57 (1.19-2.06)	
Yes	206 (38.7)	55/440	151/1612	1.06 (0.76-1.48)	
Inpatient services					.53
No	477 (89.7)	101/1103	376/7644	1.30 (1.03-1.63)	
Yes	55 (10.3)	28/290	27/592	1.48 (0.81-2.69)	
Self-reported health					.41
Good	201 (37.8)	36/633	165/5241	1.53 (1.05-2.22)	
Medium	110 (20.7)	23/309	87/1692	1.28 (0.80-2.06)	
Poor	221 (41.5)	70/451	151/1303	1.27 (0.94-1.73)	
Residence					.49
Urban	173 (32.5)	33/565	140/3750	1.25 (0.84-1.85)	
Rural	359 (67.5)	96/828	263/4486	1.36 (1.06-1.75)	
Family economic level					.07
Lowest	240 (45.1)	73/549	167/2439	1.28 (0.95-1.72)	
Lower	109 (20.5)	28/315	81/2021	1.71 (1.08-2.70)	
Higher	116 (21.8)	12/282	104/2084	0.80 (0.43-1.49)	
Highest	67 (12.6)	16/247	51/1692	1.90 (1.04-3.47)	
Socioeconomic development level					.99
Lowest	174 (32.7)	38/333	136/2158	1.29 (0.88-1.89)	
Lower	92 (17.3)	24/234	68/1258	1.28 (0.76-2.14)	
Higher	186 (35.0)	46/527	140/3199	1.57 (1.10-2.25)	
Highest	80 (15.0)	21/299	59/1621	1.20 (0.70-2.08)	



Discussion

CHE can introduce great financial pressure to a household, as household members may have to borrow money, use savings, and even sell assets [35], which may lead to a longtime negative effect on their mental health [36]. To our knowledge, this is the first cohort study on the association between CHE and the risk of depression. We found that health-related characteristics (chronic diseases, self-reported health, and outpatient and inpatient services) and socioeconomic characteristics (residence, socioeconomic development level, family size, and family economic level) had an impact on CHE. We also found that prior CHE events were strongly associated with depression experienced by household heads and that there was an interaction between CHE and outpatient services for depression.

Our results revealed that people with chronic diseases, medium and poor self-reported health, outpatient services and inpatient services, rural, family size ≥5 people, and the highest socioeconomic development level had a higher prevalence of CHE, and these findings are consistent with those of other studies [7,37]. Poor health directly leads to a higher frequency of accessing health services, further leads to a heavy financial burden of medical expenditure, and finally, leads to CHE. Nighttime light intensity is an indicator of socioeconomic development, and areas with higher economic growth may have access to diagnostic technology for most illnesses. In addition, people in areas undergoing rapid socioeconomic development may pay more attention to the early detection of diseases and other preventive health care services [38]. Consequently, health-related characteristics and socioeconomic characteristics were the main factors behind CHE.

The results of our study revealed that people with CHE had a 33% higher risk of developing depression (aHR 1.33, 95% CI 1.08-1.64) than people without CHE. Socioeconomic variables, especially financial hardship, have a great influence on depression [39]. As a form of financial hardship, CHE plays a role in depression incidence. Once a household incurred CHE, family members had to take various measures to cope with financial loss, which could take a long time to recover [35]. Given this, household heads had a higher risk of developing depression with CHE than those without CHE under financial pressure.

As CHE could trigger depression, which may further lead to adverse outcomes [40,41], it is crucial to eliminate CHE to

prevent depression. The first step to eliminating CHE is based on reducing OOP health expenditures. Because poor health is associated with an increased need and use of outpatient and inpatient care, and this negative correlation can be intensified by the number of chronic diseases [37], our results suggest that relevant departments should continue to expand health promotion, especially for chronic diseases. Additionally, governments should provide sufficient and targeted financial protection, including medical insurance and poverty subsidies, to reduce OOP payments and mitigate impoverishment caused by poor health. Most studies on CHE are interested in subpopulations that have specialized diseases like cancers, HIV, and tuberculosis [42-44]. The common features of these diseases are long duration, which may even coexist with people and cause a higher risk of CHE. To decrease individual OOP health expenditure, inclined policies and measures should be introduced for these diseases. For example, to relieve the financial burden faced by people living with chronic diseases, China launched catastrophic medical insurance (critical illness insurance) in 2012 and implemented it nationwide in 2016 after city-based testing, with the aim of reimbursing patients whose OOP health expenditure exceeded a predetermined basic medical insurance level [45]. In addition, for eligible people facing unaffordable medical expenses, China provided medical assistance and even offered treatment for free or at a reduced price for some priority diseases [46]. Of course, these policies are mostly aimed at financially vulnerable people. However, our study found that people with a higher family economic level and the highest socioeconomic development level were more likely to occur CHE, which may be explained by the concentrated distribution and utilization of health services. [47] In view of this, more financial protection policies should be implemented for the whole population regardless of their economic capacity. Additionally, as the current health care delivery in China is still fragmented and treatment-based [48], with increasing incidence of age-related diseases, integrated establishment and improvement of primary health services should be intensified in health-related policies.

To cut off the linkage between CHE and depression, the second point is to halt the progression to depression. As the evidence shows in studies on financial hardship and depression, not all people experiencing financial hardship will develop depression [39]. These differing responses to CHE may be explained by various socioeconomic and psychological variables in the stress process model, such as social support, self-esteem, personal agency, and personal ability to manage difficulties [39,49,50]. Therefore, measures to prevent and regulate depression should be extended widely. Additionally, timely financial assistance should be implemented for vulnerable populations, such as people with chronic diseases and households under the minimum living guarantee.

In this study, we found that the risk of depression was higher among participants with CHE who did not have outpatient services (aHR 1.57, 95% CI 1.10-2.25) than those who had outpatient services (aHR 1.06, 95% CI 0.76-1.48). In CFPS, the variable “outpatient services” was measured by whether participants had used outpatient services in the past 2 weeks when they felt uncomfortable. Those without outpatient services included people who did not feel uncomfortable and who felt uncomfortable but did not use health services. The latter group may have a lower socioeconomic status. Moreover, heads of households with lower social support, a lower socioeconomic position, and economic difficulties were more likely to experience depression due to life pressures [17,18]. Consequently, not accessing outpatient services intensified the association between CHE and the risk of depression.

There are several limitations to our study. First, because the 8-item CES-D is not equal to clinical diagnosis, our study could only estimate the association between CHE and the risk of depressive symptoms, which calls for future research on CHE and depressive disorders. Second, due to the limitation of the original data, households with CHE did not include those facing extreme poverty who could not seek health services and whose household medical expenditure was zero. Third, expenditure data for calculating CHE were mainly based on the participants’ memory, which may be prone to recall bias.

In conclusion, people with poor health and a higher socioeconomic position had a higher prevalence of CHE and CHE was significantly associated with a higher risk of depression. To prevent depression induced by CHE, concentrated work should be made to monitor CHE, and more efforts to ensure financial protection need to be introduced and strengthened, especially for people with higher health care needs.

Acknowledgments

We appreciate the China Family Panel Studies (CFPS) for conducting the total survey and providing data. We also thank all the interviewers, respondents, and volunteers in the survey. Moreover, we acknowledge the helpful advice of the editors and reviewers. This study was supported by the National Natural Science Foundation of China (72122001, 71934002).

Data Availability

Data from all are available to download from the official China Family Panel Studies (CFPS) website.

Authors' Contributions

JL conceptualized and designed the study. YW carried out the literature search, data analysis, and interpretation; compiled the tables and figures; and drafted the manuscript. WL, ML, and JL wrote, reviewed, and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional tables.
[DOCX File , 41 KB-Multimedia Appendix 1]

References

1. Universal health coverage. World Health Organization. URL: https://www.who.int/health-topics/universal-health-coverage#tab=tab_ [accessed 2022-08-06]
2. Tracking universal health coverage: global monitoring report. World Health Organization and World Bank. URL: https://cdn.who.int/media/docs/default-source/world-health-data-platform/events/tracking-universal-health-coverage-2021-global-monitoring-report_uhc-day.pdf?sfvrsn=fd5c65c6_5&download=true [accessed 2022-08-06]
3. Universal health coverage (UHC): fact sheets. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc) [accessed 2022-08-06]
4. Sun J, Lyu S. The effect of medical insurance on catastrophic health expenditure: evidence from China. Cost Eff Resour Alloc 2020 Feb 27;18:10 [FREE Full text] [doi: 10.1186/s12962-020-00206-y] [Medline: 32127784]
5. Yin W. To accelerate the development of people's health and promote the innovation of medical insurance system. People's Daily. URL: http://health.people.com.cn/n1/2016/1010/c398004-28764621.html [accessed 2022-08-06]
6. Poverty alleviation: China's experience and contribution. The State Council Information Office of the People's Republic of China. URL: http://www.scio.gov.cn/ztk/dtzt/44689/45216/index.htm [accessed 2022-08-07]
7. Liu C, Liu Z, Nicholas S, Wang J. Trends and determinants of catastrophic health expenditure in China 2010-2018: a national panel data analysis. BMC Health Serv Res 2021 May 29;21(1):526 [FREE Full text] [doi: 10.1186/s12913-021-06533-x] [Medline: 34051762]
8. GBD 2019 DiseasesInjuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
9. Ruo B, Rumsfeld JS, Hlatky MA, Liu H, Browner WS, Whooley MA. Depressive symptoms and health-related quality of life: the Heart and Soul Study. JAMA 2003 Jul 9;290(2):215-221 [FREE Full text] [doi: 10.1001/jama.290.2.215] [Medline: 12851276]
10. Reine G, Lançon C, Di Tucci S, Sapin C, Auquier P. Depression and subjective quality of life in chronic phase schizophrenic patients. Acta Psychiatr Scand 2003 Oct;108(4):297-303 [doi: 10.1034/j.1600-0447.2003.00132.x] [Medline: 12956831]
11. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med 2014 Jul;44(10):2029-2040 [doi: 10.1017/S0033291713002535] [Medline: 24168753]
12. Elderon L, Whooley MA. Depression and cardiovascular disease. Prog Cardiovasc Dis 2013 May;55(6):511-523 [doi: 10.1016/j.pcad.2013.03.010] [Medline: 23621961]
13. Rajan S, McKee M, Rangarajan S, Bangdiwala S, Rosengren A, Gupta R, Prospective Urban Rural Epidemiology (PURE) Study Investigators. Association of symptoms of depression with cardiovascular disease and mortality in low-, middle-, and high-income countries. JAMA Psychiatry 2020 Oct 01;77(10):1052-1063 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1351] [Medline: 32520341]
14. Meng R, Yu C, Liu N, He M, Lv J, Guo Y, China Kadoorie Biobank Collaborative Group. association of depression with all-cause and cardiovascular disease mortality among adults in China. JAMA Netw Open 2020 Feb 05;3(2):e1921043 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21043] [Medline: 32049295]
15. Heim C, Binder EB. Current research trends in early life stress and depression: review of human studies on sensitive periods, gene-environment interactions, and epigenetics. Exp Neurol 2012 Jan;233(1):102-111 [doi: 10.1016/j.expneurol.2011.10.032] [Medline: 22101006]
16. Malhi GS, Mann JJ. Depression. Lancet 2018 Nov;392(10161):2299-2312 [doi: 10.1016/s0140-6736(18)31948-2]
17. Langer, Crockett MA, Bravo-Contreras M, Carrillo-Naipayan C, Chaura-Marió M, Gómez-Curumilla B, et al. Social and Economic Factors Associated With Subthreshold and Major Depressive Episode in University Students During the COVID-19 Pandemic. Front Public Health 2022;10:893483 [FREE Full text] [doi: 10.3389/fpubh.2022.893483] [Medline: 35664111]
18. Wei D, Au Yeung SL, He J, Xiao W, Lu J, Tu S, et al. The role of social support in family socio-economic disparities in depressive symptoms during early pregnancy: Evidence from a Chinese birth cohort. J Affect Disord 2018 Oct 01;238:418-423 [doi: 10.1016/j.jad.2018.06.014] [Medline: 29913379]
19. Butterworth P, Olesen SC, Leach LS. The role of hardship in the association between socio-economic position and depression. Aust N Z J Psychiatry 2012 Apr 05;46(4):364-373 [doi: 10.1177/0004867411433215] [Medline: 22508596]

20. Hailemichael Y, Hanlon C, Tirfessa K, Docrat S, Alem A, Medhin G, et al. Catastrophic health expenditure and impoverishment in households of persons with depression: a cross-sectional, comparative study in rural Ethiopia. BMC Public Health 2019 Jul 11;19(1):930 [FREE Full text] [doi: 10.1186/s12889-019-7239-6] [Medline: 31296207]
21. Yan R, Li L, Duan X, Zhao J. Association of depressive symptoms with health service use and catastrophic health expenditure among middle-aged and older Chinese adults: analysis of population-based panel data. J Am Med Dir Assoc 2023 May;24(5):664-671.e7 [doi: 10.1016/j.jamda.2022.11.018] [Medline: 36574953]
22. Xie Y, Lu P. The sampling design of the China Family Panel Studies (CFPS). Chin J Sociol 2015 Dec;1(4):471-484 [FREE Full text] [doi: 10.1177/2057150X15614535] [Medline: 29854418]
23. China Family Panel Studies. URL: http://www.isss.pku.edu.cn/cfps/ [accessed 2022-08-06]
24. Xie Y, Hu J. An Introduction to the China Family Panel Studies (CFPS). Chinese Sociological Review Sep 2014;47(1):3-29 [doi: 10.2753/CSA2162-0555470101.2014.11082908]
25. Cylus J, Thomson S, Evetovits T. Catastrophic health spending in Europe: equity and policy implications of different calculation methods. Bull World Health Organ 2018 Sep 01;96(9):599-609 [FREE Full text] [doi: 10.2471/BLT.18.209031] [Medline: 30262941]
26. Smarr KL, Keefer AL. Measures of depression and depressive symptoms. Arthritis Care Res (Hoboken) 2020 Oct;72 Suppl 10:608-629 [doi: 10.1002/acr.24191] [Medline: 33091258]
27. Turvey CL, Wallace RB, Herzog R. A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. Int Psychogeriatr 1999 Jun;11(2):139-148 [doi: 10.1017/s1041610299005694] [Medline: 11475428]
28. Wallace R, Herzog A, Ofstedal M, Steffick D, Fonda S, Langa K. Documentation of affective functioning measures in the health and retirement study. HRS Online. URL: https://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf [accessed 2022-08-05]
29. Litwin H. The association between social network relationships and depressive symptoms among older Americans: what matters most? Int Psychogeriatr 2011 Aug;23(6):930-940 [doi: 10.1017/S1041610211000251] [Medline: 21356159]
30. PPP conversion factor, GDP (LCU per international $) - China. World Bank. URL: https://data.worldbank.org/indicator/PA.NUS.PPP?locations=CN&most_recent_year_desc=true&view=chart [accessed 2023-05-20]
31. Beyer RCM, Franco-Bedoya S, Galdo V. Examining the economic impact of COVID-19 in India through daily electricity consumption and nighttime light intensity. World Dev 2021 Apr;140:105287 [FREE Full text] [doi: 10.1016/j.worlddev.2020.105287] [Medline: 34305264]
32. Zhang G, Guo X, Li D, Jiang B. Evaluating the potential of LJ1-01 nighttime light data for modeling socio-economic parameters. Sensors (Basel) 2019 Mar 26;19(6) [FREE Full text] [doi: 10.3390/s19061465] [Medline: 30917491]
33. Chen Z, Yu B, Yang C, Zhou Y, Yao S, Qian X. An extended time-series (2000-2018) of global NPP-VIIRS-like nighttime light data. Harvard Dataverse. 2020. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YGIVCD [accessed 2022-08-04]
34. National Bureau of Statistics of the People's Republic of China. China Statistical Yearbook 2012. Beijing, China: China Statistical Publishing House; 2013:57
35. Kasahun GG, Gebretekle GB, Hailemichael Y, Woldemariam AA, Fenta TG. Catastrophic healthcare expenditure and coping strategies among patients attending cancer treatment services in Addis Ababa, Ethiopia. BMC Public Health 2020 Jun 22;20(1):984 [FREE Full text] [doi: 10.1186/s12889-020-09137-y] [Medline: 32571275]
36. Kiely KM, Leach LS, Olesen SC, Butterworth P. How financial hardship is associated with the onset of mental health problems over time. Soc Psychiatry Psychiatr Epidemiol 2015 Jun 17;50(6):909-918 [doi: 10.1007/s00127-015-1027-0] [Medline: 25683473]
37. Zhao Y, Atun R, Oldenburg B, McPake B, Tang S, Mercer SW, et al. Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data. Lancet Glob Health 2020 Jun;8(6):e840-e849 [doi: 10.1016/s2214-109x(20)30127-3]
38. Tang K, Zhang Y, Wang H, Tan SH, Bai L, Liu Y. Regional economic development, household income, gender and hypertension: evidence from half a million Chinese. BMC Public Health 2020 Jun 10;20(1):901 [FREE Full text] [doi: 10.1186/s12889-020-09002-y] [Medline: 32522178]
39. Frankham C, Richardson T, Maguire N. Psychological factors associated with financial hardship and mental health: A systematic review. Clin Psychol Rev 2020 Apr;77:101832 [doi: 10.1016/j.cpr.2020.101832] [Medline: 32088498]
40. Yoshikawa H, Aber JL, Beardslee WR. The effects of poverty on the mental, emotional, and behavioral health of children and youth: implications for prevention. Am Psychol 2012;67(4):272-284 [doi: 10.1037/a0028015] [Medline: 22583341]
41. Ridley M, Rao G, Schilbach F, Patel V. Poverty, depression, and anxiety: causal evidence and mechanisms. Science 2020 Dec 11;370(6522) [doi: 10.1126/science.aay0214] [Medline: 33303583]
42. Sun C, Shi J, Fu W, Zhang X, Liu G, Chen W, et al. Catastrophic health expenditure and its determinants among households with breast cancer patients in China: a multicentre, cross-sectional survey. Front Public Health 2021;9:704700 [FREE Full text] [doi: 10.3389/fpubh.2021.704700] [Medline: 34291034]

43. Jung H, Kwon YD, Noh J. Financial burden of catastrophic health expenditure on households with chronic diseases: financial ratio analysis. BMC Health Serv Res 2022 Apr 27;22(1):568 [FREE Full text] [doi: 10.1186/s12913-022-07922-6] [Medline: 35477404]
44. Assebe LF, Negussie EK, Jbaily A, Tolla MTT, Johansson KA. Financial burden of HIV and TB among patients in Ethiopia: a cross-sectional survey. BMJ Open 2020 Jun 01;10(6):e036892 [FREE Full text] [doi: 10.1136/bmjopen-2020-036892] [Medline: 32487582]
45. Li H, Jiang L. Catastrophic medical insurance in China. Lancet 2017 Oct 14;390(10104):1724-1725 [doi: 10.1016/S0140-6736(17)32603-X] [Medline: 29047433]
46. Cao W, Hsieh E, Li T. Optimizing treatment for adults with HIV/AIDS in China: successes over two decades and remaining challenges. Curr HIV/AIDS Rep 2020 Feb;17(1):26-34 [FREE Full text] [doi: 10.1007/s11904-019-00478-x] [Medline: 31939111]
47. Ni X, Li Z, Li X, Zhang X, Bai G, Liu Y, et al. Socioeconomic inequalities in cancer incidence and access to health services among children and adolescents in China: a cross-sectional study. Lancet 2022 Sep 24;400(10357):1020-1032 [doi: 10.1016/S0140-6736(22)01541-0] [Medline: 36154677]
48. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet 2019 Sep 28;394(10204):1192-1204 [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
49. Wickrama KAS, Surjadi FF, Lorenz FO, Conger RD, Walker C. Family economic hardship and progression of poor mental health in middle-aged husbands and wives. Fam Relat 2012 Apr 01;61(2):297-312 [FREE Full text] [doi: 10.1111/j.1741-3729.2011.00697.x] [Medline: 22577243]
50. Drentea P, Reynolds JR. Where does debt fit in the stress process model? Soc Ment Health 2015 Mar;5(1):16-32 [doi: 10.1177/2156869314554486] [Medline: 31106006]

Abbreviations

aHR: adjusted hazard ratio
aOR: adjusted odds ratio
CES-D: Center for Epidemiological Studies Depression Scale
CFPS: China Family Panel Studies
CHE: catastrophic health expenditure
CVD: cardiovascular disease
GRP: gross regional product
HR: hazard ratio
NRCMS: New Rural Cooperative Medical Scheme
OOP: out-of-pocket
UEBMI: Urban Employee Basic Medical Insurance
UHC: universal health coverage
URBMI: Urban Resident Basic Medical Insurance

Edited by A Mavragani, T Sanchez; submitted 05.09.22; peer-reviewed by T Liyuan, J Ma, S Pesälä, J Arias De La Torre; comments to author 25.04.23; revised version received 18.05.23; accepted 26.05.23; published 15.08.23

Please cite as:
Wang Y, Liang W, Liu M, Liu J
Association of Catastrophic Health Expenditure With the Risk of Depression in Chinese Adults: Population-Based Cohort Study
JMIR Public Health Surveill 2023;9:e42469
URL: https://publichealth.jmir.org/2023/1/e42469
doi: 10.2196/42469
PMID: 37581926

JAMA | Original Investigation

Extracranial-Intracranial Bypass and Risk of Stroke and Death in Patients With Symptomatic Artery Occlusion
The CMOSS Randomized Clinical Trial

Yan Ma, MD; Tao Wang, MD; Haibo Wang, PhD; Sepideh Amin-Hanjani, MD; Xiaoguang Tong, MD; Jiyue Wang, MD; Zhiyong Tong, MD; Dong Kuai, MD; Yiling Cai, MD; Jun Ren, MD; Donghai Wang, MD; Lian Duan, MD, PhD; Aisha Maimaitili, MD; Chunhua Hang, MD; Jiasheng Yu, MD; Xuesong Bai, MD; William J. Powers, MD; Colin P. Derdeyn, MD; Yangfeng Wu, PhD; Feng Ling, MD; Yuxiang Gu, MD, PhD; Liqun Jiao, MD; for the CMOSS Investigators

IMPORTANCE Prior trials of extracranial-intracranial (EC-IC) bypass surgery showed no benefit for stroke prevention in patients with atherosclerotic occlusion of the internal carotid artery (ICA) or middle cerebral artery (MCA), but there have been subsequent improvements in surgical techniques and patient selection.

OBJECTIVE To evaluate EC-IC bypass surgery in symptomatic patients with atherosclerotic occlusion of the ICA or MCA, using refined patient and operator selection.

DESIGN, SETTING, AND PARTICIPANTS This was a randomized, open-label, outcome assessor–blinded trial conducted at 13 centers in China. A total of 324 patients with ICA or MCA occlusion with transient ischemic attack or nondisabling ischemic stroke attributed to hemodynamic insufficiency based on computed tomography perfusion imaging were recruited between June 2013 and March 2018 (final follow-up: March 18, 2020).

INTERVENTIONS EC-IC bypass surgery plus medical therapy (surgical group; n = 161) or medical therapy alone (medical group; n = 163). Medical therapy included antiplatelet therapy and stroke risk factor control.

MAIN OUTCOMES AND MEASURES The primary outcome was a composite of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years after randomization. There were 9 secondary outcomes, including any stroke or death within 2 years and fatal stroke within 2 years.

RESULTS Among 330 patients who were enrolled, 324 patients were confirmed eligible (median age, 52.7 years; 257 men [79.3%]) and 309 (95.4%) completed the trial. For the surgical group vs medical group, no significant difference was found for the composite primary outcome (8.6% [13/151] vs 12.3% [19/155]; incidence difference, −3.6% [95% CI, −10.1% to 2.9%]; hazard ratio [HR], 0.71 [95% CI, 0.33-1.54]; $P = .39$). The 30-day risk of stroke or death was 6.2% (10/161) in the surgical group and 1.8% (3/163) in the medical group, and the risk of ipsilateral ischemic stroke beyond 30 days through 2 years was 2.0% (3/151) and 10.3% (16/155), respectively. Of the 9 prespecified secondary end points, none showed a significant difference including any stroke or death within 2 years (9.9% [15/152] vs 15.3% [24/157]; incidence difference, −5.4% [95% CI, −12.5% to 1.7%]; HR, 0.69 [95% CI, 0.34-1.39]; $P = .30$) and fatal stroke within 2 years (2.0% [3/150] vs 0% [0/153]; incidence difference, 1.9% [95% CI, −0.2% to 4.0%]; $P = .08$).

CONCLUSIONS AND RELEVANCE Among patients with symptomatic ICA or MCA occlusion and hemodynamic insufficiency, the addition of bypass surgery to medical therapy did not significantly change the risk of the composite outcome of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years.

TRIAL REGISTRATION ClinicalTrials.gov Identifier: NCT01758614

Visual Abstract

Editorial page 697

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The investigators in the Carotid and Middle Cerebral Artery Occlusion Surgery Study (CMOSS) are listed in Supplement 4.

Corresponding Authors: Liqun Jiao, MD, Department of Neurosurgery and Interventional Neuroradiology, Xuanwu Hospital, China International Neuroscience Institute, Capital Medical University, National Center for Neurological Disorders, 45 Changchun St, Beijing 100053, China (liqunjiao@sina.cn); Yuxiang Gu, MD, PhD, Department of Neurosurgery, Huashan Hospital, Fudan University, National Center for Neurological Disorders, 958 Jinguang St, Shanghai 201107, China (guyuxiang1972@126.com).

JAMA. 2023;330(8):704-714. doi:10.1001/jama.2023.13390

Intracranial atherosclerotic disease was a major contributor to ischemic stroke burden globally, accounting for approximately 10% of ischemic strokes in Western countries[1] and 50% in Asia in 2009.[2] Patients with symptomatic occlusion of the internal carotid artery (ICA) or middle cerebral artery (MCA) are subject to high annual recurrent stroke risks, exceeding 10% per year.[3-6] Furthermore, cerebral hemodynamic insufficiency identifies a subgroup of these patients with an even higher 2-year risk of ischemic stroke, despite medical therapy.[7]

Extracranial-intracranial (EC-IC) bypass surgery represents a plausible treatment strategy for patients with hemodynamically compromised ICA or MCA occlusion, aimed at restoring blood flow and reducing the risk of stroke by anastomosis of the superficial temporal artery to the MCA.[8] The EC/IC Bypass Study was the first randomized clinical trial demonstrating no benefit of bypass surgery in symptomatic patients with atherosclerotic stenosis or occlusion of the ICA and MCA.[9] This trial was criticized for including patients with atherosclerotic stenosis, who might be unlikely to benefit from bypass surgery, while also failing to identify patients with hemodynamic insufficiency, in whom bypass surgery might be of greatest benefit.[10] The Carotid Occlusion Surgery Study (COSS) in North America was halted early for futility, failing to demonstrate a benefit of bypass surgery in patients with hemodynamically compromised ICA occlusion.[11] Debates emerged as to whether reduction in perioperative risk could render bypass surgery effective, as the postoperative stroke rate was 15% in COSS.[12-14] Final results with adjudicated outcomes have not yet been published for a Japanese EC-IC bypass trial of 196 patients.[15]

After COSS was published in 2011, in the United States, the percentage of EC-IC bypasses performed for hospital-admitted patients with symptomatic artery occlusion decreased from 40% in 2008-2010 to 20% in 2011-2014.[16] In this context, the Carotid and Middle Cerebral Artery Occlusion Surgery Study (CMOSS) was designed as a multicenter, randomized, open-label trial comparing EC-IC bypass surgery plus medical therapy with medical therapy alone in symptomatic patients with ICA or MCA occlusion and hemodynamic insufficiency, with refined patient and operator selection.

Key Points

Question Among symptomatic patients with atherosclerotic occlusion of the internal carotid artery (ICA) or middle cerebral artery (MCA) with evidence of hemodynamic insufficiency in the affected territory, does extracranial-intracranial (EC-IC) bypass surgery plus medical therapy reduce stroke or death compared with medical therapy alone?

Findings In this randomized clinical trial that included 324 patients, the addition of bypass surgery to medical therapy did not significantly change the risk of the composite outcome of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years compared with medical therapy alone (8.6% vs 12.3%, respectively; hazard ratio, 0.71).

Meaning The findings do not support the addition of EC-IC bypass surgery to medical therapy for the treatment of patients with symptomatic atherosclerotic occlusion of the ICA or MCA.

Methods

Study Design and Oversight

CMOSS was a multicenter, randomized, open-label trial with blinded assessment of end points at 13 sites in China. Details of the study protocol have been previously published and are provided in Supplement 1.[17] A detailed statistical analysis plan is provided in Supplement 2. The trial was approved by ethics committees and research boards for each participating site and overseen by an independent data and safety monitoring board. Information on the trial design, leadership, committees, and investigators is provided in Supplement 3 and Supplement 4. All the patients or their legal representatives provided written informed consent.

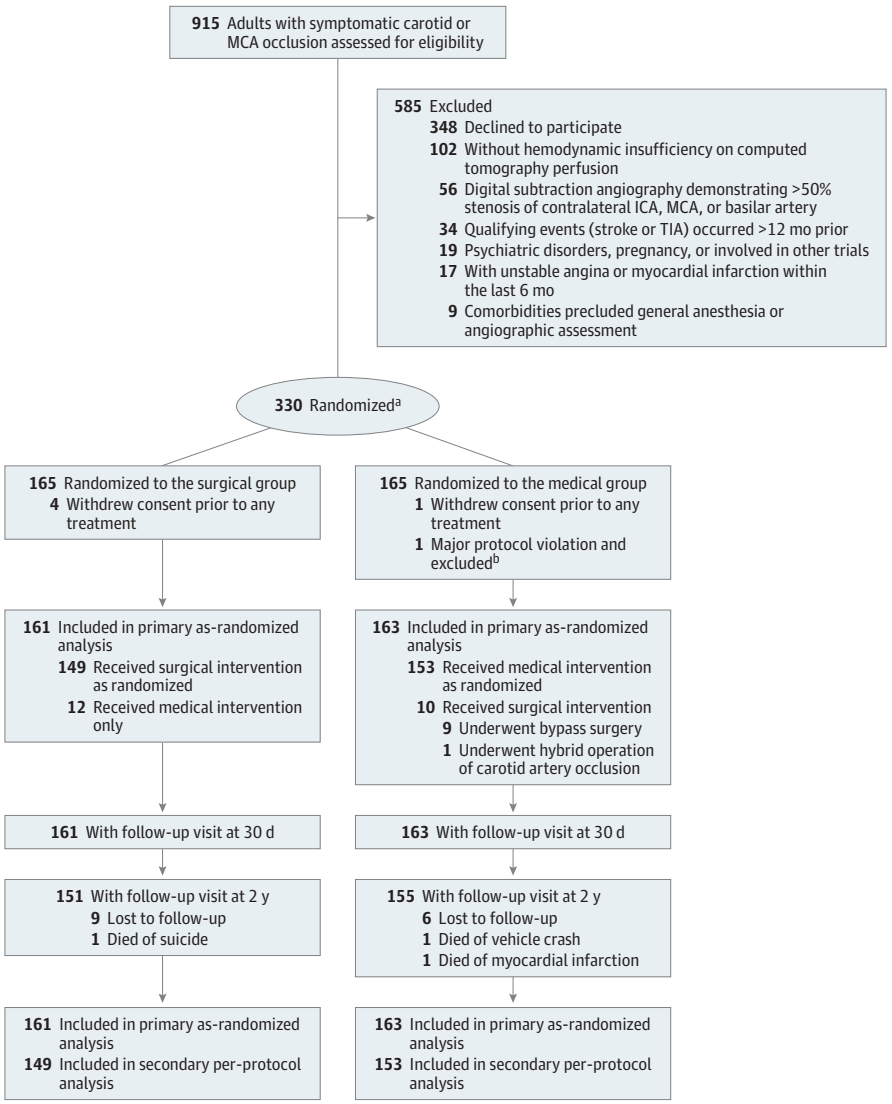
The first patient was enrolled on June 6, 2013, and the last patient on March 2, 2018. Each patient was followed up for 2 years after randomization. The 2-year follow-up for the last enrolled patient was completed on March 18, 2020.

Trial Population

Patients were eligible for inclusion in the trial if they were between 18 and 65 years, had an occlusion of a unilateral ICA or MCA demonstrated by digital subtraction angiography, and had a modified Rankin Scale (mRS) score of 0 to 2 (**Figure 1**; **Table 1**). A qualifying transient ischemic attack (TIA) or ischemic stroke in the territory of the occluded ICA or MCA must have occurred within the past 12 months, the most recent stroke must have occurred more than 3 weeks previously, and the neurologic deficit must have been stable for more than 1 month. Additionally, hemodynamic insufficiency within the MCA territory was required on imaging, defined as mean transit time (MTT) (symptomatic side) of longer than 4 seconds and relative cerebral blood flow (rCBF) (symptomatic side/asymptomatic side) of less than 0.95 on computed tomography perfusion. Detailed methods of quantitative measurement of MTT and CBF and the definition of region of interest were previously described.[18,19] Patients with more than 50% degree of stenosis of any other vessels (especially the contralateral ICA, contralateral MCA, or basilar artery) demonstrated by digital subtraction angiography, massive cerebral infarction (more than 50% of the MCA territory) demonstrated by computed tomography or magnetic resonance imaging, or other neurovascular disease likely to cause focal cerebral ischemia were excluded. Detailed eligibility criteria are provided in the eMethods in Supplement 3. An independent imaging core laboratory confirmed the imaging findings related to eligibility criteria.

Thirteen sites from 15 candidate academic tertiary centers in China met the following inclusion criteria: (1) attending neurosurgeons were certified by their roles as chief surgeon in at least 15 consecutive previous EC-IC bypass surgeries during the previous year and (2) an anastomosis patency rate greater than 95% and a perioperative stroke or death rate less than 10% according to the records from the prior year.

Figure 1. Patient Enrollment and Follow-Up in the Carotid and Middle Cerebral Artery Occlusion Surgery Study (CMOSS) Trial



ICA indicates internal carotid artery; MCA, middle cerebral artery; and TIA, transient ischemic attack.

[a] Simple randomization without block or stratification.

[b] Major protocol violation due to baseline modified Rankin Scale score of 4.

Randomization and Interventions

Patients were randomly assigned to EC-IC bypass surgery plus medical therapy (the surgical group) or medical therapy alone (the medical group) in a 1:1 ratio without block or stratification. Computer-generated random number by an interactive voice response system (Clinical Soft) was used for treatment assignment. After randomization, follow-up visits at the neurologic outpatient clinic were scheduled at 30 days as well as 6, 12, and 24 months.

All enrolled patients received optimized medical treatment including management of vascular risk factors. Risk factor management included low-density lipoprotein cholesterol level and hypertension control with medications as needed, as well as encouraging smoking cessation (eg, counseling or oral smoking cessation medications) and excess weight control (eg, behavioral change or medications), based on 2011 American Heart Association/American Stroke Association guidelines before 2014 and 2014 American Heart Association/American Stroke Association guidelines after 2014.[20,21] Aspirin (100 mg/day) or clopidogrel (75 mg/day) was prescribed to the patient and the use of any other antithrombotic therapy (specifically, oral anticoagulant agents) was discontinued until the 30-day follow-up visit.

Surgical intervention was performed within 7 days of randomization by certified surgeons. Standard end-to-side anastomosis of the superficial temporal artery to the M4 segment (MCA) was used. Major targets of perioperative management included the duration of the MCA cortical branch occlusion less than 30 minutes and steady intraprocedural arterial P_{CO_2} and periprocedural blood pressure. Specifically, continuous end-tidal P_{CO_2} needed to be calibrated with arterial blood P_{CO_2}, and if preoperative data were not available, noninvasive systolic blood pressure targets were between 130 and 150 mm Hg and P_{CO_2} targets were between 35 and 40 mm Hg.

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline

Characteristic	No. (%)[a]	
	Surgical group (n = 161)	Medical group (n = 163)
Age, median (IQR), y	52.0 (43.9-58.4)	53.5 (46.7-59.1)
Sex		
Male	125 (77.6)	132 (81.0)
Female	36 (22.4)	31 (29.0)
Body mass index, median (IQR)[b]	24.5 (23.2-26.7)	25.2 (23.0-26.8)
Medical history		
Hypertension	97 (60.2)	93 (57.1)
Diabetes	35 (21.7)	33 (20.2)
Hyperlipidemia	15 (9.3)	22 (13.5)
Previous myocardial infarction	1 (0.6)	1 (0.6)
Atrial fibrillation	1 (0.6)	0
Peripheral artery disease	1 (0.6)	0
Received medication prior to latest qualifying event		
Antiplatelet therapy	134 (83.2)	145 (89.0)
Lipid-lowering therapy	38 (23.6)	61 (37.4)
Anticoagulant therapy	3 (1.9)	1 (0.6)
Smoking history		
Current	45 (28.0)	53 (32.5)
Former	38 (23.6)	34 (20.9)
Alcohol history		
Current	35/157 (22.3)	32/162 (19.8)
Former	29/157 (18.5)	24/162 (14.8)
Qualifying event		
Stroke	90 (55.9)	90 (55.2)
TIA	71 (44.1)	73 (44.8)
Latest ischemic event to randomization, median (IQR), d		
Stroke	65.5 (46.0-102.0)	73.0 (43.0-119.0)
TIA	58.0 (36.0-122.0)	54.0 (33.0-92.0)
Qualifying artery		
Carotid artery	87 (54.0)	99 (60.7)
Middle cerebral artery	74 (46.0)	64 (39.3)
Qualifying side		
Left	87 (54.0)	88 (54.0)
Right	74 (46.0)	75 (46.0)
Contralateral artery stenosis <50%		
Carotid artery	8 (5.0)	11 (6.7)
Middle cerebral artery	1 (0.6)	3 (1.8)
Computed tomography perfusion		
Mean transit time (qualifying side), median (IQR), s[c]	6.3 (5.1-9.0) [n = 153]	5.8 (4.9-7.6) [n = 157]
Relative cerebral blood flow (qualifying/contralateral side), median (IQR)[d]	0.8 (0.7-0.9) [n = 156]	0.7 (0.6-0.8) [n = 161]
NIHSS score, median (IQR)[e]	0.0 (0.0-2.0)	0.0 (0.0-2.0)
0	81 (50.3)	93 (57.1)
≥1	80 (49.7)	70 (42.9)
mRS score, median (IQR)[f]	1.0 (1.0-1.0)	1.0 (0.0-1.0)
0	36 (22.4)	46 (28.2)
1	89 (55.3)	85 (52.2)
2	36 (22.4)	32 (19.6)
Blood pressure, median (IQR), mm Hg		
Systolic	130.0 (121.0-140.0)	130.0 (120.0-140.0)
Diastolic	80.0 (75.0-87.0)	80.0 (73.0-85.0)

(continued)

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline (continued)

Characteristic	No. (%)[a]	
	Surgical group (n = 161)	Medical group (n = 163)
Hemoglobin A_{1C}, median (IQR), %	5.8 (5.4-6.7) [n = 80]	5.8 (5.4-6.8) [n = 88]
LDL cholesterol level, median (IQR), mmol/L	1.9 (1.4-2.3) [n = 154]	2.0 (1.6-2.4) [n = 157]
Triglycerides level, median (IQR), mmol/L	1.3 (1.0-1.8) [n = 154]	1.5 (1.1-1.8) [n = 157]

Abbreviations: LDL, low-density lipoprotein; TIA, transient ischemic attack; NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin Scale.

SI conversion factors: To convert LDL cholesterol to mg/dL, divide by 0.0259; triglycerides to mg/dL, divide by 0.0113.

[a] Data are reported as No. (%) unless otherwise indicated.

[b] Body mass index is the weight in kilograms divided by height in meters squared.

[c] Mean transit time designates the mean time required for a contrast bolus to traverse the voxel and is measured in seconds. Mean transit time is inversely proportional to cerebral perfusion pressure and prolonged value indicates arterial stenosis or occlusion.

[d] Cerebral blood flow refers to the volume of blood flowing in a unit of brain mass during a unit of time (measured in mL/100 g/min). Relative cerebral blood flow denotes symptomatic side/asymptomatic side.

[e] Scores on the NIHSS range from 0 to 42, with higher scores indicating a more severe neurologic deficit.

[f] Scores on the mRS range from 0 (no functional limitations) to 6 (death), with higher scores indicating more severe functional disability. A score of 2 or less indicates functional independence.

Outcome Measures and Definitions

The primary outcome was a composite of any stroke or death within 30 days after randomization or ipsilateral ischemic stroke beyond 30 days through 2 years after randomization. Secondary outcomes included (1) any stroke within 2 years; (2) disabling stroke within 2 years; (3) fatal stroke within 2 years; (4) death within 2 years; (5) any stroke or death within 2 years; (6) TIA within 2 years; (7) National Institutes of Health Stroke Scale (NIHSS) score and modified Rankin Scale (mRS) score within 2 years; (8) complications associated with the surgical procedures; and (9) anastomosis patency at 2 years.

A stroke was defined as rapidly developed clinical signs of focal or global disturbance of cerebral function that lasted more than 24 hours due to cerebral ischemia or hemorrhage, according to World Health Organization criteria. Ipsilateral ischemic stroke was further defined as the clinical diagnosis of a focal or global disturbance of cerebral function due to cerebral ischemia that was clinically localizable within the territory of the symptomatic occluded ICA or MCA and that lasted for more than 24 hours. Brain computed tomography or magnetic resonance image scanning was used to identify stroke. The mRS is a standard global 7-level measure of disability, ranging from 0 (no functional limitations) to 6 (death), with higher scores indicating more severe functional disability. Scores on the NIHSS range from 0 to 42, with higher scores indicating a more severe neurologic deficit. Disabling stroke was defined by any of the following: (1) an mRS score of 3 or more; (2) an increase of at least 1 point in the mRS score from prestroke baseline; (3) a score on the composite NIHSS of 7 or more; or (4) an increase of at least 4 points in the NIHSS score from prestroke baseline. Detailed definitions are provided in Supplement 1. An independent outcome committee and imaging core laboratory determined the primary and secondary outcomes, blinded to treatment assignment.

Sample Size Calculation

We assumed that the true primary outcome rates would be 28% in the medical group,[22-24] and there would be a 50% relative risk reduction in the surgical group. This meant that the primary outcome rate was assumed to be 14% in the surgical group. We estimated that 330 patients (165 per group) would provide 80% power using a 2-tailed α of 5% to detect an absolute 14% difference in the rate of primary outcome events in the surgical group (estimated as 14%) from that in the medical group (estimated as 28%), with an attrition rate of 20% during 2 years of follow-up.

Statistical Analysis

Outcome evaluation was conducted in both the full analysis set and per-protocol set, with analysis in the full analysis set as the primary analysis. The full analysis set population included all the patients who underwent randomization and treatment, and patients were analyzed according to their randomized groups. The difference in the primary outcome between groups was tested using log-rank test, with the center information (site effect) as a stratification factor. The same test was used to compare the secondary outcomes including 2-year any stroke, disabling stroke, fatal stroke, death, any stroke or death, and TIA between 2 groups. Kaplan-Meier curves were used to show the risk of time-to-event outcomes over time. Participants who were lost to follow-up or who died from nonstroke causes were censored at the time of occurrence. For other secondary outcomes and baseline characteristics, χ^2 or Fisher exact tests were used for categorical variables, and *t* or Wilcoxon rank tests for quantitative variables.

We estimated hazard ratios (HRs) with Cox proportional-hazards models for time-to-event analyses of the primary outcome and secondary outcomes. However, we also performed post hoc relative risk (RR) analyses for both primary and secondary outcomes as the underlying assumption of proportional hazards was not met with supremum test ($P < .001$). Participants lost to follow-up were presumed to not have experienced the event of interest for primary outcome and secondary outcomes in the RRs calculation. Other post hoc analyses included (1) analysis of the individual components of the primary outcome (ie, stroke or death within 30 days after randomization, ipsilateral ischemic stroke beyond 30 days through 2 years after randomization);

(2) subgroup analyses by age, sex, received medication prior to latest qualifying event (antiplatelet therapy, lipid-lowering therapy), qualifying event, time interval between the latest ischemic event to randomization (≤6 weeks, >6 weeks), qualifying artery, MTT (≤6 seconds, >6 seconds), and rCBF (≤0.8, >0.8). The interaction effects between interventions and the factors mentioned above were evaluated using the generalized linear regression model with the binomial distribution and log link function in the RRs subgroup analyses. The trial was not powered for and had no prespecified correction for multiple comparisons for a definitive analysis of secondary end points or subgroups. The findings for these analyses should be interpreted as exploratory.

All statistical tests were performed by 2-sided tests. A *P* value of .05 was considered to demonstrate a statistically significant difference. All analyses were performed with SAS software version 9.4 (SAS Institute).

Results

Between June 6, 2013, and March 2, 2018, a total of 915 patients with symptomatic ICA or MCA occlusion were screened at 13 sites; 330 patients were enrolled, with 165 each randomly assigned to the surgical and medical groups (eFigure 1 in Supplement 3). A total of 6 patients (4 in the surgical group and 2 in the medical group) immediately withdrew from the study without any treatment or data collection and were excluded from the full analysis set for final analysis. The remaining 324 patients (161 in the surgical and 163 in the medical group) were included in the full analysis set for final analysis. A total of 15 patients (4.6%) had incomplete follow-up for the primary outcome (Figure 1). The progress of the trial, such as on-site clarification and confirmation for data inquiries, was significantly delayed due to the COVID-19 pandemic.

Baseline characteristics were generally balanced between the 2 groups, except for the percentage of patients with previous lipid-lowering therapy and the level of triglycerides, which were lower in the surgical group, and the MTT of the qualifying side and rCBF, which were higher in the surgical group. The median age of the participants was 52.7 years (IQR, 44.6-58.7), and 257 (79.3%) were male. Among all 324 patients, 180 patients (55.6%) presented with index stroke as the qualifying event, and 186 (57.4%) were diagnosed as having ICA occlusion vs 138 (42.6%) as having MCA occlusion as the qualifying artery (Table 1). At the last follow-up visit, risk factor control was similar in both groups (eTable 1 in Supplement 3).

Primary Outcome

For the surgical group vs medical group, no significant difference was found for the primary outcome of risk of stroke or death (8.6% [13/151] vs 12.3% [19/155]; incidence difference, −3.6% [95% CI, −10.1% to 2.9%]; HR, 0.71 [95% CI, 0.33-1.54]; *P* = .39) (**Table 2** and **Figure 2**). The primary outcome was missing for 18 patients (10 in the surgical group and 8 in the medical group). Among patients assigned to the medical group, 10 patients crossed over to surgical procedures. Four patients were further lost to follow-up, 1 died of a vehicle crash, and 1 died of myocardial infarction beyond 30 days. Among patients assigned to the surgical group, 12 patients crossed over to medical treatment only. Nine patients were further lost to follow-up and 1 died of suicide beyond 30 days. The per-protocol analysis yielded similar results (9.4% [13/139] vs 10.9% [16/147]; incidence difference, −1.7% [95% CI, −8.4% to 4.9%]; HR, 1.09 [95% CI, 0.50-2.37]; *P* = .85) (eTable 2 in Supplement 3).

Secondary Outcomes

All secondary outcomes showed no significant difference between groups (Table 2; eFigures 2-8 in Supplement 3). The 2-year risk of any stroke or death was 9.9% (15/152) in the surgical group vs 15.3% (24/157) in the medical group (incidence difference, −5.4% [95% CI, −12.5% to 1.7%]; HR, 0.69 [95% CI, 0.34-1.39]; *P* = .30). The risk of fatal stroke within 2 years was 2.0% (3/150) in the surgical group vs 0% (0/153) in the medical group (incidence difference, 1.9% [95% CI, −0.2% to 4.0%]; *P* = .08). The risk of disabling stroke within 2 years was 4.1% (6/147) in the surgical group and 2.0% (3/153) in the medical group (incidence difference, 1.9% [95% CI, −1.7% to 5.5%]; HR, 1.75 [95% CI, 0.41-7.51]; *P* = .43). The 2-year mortality risk was 2.7% (4/151) in the surgical group vs 1.3% (2/155) in the medical group (incidence difference, 1.3% [95% CI, −1.7% to 4.2%]; HR, 3.72 [95% CI, 0.68-20.38]; *P* = .12). Anastomosis patency at 2 years in the surgical group was 93.6% (103/110). Other secondary outcomes are shown in Table 2.

Post Hoc Outcomes and Analyses

Post hoc analysis of the individual components of the primary outcome showed the 30-day risk of stroke or death was 6.2% (10/161) in the surgical group vs 1.8% (3/163) in the medical group, and the risk of ipsilateral ischemic stroke beyond 30 days through 2 years was 2.0% (3/151) in the surgical group vs 10.3% (16/155) in the medical group (Table 2; eFigures 9-10 in Supplement 3). In patients with MTT longer than 6 seconds, the primary outcome occurred in 7 (9.2%) of 76 patients in the surgical group and 12 (17.4%) of 69 patients in the medical group (incidence difference, −8.2% [95% CI, −18.7% to 2.3%]; HR, 0.57 [95% CI, 0.21-1.59]) (**Figure 3**). In patients with rCBF of 0.8 or less, the primary outcome occurred in 5 (6.4%) of 78 patients in the surgical group and 15 (14.0%) of 107 patients in the medical group (incidence difference, −7.4% [95% CI, −15.5% to 0.7%]; HR, 0.38 [95% CI, 0.13-1.18]) (Figure 3; eFigure 11 and eTable 3 in Supplement 3). Details of perioperative management in the surgical group are shown in eTable 4 in Supplement 3.

Discussion

This randomized clinical trial showed that, among patients with TIA or nondisabling ischemic stroke due to ICA or MCA occlusion and hemodynamic insufficiency, the primary outcome (a composite of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years)

Table 2. Primary and Secondary Outcomes

Outcome	No. (%)		Hazard ratio (95% CI)[a]	Incidence difference, % (95% CI)[b]	Relative risk (95% CI)[c]	*P* value[d]
	Surgical group (n = 161)	Medical group (n = 163)				
Primary outcome						
Stroke or death within 30 d or ischemic stroke in territory of qualifying artery beyond 30 d through 2 y	13/151 (8.6)	19/155 (12.3)	0.71 (0.33 to 1.54)	−3.6 (−10.1 to 2.9)	0.69 (0.35 to 1.36)	.39
Components of the primary outcome[e]						
Stroke or death within 30 d[f]	10/161 (6.2)	3/163 (1.8)				
Ischemic stroke in territory of qualifying artery beyond 30 d through 2 y	3/151 (2.0)	16/155 (10.3)				
Secondary outcomes						
Any stroke or death within 2 y	15/152 (9.9)	24/157 (15.3)	0.69 (0.34 to 1.39)	−5.4 (−12.5 to 1.7)	0.63 (0.34 to 1.16)	.30
Any stroke within 2 y	14/151 (9.3)	22/155 (14.2)	0.67 (0.33 to 1.39)	−4.8 (−11.6 to 2.0)	0.64 (0.34 to 1.21)	.29
Disabling stroke within 2 y[g]	6/147 (4.1)	3/153 (2.0)	1.75 (0.41 to 7.51)	1.9 (−1.7 to 5.5)	2.02 (0.52 to 7.96)	.43
Deaths within 2 y	4/151 (2.7)	2/155 (1.3)	3.72 (0.68 to 20.38)	1.3 (−1.7 to 4.2)	2.02 (0.38 to 10.91)	.12
TIA within 2 y	3/147 (2.0)	6/154 (3.9)	0.31 (0.08 to 1.30)	−1.8 (−5.4 to 1.8)	0.51 (0.13 to 1.99)	.10
Fatal stroke within 2 y	3/150 (2.0)	0/153	NA	1.9 (−0.2 to 4.0)	NA	.08
NHISS score at 2 y, median (IQR)[h]	0.0 (0.0 to 1.0) [n = 139]	0.0 (0.0 to 0.0) [n = 148]				.44[i]
0	101/139 (72.7)	113/148 (76.4)				.47[j]
≥1	38/139 (27.3)	35/148 (23.6)				
mRS score at 2 y, median (IQR)[k]	0.0 (0.0 to 1.0) [n = 148]	0.0 (0.0 to 1.0) [n = 151]				.51[i]
0	88/148 (59.5)	84/151 (55.6)				.51[i]
1	42/148 (28.4)	44/151 (29.1)				
2	8/148 (5.4)	18/151 (11.9)				
3	2/148 (1.4)	3/151 (2.0)				
4	4/148 (2.7)	0/151				
5	0/148	0/151				
6	4/148 (2.7)	2/151 (1.3)				
Complications associated with the surgical procedures[l]	6/149 (4.0)	NA				NA
Anastomosis patency at 2 y	103/110 (93.6)	NA				NA

Abbreviations: mRS, modified Rankin Scale; NA, not applicable; NIHSS, National Institutes of Health Stroke Scale; TIA, transient ischemic attack.

[a] Cox proportional hazard model adjusted for site effect.

[b] Absolute incidence difference without adjustment.

[c] Post hoc analysis of relative risk adjusted for site effect based on the assumption that patients with missing data did not have the primary or secondary outcomes.

[d] Log-rank test adjusted for site effect.

[e] Individual elements of the primary outcome were analyzed post hoc.

[f] The mortality rate was 1.9% (3/161) in the surgical group and 0% (0/163) in the medical group. The stroke rate was 6.2% (10/161) and 1.8% (3/163), respectively.

[g] The rate of disabling stroke within 30 days was 1.9% (3/158) in the surgical group and 0% (0/163) in the medical group.

[h] Scores on the NIHSS range from 0 to 42, with higher scores indicating a more severe neurologic deficit. Twenty-two lost to follow-up in the surgical group and 15 in the medical group.

[i] The *P* value was calculated with the use of the nonparametric Wilcoxon test.

[j] χ^2 test.

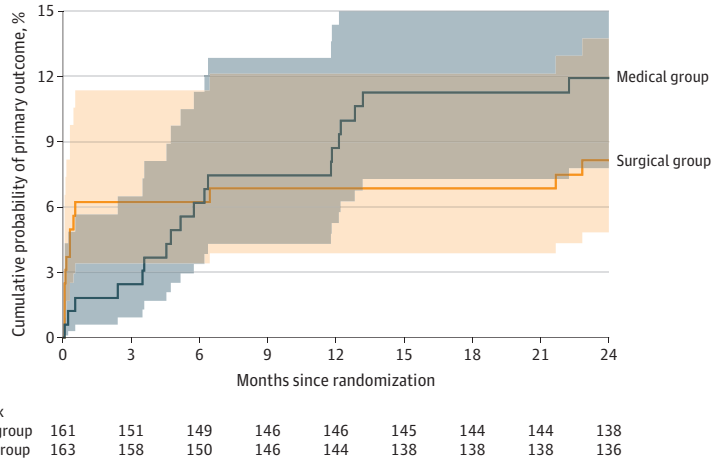
[k] Scores on the mRS range from 0 (no functional limitations) to 6 (death), with higher scores indicating more severe functional disability. A score of 2 or less indicates functional independence. Seventeen lost to follow-up in the surgical group and 14 in the medical group.

[l] There were 3 cases of asymptomatic cerebral infarction, 1 epidural hematoma, 1 hyperperfusion syndrome, and 1 subdural hematoma resulting in TIA.

did not differ significantly in patients who received EC-IC bypass surgery plus medical treatment compared with medical treatment alone. The results on all prespecified secondary outcomes were consistent with the primary result.

This trial demonstrated a lower rate of the primary outcome in the medical therapy alone group compared with previous studies. In this trial, the 2-year primary outcome rate was 12.3%, nearly half of that reported in COSS a decade ago (21.0%).[11]

Figure 2. Cumulative Probability of the Primary Outcome, According to Treatment Assignment



No. at risk									
Surgical group	161	151	149	146	146	145	144	144	138
Medical group	163	158	150	146	144	138	138	138	136

The primary outcome was a composite of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years after randomization. Nine patients lost to follow-up within 2 years in the surgical group and 6 patients in the medical group were treated as censored data. All other patients were followed up to event or 2 years. The median time of observation was 24.0 months (IQR, 24.0-24.0) for the surgical group and 24.0 months (IQR, 24.0-24.0) for the medical group. *P* = .39 for log-rank testing between the surgical group and medical group with the center information (site effect) as a stratification factor. The shading indicates 95% CI of the primary outcome.

Also, the rate was much lower than the anticipated 28% in the trial design.[17] Several reasons may have contributed to this discrepancy. First, this may have been due to improved efficacy of medical treatment, and similar trends of improved efficacy of medical treatment for other cerebrovascular diseases have been frequently reported. For example, for patients with asymptomatic carotid artery stenosis, ipsilateral ischemic stroke within 5 years in the medical group was lower in the SPACE-2 trial than that in the ACAS trial almost 30 years prior (3.1% vs 11.0%).[25,26] Also, in one study comparing stroke rates in the WASID and SAMMPRIS trials for patients with intracranial stenosis, a lower stroke rate at 12 months (12.6% vs 21.9%) was found in SAMMPRIS, which featured more intensive medical therapy targets.[27] Second, compared with COSS, the present study achieved an overall better control of risk factors for atherosclerosis such as hypertension and hyperlipidemia (eTable 1 in Supplement 3). Third, the imaging selection criteria may have identified less severe hemodynamic insufficiency than would be indicative of the target high-risk population. In contrast to COSS,[11] which relied on stage 2 hemodynamic failure on positron emission tomography imaging, the present study defined it on computed tomography perfusion with MTT longer than 4 seconds and rCBF ratio less than 0.95. However, in subgroup analysis, patients with MTT longer than 6 seconds (median of MTT) had higher primary outcome rates (Figure 3).

The perioperative stroke rate of 6.2% in this trial was substantially lower than that of 15% in COSS and that of 12% in the prior EC-IC Bypass Study.[9,11] In recent observational studies, the reported rate of perioperative stroke following bypass surgery for atherosclerotic vessel occlusion ranged from 4.3% to 8.9%,[12,28] similar to that of the present study. Besides the use of stricter criteria for surgeon certification in this trial compared with COSS,[17] intraoperative temporary occlusion time was also lower in this trial than COSS (21.0 vs 55.9 minutes [with stroke] or 45.4 minutes [without stroke]; eTable 4 in Supplement 3).[14] Strict periprocedural care implemented in this trial could be a further key factor in reducing perioperative risk. For example, the trial required monitoring intraprocedural P_{CO_2} and periprocedural blood pressure level, with protocol-defined targets (eTable 4 in Supplement 3). The inclusion criteria relating the time interval from the latest ischemic event to randomization was also different between this trial and COSS (12 months in CMOSS vs 120 days in COSS), but the actual time interval in CMOSS was comparable with that of COSS. Thus, this factor may not be a reason for the reduction of complication rates. This trial also limited enrollment to patients younger than 65 years compared with younger than 85 years in COSS, with the mean 52 vs 58 years of age. This may have contributed to the reduced risk in both groups of the trial, comparing this trial with the earlier study.

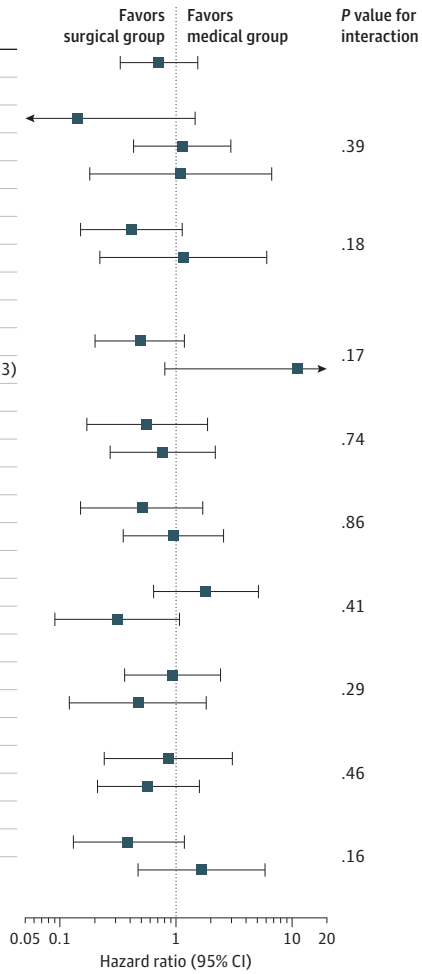
Despite the lower rates of perioperative risk than prior trials and the numerically lower rates of ipsilateral stroke beyond 30 days in the bypass surgery group in this trial, this study failed to show overall superiority of bypass surgery over medical treatment.[17] If bypass surgery is investigated in future trials, there are several factors to consider. Further refinement of patient selection may be needed to identify patients who could benefit most from bypass surgery. The subgroup of patients with MCA occlusion, who were not included in COSS, had a numerically lower rate of the primary outcome at 2 years compared with patients with ICA occlusion (5.8% [4/69] vs 11.0% [9/82]; *P* = .38; Figure 3), but the difference was not statistically significant and the test for interaction was not statistically significant (*P* for interaction = .29). The timing of bypass surgery may also need to be considered further; the perioperative risk of bypass surgery may be increased if surgery is performed shortly after the latest ischemic event, and the benefit of surgery may decrease in patients if surgery is performed much later after their event because hypoperfusion is compensated or medically managed. However, the optimal timing of bypass surgery is still unknown and may warrant further studies. In sum, future trials of EC-IC bypass may need larger sample sizes, refined patient selection (eg, MCA occlusion, MTT >6 seconds or rCBF ≤0.8), and longer follow-up time.

Limitations

This trial had several limitations. First, because sham surgery was not performed, there is the potential for bias in individual

Figure 3. Post Hoc Subgroup Analysis for the Primary Outcome[a]

	No./total (%)				
	Surgical group	Medical group	Incidence difference (95% CI)	Hazard ratio (95% CI)	*P* value for interaction
Overall	13/151 (8.6)	19/155 (12.3)	-3.6 (-10.1 to 2.9)	0.71 (0.33-1.54)	
Age, y					
<50	2/62 (3.2)	6/59 (10.2)	-6.6 (-15.1 to 1.9)	0.14 (0.01-1.46)	
50-60	9/67 (13.4)	9/64(14.1)	-0.9 (-11.9 to 10.1)	1.14 (0.43-2.97)	.39
>60	2/22 (9.1)	4/32 (12.5)	-3.4 (-19.4 to 12.6)	1.10 (0.18-6.69)	
Sex					
Male	7/116 (6.0)	15/124 (12.1)	-5.8 (-12.5 to 1.0)	0.41 (0.15-1.13)	
Female	6/35 (17.1)	4/31 (12.9)	3.8 (-13.2 to 20.7)	1.15 (0.22-6.06)	.18
Received medication prior to latest qualifying event					
Antiplatelet therapy					
Yes	9/125 (7.2)	18/140 (12.9)	-5.7 (-12.5 to 1.1)	0.49 (0.20-1.18)	
No	4/26 (15.4)	1/15 (6.7)	9.3 (-7.8 to 26.3)	11.15 (0.80-155.33)	.17
Lipid-lowering therapy					
Yes	4/36 (11.1)	10/57 (17.5)	-5.9 (-19.3 to 7.6)	0.56 (0.17-1.87)	
No	9/115 (7.8)	9/98 (9.2)	-1.5 (-8.7 to 5.7)	0.77 (0.27-2.18)	.74
Qualifying event					
Stroke	6/81 (7.4)	8/86 (9.3)	-2.2 (-10.0 to 5.6)	0.51 (0.15-1.70)	
TIA	7/70 (10.0)	11/69 (15.9)	-5.2 (-16.0 to 5.5)	0.95 (0.35-2.57)	.86
Latest ischemic event to randomization, wk					
≤6	7/47 (14.9)	8/47 (17.0)	-1.1 (-15.2 to 13.0)	1.81 (0.64-5.14)	
>6	6/104 (5.8)	11/108 (10.2)	-4.5 (-11.4 to 2.4)	0.31 (0.09-1.07)	.41
Qualifying artery					
Carotid artery	9/82 (11.0)	11/95 (11.6)	-0.8 (-9.7 to 8.1)	0.94 (0.36-2.42)	
Middle cerebral artery	4/69 (5.8)	8/60 (13.3)	-7.1 (-16.7 to 2.5)	0.47 (0.12-1.82)	.29
Mean transit time (qualifying side), s[b]					
≤6	5/67 (7.5)	7/80 (8.8)	-1.1 (-9.5 to 7.3)	0.86 (0.24-3.06)	
>6	7/76 (9.2)	12/69 (17.4)	-8.2 (-18.7 to 2.3)	0.57 (0.21-1.59)	.46
Relative cerebral blood flow[c]					
≤0.8	5/78 (6.4)	15/107 (14.0)	-7.4 (-15.5 to 0.7)	0.38 (0.13-1.18)	
>0.8	8/68 (11.8)	4/46 (8.7)	3.0 (-7.6 to 13.5)	1.66 (0.47-5.86)	.16



HR indicates hazard ratio; and TIA, transient ischemic attack.

[a] The primary outcome was a composite of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years after randomization.

[b] Mean transit time designates the mean time required for a contrast bolus to traverse the voxel and is measured in seconds. Mean transit time is inversely proportional to cerebral perfusion pressure and prolonged value indicates arterial stenosis or occlusion.

[c] Cerebral blood flow refers to the volume of blood flowing in a unit of brain mass during a unit of time (measured in mL/100 g/min). Relative cerebral blood flow denotes symptomatic side/asymptomatic side.

sites reporting potential end points for adjudication. Second, this study was conducted only in centers in China, and its generalizability to other populations outside of China is uncertain. Third, the enrolled population had a lower risk of events than that anticipated in the trial design, and the study may be underpowered to detect clinically relevant differences between groups. Fourth, the low-density lipoprotein cholesterol target level (<2.58 mmol/L [100 mg/dL]) according to guidelines at the time the trial began was relatively conservative compared with a lower low-density lipoprotein cholesterol target for those with large artery atherosclerosis in more recent trials. Even lower stroke rates in the medical therapy alone group may be observed in the future because only 38.1% of the medically treated patients reached the modern target low-density lipoprotein level of less than 70 mg/dL. Fifth, the enrollment of women was suboptimal and future research is needed to focus on the sex difference in atherosclerotic ICA or MCA occlusion and its prognosis.

Conclusions

Among patients with symptomatic ICA or MCA occlusion with hemodynamic insufficiency, the addition of bypass surgery to medical therapy did not significantly change the risk of stroke or death within 30 days or ipsilateral ischemic stroke beyond 30 days through 2 years. The findings do not support the addition of EC-IC bypass surgery to medical therapy for the treatment of patients with symptomatic atherosclerotic occlusion of the ICA or MCA.

ARTICLE INFORMATION

Accepted for Publication: June 29, 2023.

Author Affiliations: Department of Neurosurgery, Xuanwu Hospital, China International Neuroscience Institute, Capital Medical University, National Center for Neurological Disorders, Beijing, China (Ma, T. Wang, Bai); Peking University Clinical Research Institute, Peking University First Hospital, Beijing, China (H. Wang, Wu); Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China (H. Wang); Cerebrovascular and Skull Base Surgery, Department of Neurological Surgery, University Hospitals Cleveland Medical Center, Cleveland, Ohio (Amin-Hanjani); Department of Neurosurgery, Huanhu Hospital, Tianjin, China (X. Tong); Department of Neurosurgery, Liaocheng People's Hospital, Shandong First Medical University and Shandong Academy of Medical Sciences, Liaocheng City, Shandong, China (J. Wang); Department of Neurosurgery, The First Hospital of China Medical University, Shenyang, Liaoning, China (Z. Tong); Department of Neurosurgery, The Affiliated Cardiovascular Hospital of Shanxi Medical University and Shanxi Cardiovascular Hospital (Institute), Taiyuan, Shanxi, China (Kuai); Department of Neurology, Strategic Support Force Medical Center, Beijing, China (Cai); Department of Neurosurgery, The Second Hospital of Lan Zhou University, Lan Zhou, China (Ren); Department of Neurosurgery, Qilu Hospital of Shandong University, Jinan, China (D. Wang); Qilu Hospital of Shandong University Dezhou Hospital, Dezhou, China (D. Wang); Department of Neurosurgery, The First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, HaiDian District, Beijing, China (Duan); Department of Neurosurgery, The First Affiliated Hospital of Xinjiang Medical University, Xinshi District, Urumqi, Xinjiang,China (Maimaitili); Department of Neurosurgery, Nanjing Drum Tower Hospital, Neurosurgical Institute of Nanjing University, Nanjing University Medical School, Nanjing, China (Hang); Department of Neurosurgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China (Yu); Department of Neurology, Duke University School of Medicine, Duke South, Durham, North Carolina (Powers); Departments of Radiology and Neurology, University of Iowa Hospitals and Clinics, Iowa City (Derdeyn); Department of Neurosurgery and Interventional Neuroradiology, Xuanwu Hospital, China International Neuroscience Institute, Capital Medical University, National Center for Neurological Disorders, Beijing, China (Ling, Jiao); Department of Neurosurgery, Huashan Hospital, Fudan University, National Center for Neurological Disorders, Shanghai, China (Gu).

Author Contributions: Dr Jiao had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Ma, T. Wang, H. Wang, and Amin-Hanjani are co-first authors.
Concept and design: Ma, Jiao.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Ma, T. Wang, H. Wang, X. Tong, J. Wang, Cai, Bai.
Critical review of the manuscript for important intellectual content: T. Wang, H. Wang, Amin-Hanjani, X. Tong, J. Wang, Z. Tong, Kuai, Cai, Ren, D. Wang, Duan, Maimaitili, Hang, Yu, Bai, Powers, Derdeyn, Wu, Ling, Gu, Jiao.
Statistical analysis: H. Wang, D. Wang, J. Powers.
Obtained funding: Jiao.
Administrative, technical, or material support: T. Wang, X. Tong, J. Wang, Z. Tong, Kuai, Cai, Ren, D. Wang, Duan, Maimaitili, Hang, Yu, Bai, Ling, Gu, Jiao.
Supervision: Ma, D. Wang, Wu, Gu, Jiao.

Conflict of Interest Disclosures: Dr Derdeyn reported receiving personal fees from Penumbra (data and safety monitoring board for MIND and THUNDER trials), NoNo Inc (data and safety monitoring board for ESCAPE NEXT and FRONTIER trials), and Euphrates Vascular Inc and grants from Siemens Healthineers outside the submitted work. No other disclosures were reported.

Funding/Support: This work was supported by a research grant (2011BAI08B00) from the National Health Commission of the People's Republic of China.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The Carotid and Middle Cerebral Artery Occlusion Surgery Study (CMOSS) investigators are listed in Supplement 4.

Data Sharing Statement: See Supplement 5.

Additional Contributions: We thank the patients and their families for participating in this trial.

REFERENCES

1. White H, Boden-Albala B, Wang C, et al. Ischemic stroke subtype incidence among Whites, Blacks, and Hispanics: the Northern Manhattan Study. *Circulation*. 2005;111(10):1327-1331. doi:10.1161/01.CIR.0000157736.19739.D0

2. Wang Y, Zhao X, Liu L, et al; CICAS Study Group. Prevalence and outcomes of symptomatic intracranial large artery stenoses and occlusions in China: the Chinese Intracranial Atherosclerosis (CICAS) Study. *Stroke*. 2014;45(3):663-669. doi:10.1161/STROKEAHA.113.003508

3. Grubb RL Jr, Derdeyn CP, Fritsch SM, et al. Importance of hemodynamic factors in the prognosis of symptomatic carotid occlusion. *JAMA*. 1998;280(12):1055-1060. doi:10.1001/jama.280.12.1055

4. Kern R, Steinke W, Daffertshofer M, Prager R, Hennerici M. Stroke recurrences in patients with symptomatic vs asymptomatic middle cerebral artery disease. *Neurology*. 2005;65(6):859-864. doi:10.1212/01.wnl.0000175983.76110.59

5. Flaherty ML, Flemming KD, McClelland R, Jorgensen NW, Brown RD Jr. Population-based study of symptomatic internal carotid artery occlusion: incidence and long-term follow-up. *Stroke*. 2004;35(8):e349-e352. doi:10.1161/01.STR.0000135024.54608.3f

6. Reynolds MR, Derdeyn CP, Grubb RL Jr, Powers WJ, Zipfel GJ. Extracranial-intracranial bypass for ischemic cerebrovascular disease: what have we learned from the Carotid Occlusion Surgery Study? *Neurosurg Focus*. 2014;36(1):E9. doi:10.3171/2013.10.FOCUS13427

7. Grubb RL Jr, Derdeyn CP, Videen TO, Carpenter DA, Powers WJ. Relative mean transit time predicts subsequent stroke in symptomatic carotid occlusion. *J Stroke Cerebrovasc Dis*. 2016;25(6):1421-1424. doi:10.1016/j.jstrokecerebrovasdis.2015.12.041

8. Vilela MD, Newell DW. Superficial temporal artery to middle cerebral artery bypass: past, present, and future. *Neurosurg Focus*. 2008;24(2):E2. doi:10.3171/FOC/2008/24/2/E2

9. EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke: results of an international randomized trial. *N Engl J Med*. 1985;313(19):1191-1200. doi:10.1056/NEJM198511073131904

10. Schmiedek P, Piepgras A, Leinsinger G, Kirsch CM, Einhüpl K. Improvement of cerebrovascular reserve capacity by EC-IC arterial bypass surgery in patients with ICA occlusion and hemodynamic cerebral ischemia. *J Neurosurg*. 1994;81(2):236-244. doi:10.3171/jns.1994.81.2.0236

11. Powers WJ, Clarke WR, Grubb RL Jr, Videen TO, Adams HP Jr, Derdeyn CP; COSS Investigators. Extracranial-intracranial bypass surgery for stroke prevention in hemodynamic cerebral ischemia: the Carotid Occlusion Surgery Study randomized trial. *JAMA*. 2011;306(18):1983-1992. doi:10.1001/jama.2011.1610

12. Wessels L, Hecht N, Vajkoczy P. Patients receiving extracranial to intracranial bypass for atherosclerotic vessel occlusion today differ significantly from the COSS population. *Stroke*. 2021;52(10):e599-e604. doi:10.1161/STROKEAHA.120.033991

13. Amin-Hanjani S, Barker FG II, Charbel FT, Connolly ES Jr, Morcos JJ, Thompson BG; Cerebrovascular Section of the American Association of Neurological Surgeons; Congress of Neurological Surgeons. Extracranial-intracranial bypass for stroke-is this the end of the line or a bump in the road? *Neurosurgery*. 2012;71(3):557-561. doi:10.1227/NEU.0b013e3182621488

14. Grubb RL Jr, Powers WJ, Clarke WR, Videen TO, Adams HP Jr, Derdeyn CP; Carotid Occlusion Surgery Study Investigators. Surgical results of the Carotid Occlusion Surgery Study. *J Neurosurg*. 2013;118(1):25-33. doi:10.3171/2012.9.JNS12551

15. Group JS. Japanese EC-IC Bypass Trial (JET Study): the second interim analysis. *Surg Cereb Stroke*. 2002;30:434-437. doi:10.2335/scs.30.434

16. Winkler EA, Yue JK, Deng H, et al. National trends in cerebral bypass surgery in the United States, 2002-2014. *Neurosurg Focus*. 2019;46(2):E4. doi:10.3171/2018.11.FOCUS18530

17. Ma Y, Gu Y, Tong X, et al. The Carotid and Middle cerebral artery Occlusion Surgery Study (CMOSS): a study protocol for a randomised controlled trial. *Trials*. 2016;17(1):544. doi:10.1186/s13063-016-1600-1

18. Chen A, Shyr MH, Chen TY, Lai HY, Lin CC, Yen PS. Dynamic CT perfusion imaging with acetazolamide challenge for evaluation of patients with unilateral cerebrovascular steno-occlusive disease. *AJNR Am J Neuroradiol*. 2006;27(9):1876-1881.

19. Latchaw RE, Yonas H, Hunter GJ, et al; Council on Cardiovascular Radiology of the American Heart Association. Guidelines and recommendations for

perfusion imaging in cerebral ischemia: a scientific statement for healthcare professionals by the writing group on perfusion imaging, from the Council on Cardiovascular Radiology of the American Heart Association. *Stroke*. 2003;34(4):1084-1104. doi:10.1161/01.STR.0000064840.99271.9E

20. Furie KL, Kasner SE, Adams RJ, et al; American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Clinical Cardiology, and Interdisciplinary Council on Quality of Care and Outcomes Research. Guidelines for the prevention of stroke in patients with stroke or transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(1):227-276. doi:10.1161/STR.0b013e3181f7d043

21. Kernan WN, Ovbiagele B, Black HR, et al; American Heart Association Stroke Council, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Peripheral Vascular Disease. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(7):2160-2236. doi:10.1161/STR.0000000000000024

22. Kuroda S, Houkin K, Kamiyama H, Mitsumori K, Iwasaki Y, Abe H. Long-term prognosis of medically treated patients with internal carotid or middle cerebral artery occlusion: can acetazolamide test predict it? *Stroke*. 2001;32(9):2110-2116. doi:10.1161/hs0901.095692

23. Ogasawara K, Ogawa A, Yoshimoto T. Cerebrovascular reactivity to acetazolamide and outcome in patients with symptomatic internal carotid or middle cerebral artery occlusion: a xenon-133 single-photon emission computed tomography study. *Stroke*. 2002;33(7):1857-1862. doi:10.1161/01.STR.0000019511.81583.A8

24. Yamauchi H, Fukuyama H, Nagahama Y, et al. Evidence of misery perfusion and risk for recurrent stroke in major cerebral arterial occlusive diseases from PET. *J Neurol Neurosurg Psychiatry*. 1996;61(1):18-25. doi:10.1136/jnnp.61.1.18

25. Reiff T, Eckstein H-H, Mansmann U, et al; SPACE-2 Investigators. Carotid endarterectomy or stenting or best medical treatment alone for moderate-to-severe asymptomatic carotid artery stenosis: 5-year results of a multicentre, randomised controlled trial. *Lancet Neurol*. 2022;21(10):877-888. doi:10.1016/S1474-4422(22)00290-3

26. Endarterectomy for asymptomatic carotid artery stenosis: executive committee for the Asymptomatic Carotid Atherosclerosis Study. *JAMA*. 1995;273(18):1421-1428. doi:10.1001/jama.1995.03520420037035

27. Chaturvedi S, Turan TN, Lynn MJ, et al; SAMMPRIS Trial Investigators. Do patient characteristics explain the differences in outcome between medically treated patients in SAMMPRIS and WASID? *Stroke*. 2015;46(9):2562-2567. doi:10.1161/STROKEAHA.115.009656

28. Gunawardena M, Rogers JM, Stoodley MA, Morgan MK. Revascularization surgery for symptomatic non-moyamoya intracranial arterial stenosis or occlusion. *J Neurosurg*. 2019;132(2):415-420. doi:10.3171/2018.9.JNS181075

ORIGINAL ARTICLE

Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality

The Global Cardiovascular Risk Consortium

ABSTRACT

BACKGROUND

Five modifiable risk factors are associated with cardiovascular disease and death from any cause. Studies using individual-level data to evaluate the regional and sex-specific prevalence of the risk factors and their effect on these outcomes are lacking.

METHODS

We pooled and harmonized individual-level data from 112 cohort studies conducted in 34 countries and 8 geographic regions participating in the Global Cardiovascular Risk Consortium. We examined associations between the risk factors (body-mass index, systolic blood pressure, non–high-density lipoprotein cholesterol, current smoking, and diabetes) and incident cardiovascular disease and death from any cause using Cox regression analyses, stratified according to geographic region, age, and sex. Population-attributable fractions were estimated for the 10-year incidence of cardiovascular disease and 10-year all-cause mortality.

RESULTS

Among 1,518,028 participants (54.1% of whom were women) with a median age of 54.4 years, regional variations in the prevalence of the five modifiable risk factors were noted. Incident cardiovascular disease occurred in 80,596 participants during a median follow-up of 7.3 years (maximum, 47.3), and 177,369 participants died during a median follow-up of 8.7 years (maximum, 47.6). For all five risk factors combined, the aggregate global population-attributable fraction of the 10-year incidence of cardiovascular disease was 57.2% (95% confidence interval [CI], 52.4 to 62.1) among women and 52.6% (95% CI, 49.0 to 56.1) among men, and the corresponding values for 10-year all-cause mortality were 22.2% (95% CI, 16.8 to 27.5) and 19.1% (95% CI, 14.6 to 23.6).

CONCLUSIONS

Harmonized individual-level data from a global cohort showed that 57.2% and 52.6% of cases of incident cardiovascular disease among women and men, respectively, and 22.2% and 19.1% of deaths from any cause among women and men, respectively, may be attributable to five modifiable risk factors. (Funded by the German Center for Cardiovascular Research (DZHK); ClinicalTrials.gov number, NCT05466825.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Dr. Blankenberg can be contacted at s.blankenberg@uke.de or at the University Heart and Vascular Center, Department for Cardiology, Center of Population Health Innovation, University Medical Center Hamburg–Eppendorf, Martinistr. 52, 20246 Hamburg, Germany.

A list of the investigators in the Global Cardiovascular Risk Consortium is provided in the Supplementary Appendix, available at NEJM.org.

Drs. Magnussen, Ojeda, and Leong contributed equally to this article.

This article was published on August 26, 2023, at NEJM.org.

This is the *New England Journal of Medicine* version of record, which includes all *Journal* editing and enhancements. The Author Accepted Manuscript, which is the author's version after external peer review and before publication in the *Journal*, is available at PubMed Central.

DOI: 10.1056/NEJMoa2206916

Cardiovascular diseases are the most common noncommunicable conditions worldwide and account for approximately one third of all deaths globally.[1] Modifiable risk factors such as body-mass index, systolic blood pressure, low-density lipoprotein cholesterol level, tobacco smoking, and diabetes account for a percentage of the prevalence and incidence of cardiovascular disease; however, the percentage varies according to the populations studied and the methods used.[2,3] These risk factors are used to derive contemporary risk scores[4-6] for the estimation of the 10-year risk of cardiovascular disease, although they are given different weights. These cardiovascular risk factors also have different associations with cardiovascular and noncardiovascular outcomes. Tobacco use is strongly associated with premature death, whereas elevated blood pressure and non–high-density lipoprotein (HDL) cholesterol level are more specifically related to cardiovascular disease.[7]

A tailored reduction in the burden of cardiovascular disease and death from any cause for persons and populations can be achieved with a better understanding of the region- and sex-specific associations of these cardiovascular risk factors with the development of cardiovascular disease. The Global Cardiovascular Risk Consortium analyzed a global harmonized individual-level data set of population-based cohorts to overcome the limitations of summary data and methodologic heterogeneity.

Methods

Study Design and Oversight

The study was designed by the Global Cardiovascular Risk Consortium Management Group, whose members are listed in the Supplementary Appendix, available with the full text of this article at NEJM.org. Data were collected by the Hamburg Data Center. Analyses were performed by the second author and reviewed by the Global Cardiovascular Risk Consortium Statistical Working Group. The first draft of the manuscript was prepared by the first, second, and last authors and was reviewed and edited by all the authors. The authors jointly agreed to submit the manuscript for publication and vouch for the accuracy and completeness of the data and for the fidelity of the study to the protocol.

Study Population

We pooled and harmonized individual-level data from 1,518,028 participants in 112 cohort studies conducted in eight geographic regions (North America, Latin America, Western Europe, Eastern Europe and Russia, North Africa and the Middle East, sub-Saharan Africa, Asia, and Australia) participating in the Global Cardiovascular Risk Consortium. Data were harmonized by applying the variable definitions used by the MORGAM (MONICA [Multinational Monitoring of Trends and Determinants in Cardiovascular Diseases] Risk, Genetics, Archiving, and Monograph) project.[8] Investigators in the studies that were not part of the MORGAM project were sent a list of study-related variables with definitions and were asked to provide these data. A description of each study cohort and information about the local ethics committees and participant informed consent are provided in the Supplementary Appendix. The cohorts that were included in the Global Cardiovascular Risk Consortium were selected on the basis of literature review, existing collaborations among investigators, and the availability of the variables of interest (Table S1). The flow of participants through the study is shown in Figure S1 in the Supplementary Appendix.

Cardiovascular Risk Factors and Outcome Definitions

Five risk factors (body-mass index, systolic blood pressure, non-HDL cholesterol, current smoking, and diabetes) and two outcomes (cardiovascular disease and death from any cause) were assessed in the study because of the heterogeneity of the effects of the risk factors on outcomes and the widespread availability of these data in the population. Further, these risk factors can be modified with interventions. Information on these five modifiable risk factors was collected at baseline according to the protocols of the respective studies included in the Global Cardiovascular Risk Consortium. The standardized definitions that were used to classify cardiovascular disease events are provided in Table S2, and the representativeness of the study population is shown in Table S3.

Statistical Analysis

Missing data were imputed by means of multiple imputation with chained equations (Table S4).[9] Both crude and age- and sex-standardized base-

line characteristics were calculated according to region with the use of direct standardization, with the age and sex distribution of the Global Cardiovascular Risk Consortium data set as the standard. Age-standardized event rates stratified according to region were also estimated and reported per 1000 person-years. Cumulative incidence curves were generated for cardiovascular disease and death from any cause. Associations between risk factors and outcome events were evaluated with the use of a two-stage, multivariate, random-effects meta-analysis of individual participant data.[10] Sex-specific Cox models, with age as the time scale,[11] were computed for each study, and coefficients were pooled across studies according to region as well as globally. Covariates (body-mass index, systolic blood pressure, non-HDL cholesterol level, current smoking, diabetes, and the use of antihypertensive medications) were included simultaneously in the models. Both linear and restricted cubic spline models for continuous covariates and models that allowed for time-varying effects were performed. Models that included receipt of lipid-lowering medications as an additional covariate were also computed with the use of data from the studies in which this information was available (these data were missing for approximately 20% of participants).

For the five modifiable risk factors, region- and sex-specific population-attributable fractions were estimated for the 10-year incidence of cardiovascular disease and 10-year all-cause mortality (see the Supplementary Appendix). Population-attributable fraction is an estimate of the proportion of an outcome that could be prevented if the value of a risk factor were replaced by a hypothetical, ideal value. The approach used by Laaksonen and colleagues,[12] which takes into account the time-to-event nature of the data, was applied to calculate the population-attributable fractions. Weibull models were used in the estimation, and their distributional assumptions were assessed graphically. Reference categories for the risk factors are provided in the Supplementary Appendix.

All the models that were used in the analyses of associations and population-attributable fractions were run with the exclusion of first-year follow-up data (1-year landmark analysis). Two-year landmark analyses that excluded data from the first 2 years of follow-up were performed as sensitivity analyses. The widths of the confidence intervals have not been adjusted for multiple comparisons and should not be used in place of hypothesis testing. All statistical analyses were performed with the use of R software, version 4.1.3.[13]

RESULTS

PARTICIPANT CHARACTERISTICS AND PREVALENCE OF RISK FACTORS

The baseline examination for all study cohorts included in the Global Cardiovascular Risk Consortium took place between 1963 and 2020. In the age- and sex-standardized analysis of data from 1,518,028 participants (54.1% of whom were women) with a median age of 54.4 years, the median body-mass index (the weight in kilograms divided by the square of the height in meters) was 26.4 (interquartile range, 23.7 to 29.7), the median systolic blood pressure 130 mm Hg (interquartile range, 118 to 144), and the median non-HDL cholesterol level 156.9 mg per deciliter (4.06 mmol per liter; interquartile range, 128.8 to 187.9 mg per deciliter [3.33 to 4.86 mmol per liter]); 21.6% were current smokers and 8.3% had diabetes. The age- and sex-standardized prevalence of the five risk factors and the use of antihypertensive and lipid-lowering medications across geographic regions are shown in Table 1 and Table S5. Baseline characteristics without age and sex standardization are shown in Table S6, and the distributions of the risk factors according to sex are shown in Tables S7 and S8. The prevalence of modifiable risk factors across contemporary national health examination surveys, which were used in the population-attributable fraction analyses, is shown in Tables S9, S10, and S11.

CARDIOVASCULAR DISEASE AND DEATH FROM ANY CAUSE

The median duration of follow-up among participants was 7.3 years (interquartile range, 5.9 to 11.8; maximum, 47.3) for incident cardiovascular disease and 8.7 years (interquartile range, 7.0 to 15.9; maximum, 47.6) for death from any cause. The follow-up times for each of the individual cohorts are provided in Table S12. A total of 80,596 cardiovascular disease events (30,033 in women and 50,563 in men) and 177,369 deaths from any cause (78,608 in women and 98,761 in men) were observed during the follow-up period

Table 1. Characteristics of the Cohort Studies and Age- and Sex-Standardized Characteristics of the Participants at Baseline According to Geographic Region.*

Characteristic	Global	North America	Latin America	Western Europe	Eastern Europe and Russia	North Africa and the Middle East	Sub-Saharan Africa	Asia	Australia
Cohort studies									
Cohort studies — no.	112	11	10	58	16	5	2	4	6
Participants — no.	1,518,028	65,182	191,244	907,760	51,133	185,608	10,390	59,802	46,909
Range of survey years†	1963–2020	1971–2011	1990–2013	1970–2015	1983–2014	1963–2020	2011–2017	1988–2015	1983–2007
Participants									
Median age (IQR) — yr‡	54.4 (4.2–63.0)	54.0 (45.0–63.0)	54.0 (45.0–63.0)	54.6 (45.5–63.0)	54.1 (45.5–63.0)	54.0 (45.0–62.6)	54.0 (45.0–63.0)	54.0 (45.0–63.0)	54.6 (45.5–63.0)
Male sex — %‡	45.9	45.9	45.9	45.9	45.9	45.9	45.9	45.9	45.9
Median BMI (IQR)	26.4 (23.7–29.7)	27.2 (24.1–31.0)	28.2 (25.4–31.5)	26.1 (23.6–29.2)	27.2 (24.3–30.6)	27.0 (24.0–30.3)	21.0 (19.0–23.4)	22.8 (20.5–25.2)	26.4 (23.7–29.5)
Median SBP (IQR) — mm Hg	130.0 (118.0–144.0)	122.0 (111.0–136.0)	126.7 (118.0–138.7)	134.0 (122.0–148.0)	132.0 (120.0–148.0)	115.0 (105.0–130.0)	125.0 (113.0–140.0)	123.5 (112.0–136.0)	127.0 (116.5–139.0)
Median DBP (IQR) — mm Hg	80.0 (72.0–87.5)	74.0 (67.0–81.0)	82.7 (76.7–90.0)	81.0 (74.0–89.0)	82.0 (75.0–91.0)	75.0 (67.5–80.0)	75.0 (69.0–83.0)	76.0 (68.0–84.0)	72.5 (64.5–80.5)
Median non-HDL cholesterol (IQR) — mg/dl	156.9 (128.8–187.9)	150.0 (123.0–179.4)	156.2 (131.1–185.2)	162.8 (134.8–193.8)	162.4 (135.0–191.8)	140.1 (115.3–167.8)	116.0 (77.3–154.7)	140.0 (117.6–167.0)	151.2 (124.5–181.0)
Current smoking — %	21.6	22.5	30.8	20.9	29.2	14.2	18.6	23.5	14.3
Diabetes — %	8.3	13.0	15.3	4.8	9.0	18.3	2.0	5.1	4.8
Antihypertensive medications — %	19.4	27.5	19.3	17.9	28.8	24.7	18.5	11.6	13.7
Lipid-lowering medications — %	9.6	8.0	2.3	11.5	8.8	11.6	NA	4.4	4.1
History of CVD — %	5.6	7.2	3.6	5.6	11.2	5.6	0	6.3	7.2

* Percentages are presented for binary variables. Percentages, medians, and interquartile ranges (IQRs) are based on data from the available number of cases per variable. Percentages, medians, and IQRs per geographic region were computed with the use of direct standardization according to age (≤40 years, >40 to ≤45 years, >45 to ≤50 years, >50 to ≤55 years, >55 to ≤60 years, >60 to ≤65 years, >65 to ≤70 years, and >70 years) and sex distribution in the Global Cardiovascular Risk Consortium data set. To convert the values for non–high-density lipoprotein (HDL) cholesterol to millimoles per liter, multiply by 0.02586. BMI denotes body-mass index, CVD cardiovascular disease, DBP diastolic blood pressure, NA data not available, and SBP systolic blood pressure.

† The approximate number of observations according to categorized examination year were 198,517 (13.1%) between 1963 and 1989, 227,002 (15.0%) between 1990 and 1999, 746,074 (49.1%) between 2000 and 2009, and 342,887 (22.8%) between 2010 and 2020.

‡ Similar values for the characteristics of age and sex across the geographic regions are due to the age and sex standardization.

Table 2. Age-Standardized Outcomes According to Geographic Region and Sex.*

Region	CVD Events among Women			CVD Events among Men			Deaths from Any Cause among Women			Deaths from Any Cause among Men		
	No. of Events	10-yr Incidence (95% CI)	Events per 1000 person-yr (95% CI)	No. of Events	10-yr Incidence (95% CI)	Events per 1000 person-yr (95% CI)	No. of Events	10-yr Incidence (95% CI)	Events per 1000 person-yr (95% CI)	No. of Events	10-yr Incidence (95% CI)	Events per 1000 person-yr (95% CI)
		%	*no.*		%	*no.*		%	*no.*		%	*no.*
Global	30,033	4.0 (4.0–4.1)	5.2 (5.2–5.3)	50,563	7.8 (7.7–7.9)	9.9 (9.8–10.0)	78,608	6.1 (6.0–6.2)	9.0 (9.0–9.1)	98,761	9.3 (9.2–9.4)	13.4 (13.3–13.4)
North America	4,702	7.4 (7.1–7.8)	10.1 (9.8–10.4)	5,321	13.7 (13.1–14.2)	16.6 (16.2–17.1)	8,674	7.6 (7.3–8.0)	16.7 (16.4–17.0)	8,128	11.3 (10.8–11.8)	21.4 (20.9–21.8)
Latin America†	71	—	2.4 (1.9–3.1)	89	—	4.1 (3.3–5.0)	12,488	6.8 (6.6–7.0)	7.5 (7.4–7.7)	9,733	9.7 (9.5–10.0)	10.7 (10.4–10.9)
Western Europe	22,212	3.7 (3.6–3.8)	4.9 (4.8–4.9)	40,942	7.3 (7.2–7.5)	9.6 (9.5–9.7)	42,676	5.6 (5.5–5.7)	8.9 (8.8–9.0)	59,447	8.4 (8.3–8.5)	12.7 (12.6–12.8)
Eastern Europe and Russia	1,078	5.7 (5.0–6.4)	8.7 (7.9–9.5)	1,508	9.9 (8.9–10.9)	13.5 (12.4–14.6)	3,255	10.1 (9.5–10.7)	12.7 (12.2–13.3)	4,827	17.9 (17.1–18.6)	22.1 (21.4–22.8)
North Africa and the Middle East	1,146	6.4 (5.6–7.2)	4.0 (3.7–4.3)	1,805	9.4 (8.5–10.2)	6.8 (6.5–7.2)	1,650	8.1 (7.2–9.0)	6.1 (5.7–6.5)	8,615	11.9 (11.3–12.5)	16.2 (15.8–16.6)
Sub-Saharan Africa‡	3	—	—	1	—	—	431	27.2 (16.0–36.9)	14.1 (12.7–15.6)	456	34.6 (25.1–42.9)	26.7 (24.1–29.6)
Asia	311	2.5 (1.8–3.2)	2.5 (2.2–2.9)	353	4.2 (3.0–5.4)	5.1 (4.5–5.9)	6,399	11.0 (9.9–12.0)	7.6 (7.1–8.1)	4,751	16.7 (15.1–18.4)	12.0 (11.2–13.0)
Australia	510	4.9 (4.3–5.5)	6.0 (5.4–6.6)	544	9.2 (8.3–10.1)	10.3 (9.4–11.4)	3,035	4.7 (4.5–5.0)	5.6 (5.4–5.9)	2,804	7.2 (6.8–7.6)	8.9 (8.5–9.3)

* Computations were performed with the use of data from 1,088,670 participants in the analysis of cardiovascular disease and from 1,419,699 participants in the analysis of all-cause mortality. The 10-year incidence of events was estimated with the use of the Kaplan–Meier estimator. Events per 1000 person-years were estimated with the use of data obtained during the complete follow-up, and a Poisson regression with log-transformed follow-up time was used as an offset. Direct standardization according to the age distribution in the Global Cardiovascular Risk Consortium data set (≤40 years, >40 to ≤45 years, >45 to ≤50 years, >50 to ≤55 years, >55 to ≤60 years, >60 to ≤65 years, >65 to ≤70 years, and >70 years) was used in the computation of the 10-year incidence of events and events per 1000 person-years per geographic region. The widths of the 95% confidence intervals have not been adjusted for multiplicity and should not be used in place of hypothesis testing.

† Because the CVD follow-up in Latin America is shorter than 10 years, it is not possible to obtain an estimate of the 10-year incidence of the event with the Kaplan–Meier estimator.

‡ Because of the low number of CVD events recorded, the 10-year incidence of events and events per 1000 person-years were not estimated.

(Table 2). The age-standardized 10-year incidence of cardiovascular disease was 7.4% for women and 13.7% for men in North America, 6.4% for women and 9.4% for men in North Africa and the Middle East, 5.7% for women and 9.9% for men in Eastern Europe and Russia, 3.7% for women and 7.3% for men in Western Europe, and 2.5% for women and 4.2% for men in Asia. The global 10-year incidence of cardiovascular disease was 4.0% among women and 7.8% among men (Table 2). Cardiovascular disease appeared to develop in women at older ages than men (Fig. S2). The age-standardized 10-year all-cause mortality was 27.2% for women and 34.6% for men in sub-Saharan Africa, 10.1% for women and 17.9% for men in Eastern Europe and Russia, 11.0% for women and 16.7% for men in Asia, and 4.7% for women and 7.2% for men in Australia (Table 2).

EFFECTS OF MODIFIABLE RISK FACTORS

The risk factor–associated hazard ratios for cardiovascular disease and death from any cause according to geographic region and sex, as calculated with the exclusion of data from the first year of follow-up (1-year landmark analysis), are shown in Table S13 and Figures S3 through S7. Subdistribution hazard ratios for cardiovascular disease were calculated with death from noncardiovascular causes as the competing event, and the results were similar to those in the 1-year landmark analysis (Table S14). The associations between the risk factors and cardiovascular disease and death from any cause in the models that used continuous risk factors and allowed for nonlinear effects are shown in Figure 1 and Figure S4. In a 2-year landmark analysis that excluded data from the first 2 years of follow-up, the observed associations appeared to be similar to those in the 1-year landmark analysis (Fig. S5), as were the results of other sensitivity analyses (2-year landmark Cox regression models with a linear exposure–effect assumption, models restricted to cohorts starting in the year 2000 or later, models restricted to participants with data on the use of lipid-lowering medications, and

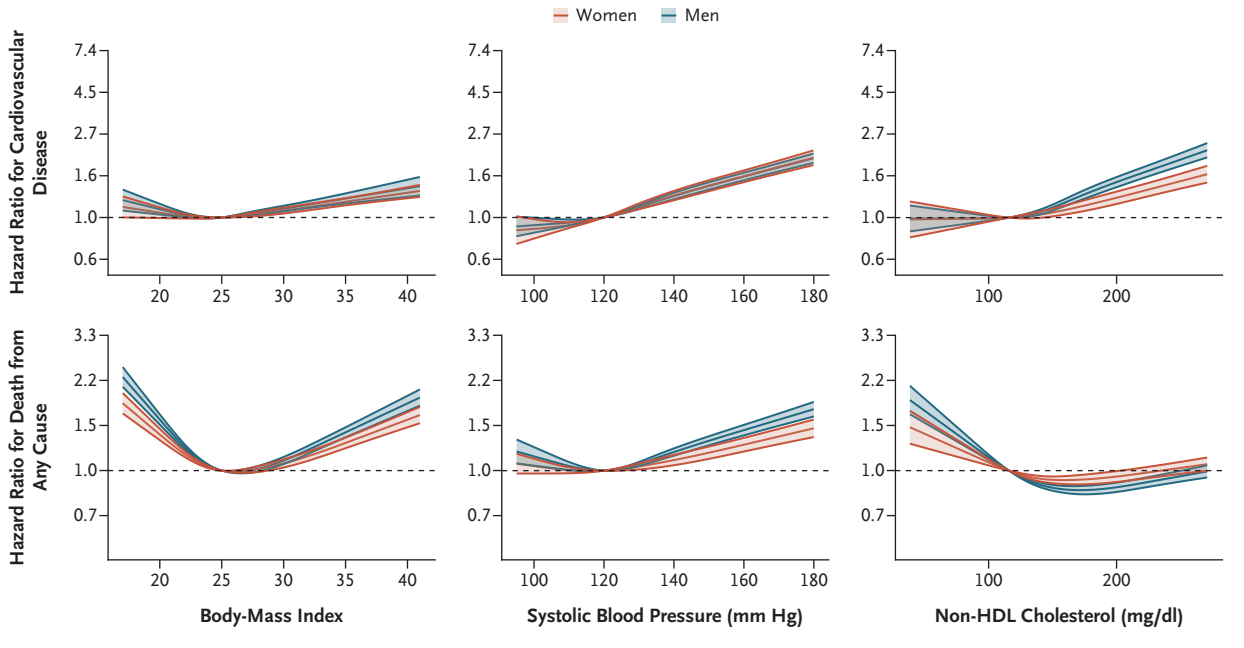


Figure 1. Associations of Continuous Risk Factors with Cardiovascular Disease and Death from Any Cause.

Shown are the results of a global 1-year landmark analysis that allowed for nonlinear effects. Participants with cardiovascular disease at baseline were excluded. Age was used as the time scale. All five risk factors, together with the use of antihypertensive medications, were included as covariates in the models. The widths of the 95% confidence intervals (shaded areas) have not been adjusted for multiplicity and should not be used in place of hypothesis testing. To convert the values for non–high-density lipoprotein (HDL) cholesterol to millimoles per liter, multiply by 0.02586.

models with an alternative definition of cardiovascular disease [a composite of fatal or nonfatal myocardial infarction, ischemic or hemorrhagic stroke, or cardiovascular death]) (Tables S15 through S18 and Fig. S6). Unadjusted risk factor–associated hazard ratios for cardiovascular disease and death from any cause are provided in Table S19. For both cardiovascular disease and death from any cause, the association with body-mass index appeared to be consistent across all ages, whereas the strength of the associations with systolic blood pressure, current smoking (after a steady increase up to the second half of life for death from any cause), and diabetes decreased with age (Fig. 2 and Fig. S7). The strength of the association between non-HDL cholesterol level and cardiovascular disease seemed to decline with age but appeared to be stable for death from any cause.

PREVENTABLE CARDIOVASCULAR DISEASE AND DEATH FROM ANY CAUSE

The distributional assumptions of the models used in the estimations of the population-attributable fractions were examined graphically (Fig. S8). The five modifiable risk factors accounted for an aggregate global population-attributable fraction of the 10-year incidence of cardiovascular disease of 57.2% (95% confidence interval [CI], 52.4 to 62.1) among women and 52.6% (95% CI, 49.0 to 56.1) among men. In comparison, the five risk factors accounted for an aggregate global population-attributable fraction of 10-year all-cause mortality of 22.2% (95% CI, 16.8 to 27.5) among women and 19.1% (95% CI, 14.6 to 23.6) among men (Fig. 3).

For all modifiable risk factors combined, the aggregate population-attributable fraction of the 10-year incidence of cardiovascular disease was

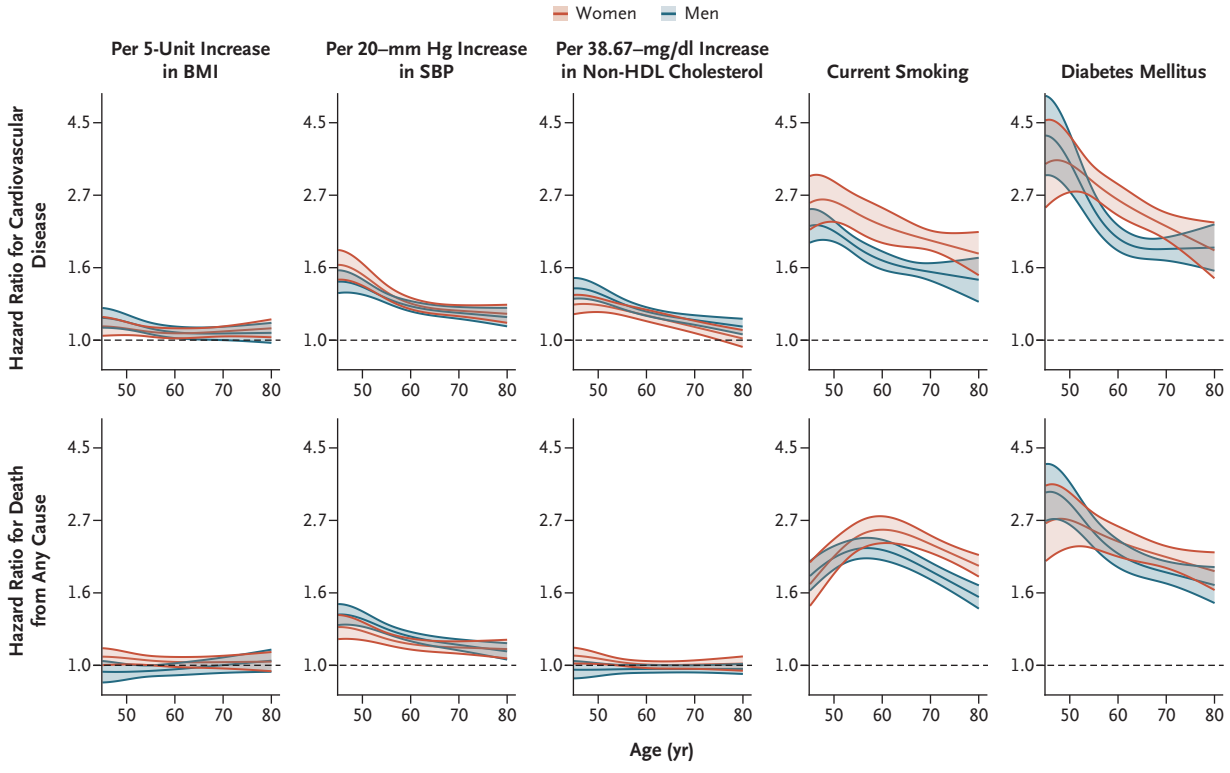


Figure 2. Associations of Risk Factors with Cardiovascular Disease and Death from Any Cause.

Shown are the results of a global 1-year landmark analysis that allowed for effects to change with age. Participants with cardiovascular disease at baseline were excluded. Age was used as the time scale. All five risk factors, together with the use of antihypertensive medications, were included as covariates in the models. The widths of the 95% confidence intervals (shaded areas) have not been adjusted for multiplicity and should not be used in place of hypothesis testing. BMI denotes body-mass index, and SBP systolic blood pressure.

64.2% (95% CI, 59.8 to 68.6) among women and 60.5% (95% CI, 57.2 to 63.9) among men in North Africa and the Middle East and 55.5% (95% CI, 50.7 to 60.3) and 50.3% (95% CI, 46.8 to 53.8), respectively, among those in North America. The aggregate population-attributable fraction of 10-year all-cause mortality was 34.3% (95% CI, 29.7 to 38.9) among women and 43.2% (95% CI, 39.8 to 46.6) among men in Asia; 13.7% (95% CI, 7.1 to 20.3) and 2.9% (95% CI, −3.7 to 9.5), respectively, among those in Australia; and 15.7% (95% CI, 9.3 to 22.0) and 2.1% (95% CI, −4.3 to 8.6), respectively, among those in Western Europe (Fig. 3).

The population-attributable fractions calculated for the individual five modifiable risk factors are shown in Figure S9. The population-attributable fraction of the 10-year incidence of cardiovascular disease associated with systolic blood pressure was 29.3% (95% CI, 25.4 to 33.2) among women and 21.6% (95% CI, 18.7 to 24.5) among

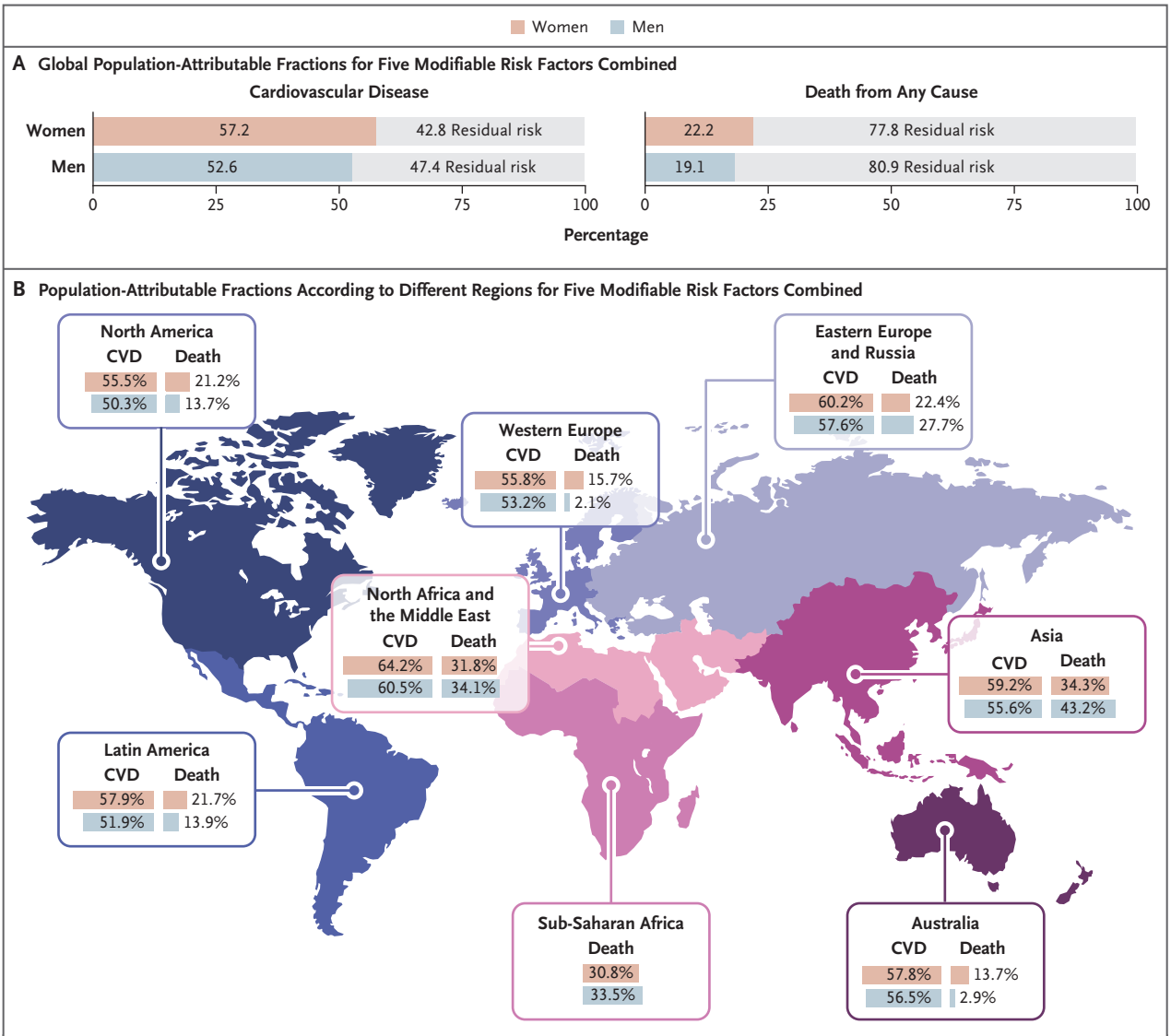


Figure 3. Population-Attributable Fractions of the 10-Year Incidence of Cardiovascular Disease (CVD) and 10-Year All-Cause Mortality.

Population-attributable fractions were estimated with the use of Weibull models. For the five modifiable risk factors combined, the aggregate global population-attributable fractions are shown in Panel A and the aggregate population-attributable fractions according to geographic region are shown in Panel B. Population-attributable fractions for the individual risk factors according to geographic region are provided in Figure S9 in the Supplementary Appendix.

men; the corresponding values were 15.4% (95% CI, 10.9 to 19.8) and 16.6% (95% CI, 12.6 to 20.6) for cardiovascular disease associated with non-HDL cholesterol level and 15.2% (95% CI, 13.3 to 17.1) and 10.2% (95% CI, 9.2 to 11.2) for cardiovascular disease associated with diabetes. The population-attributable fraction of the 10-year incidence of cardiovascular disease among women was 6.7% (95% CI, 5.8 to 7.6) for current smoking and 7.6% (95% CI, 5.1 to 10.1) for body-mass index; the corresponding values among men were 10.7% (95% CI, 9.6 to 11.7) and 7.6% (95% CI, 5.6 to 9.7). Among women, the population-attributable fraction of 10-year all-cause mortality associated with diabetes was 12.2% (95% CI, 11.1 to 13.3), whereas among men, the population-attributable fraction of 10-year all-cause mortality associated with current smoking was 14.4% (95% CI 13.3 to 15.4). The population-attributable fractions of the 10-year incidence of cardiovascular disease and 10-year all-cause mortality, according to modifiable risk factor level or status, are shown in Tables S20 and S21.

DISCUSSION

The Global Cardiovascular Risk Consortium harmonized individual-level data from 1,518,028 participants who participated in 112 cohort studies conducted in 34 countries in North America, Latin America, Western Europe, Eastern Europe and Russia, North Africa and the Middle East, sub-Saharan Africa, Asia, and Australia to assess the effect of five modifiable risk factors on the incidence of cardiovascular disease and all-cause mortality. The study showed that the prevalence of the five modifiable risk factors and the incidence of cardiovascular disease and all-cause mortality varied across geographic regions worldwide, and women had consistently lower event rates than men. The association between individual modifiable risk factors and both incident cardiovascular disease and death from any cause also varied across regions. The five modifiable risk factors accounted for an aggregate population-attributable fraction of the 10-year incidence of cardiovascular disease of 57.2% among women and 52.6% among men, and the corresponding values for 10-year all-cause mortality were 22.2% and 19.1%. Population-attributable fractions of the incidence of cardiovascular disease and all-cause mortality varied according to geographic region. Elevated systolic blood pressure appeared to be the largest contributor to the population-attributable fraction of incident cardiovascular disease events in all regions.

The Global Cardiovascular Risk Consortium and other studies[14-16] confirmed apparent differences in cardiovascular risk factor profile and event rates between women and men, irrespective of geographic region. Differences in risk factor level have been shown to translate into lifetime risk of cardiovascular disease[17] but not necessarily to affect other fatal outcomes. Cardiovascular risk factors are known to increase cardiovascular disease risk differently across various geographic regions.[2,18] Among them, high blood pressure is associated with up to 13.5% of all deaths annually worldwide and is considered to be the leading risk factor for cardiovascular disease.[19] Strict blood-pressure control to a systolic blood pressure of less than 120 mm Hg has been associated with lower rates of cardiovascular events and all-cause mortality.[20] Our data corroborate this observation; of the five risk factors studied, systolic blood pressure may offer the greatest potential for the prevention of cardiovascular disease. Although there is a strong continuous association between non-HDL cholesterol level and incident cardiovascular disease,[21] we and others[3,22,23] observed an inverted J-shaped association of non-HDL cholesterol level with all-cause mortality. Although very low levels of non-HDL cholesterol are related to a reduction in cardiovascular disease events,[24,25] some observations point toward higher all-cause mortality among participants with very low levels, at least in longer-term follow-up.[26] In contrast to what was previously reported,[3] body-mass index and current smoking (at least in some parts of the world) were associated with comparatively modest population-attributable fractions of cardiovascular disease events in the populations participating in the Global Cardiovascular Risk Consortium. These associations may be related to underlying differences in population characteristics, risk-factor definition and prevalence, or methods used to estimate population-attributable fractions.

Case–control studies such as INTERHEART may have overestimated the population-attributable fraction of the incidence of cardiovascular disease subtypes by attributing 90% of the risk of myocardial infarction to nine targetable risk factors.[2] Data from 155,722 participants who were

evaluated prospectively in the Prospective Urban Rural Epidemiology (PURE) study suggested that 71% of cardiovascular disease cases are attributable to 14 potentially modifiable metabolic and behavioral risk factors, a result that is consistent with our findings.[3] Our study focused on five modifiable risk factors for which strict control could potentially prevent 57.2% of all cases of cardiovascular disease in women and 52.6% of all cases in men globally. The varying effect of individual risk factors across different regions could enable ranking and prioritization of risk-factor control for public health action within those regions. However, there is substantial scope for a more complete characterization of the risk of cardiovascular disease. Environmental and exposure-related factors such as physical activity,[2] alcohol consumption,[27] air pollution,[28] climate and noise,[29] educational level,[3] or psychosocial risk factors, including depression,[30] have an effect on the risk of cardiovascular disease. Biomarkers[31,32] and genetic variants most likely would add to the prediction of cardiovascular disease risk.

The analysis by the Global Cardiovascular Risk Consortium differs from other global initiatives that combine different data sources such as registries, population surveys, and health system administrative data to produce meta-analytic summaries.[33,34] The Global Cardiovascular Risk Consortium maintains a large and comprehensive database of harmonized, observational, individual-level, prospectively collected data. This database allows for multiple prespecified statistical analyses on large-scale individual-level data. This study relates major modifiable cardiovascular risk factors to the incidence of cardiovascular disease and all-cause mortality. The inclusion of cohorts with a large spectrum of follow-up times enabled robust sex-specific analyses and the evaluation of differences across geographic regions.

Our study has several limitations. The Global Cardiovascular Risk Consortium database includes cohorts with varying representativeness, data quality and quantity, dates of baseline assessments, follow-up times, end-point definitions, and use of clinical interventions. Variation in the adjudication of causes of death or surrogates of nonfatal myocardial infarction is plausible across regions, but an analysis that included the use of a secondary definition of cardiovascular disease that excluded unclassifiable death, unstable angina, and coronary revascularization did not change the results. Structured harmonization was used to reduce variation, and sensitivity analyses provided results similar to those for the overall study population. Standardized event rates should be interpreted as descriptive measures rather than actual incidences in a population. To overcome bias resulting from deaths from noncardiovascular diseases that were present at the time of the baseline examination, analyses were performed with the exclusion of first-year follow-up data. Information about modifiable risk factors was available from the baseline examination, and the effect of changes in exposure over time are not known; the analyses were not corrected for regression dilution bias. Residual confounding cannot be completely excluded. The effects of overweight and obesity may be mediated by hyperlipidemia, hypertension, and diabetes.[35] Models that included body-mass index, systolic blood pressure, and diabetes attribute the share of the effect of body-mass index to systolic blood pressure and diabetes, even if overweight or obesity is the real underlying cause. The definition of current smoking may not capture the entire spectrum and dose of tobacco exposure, and smoking cessation during follow-up might have led to an underestimation of tobacco smoking as a risk factor. It was also assumed that risk factor effects and prevalence within a region are homogeneous; however, intraregional differences may exist. Information about ethnic group is not provided, because definitions differed among the cohorts or collection of the variable was incomplete or not available to a comparable standard. The categorization of geographic regions by the World Health Organization and United Nations was adapted to accommodate cohort size and representativeness of a geographic region, so a different categorization of regions may produce different results.

In our study, harmonized individual-level data from the Global Cardiovascular Risk Consortium showed that 57.2% and 52.6% of cases of incident cardiovascular disease among women and men, respectively, and 22.2% and 19.1% of deaths from any cause among women and men, respectively, may be attributable to five modifiable risk factors. The prevalence and effect of these risk factors on the incidence of cardiovascular disease and all-cause mortality vary according to sex and geographic region.

Supported by the German Center for Cardiovascular Research (DZHK).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

APPENDIX

The authors' full names and academic degrees are as follows: Christina Magnussen, M.D., Francisco M. Ojeda, Ph.D., Darryl P. Leong, M.B., B.S., Ph.D., Jesus Alegre-Diaz, M.D., Philippe Amouyel, M.D., Ph.D., Larissa Aviles-Santa, M.D., Dirk De Bacquer, Ph.D., Christie M. Ballantyne, M.D., Antonio Bernabé-Ortiz, M.D., Ph.D., Martin Bobak, Ph.D., Hermann Brenner, M.D., Rodrigo M. Carrillo-Larco, M.D., Ph.D., James de Lemos, M.D., Annette Dobson, Ph.D., Marcus Dörr, M.D., Chiara Donfrancesco, Ph.D., Wojciech Drygas, M.D., Ph.D., Robin P. Dullaart, M.D., Ph.D., Gunnar Engström, M.D., Ph.D., Marco M. Ferrario, M.D., Ph.D., Jean Ferrières, M.D., Ph.D., Giovanni de Gaetano, M.D., Ph.D., Uri Goldbourt, Ph.D., Clicerio Gonzalez, M.D., Guido Grassi, M.D., Allison M. Hodge, Ph.D., Kristian Hveem, M.D., Ph.D., Licia Iacoviello, M.D., Ph.D., M. Kamran Ikram, M.D., Ph.D., Vilma Irazola, M.D., Ph.D., Modou Jobe, M.D., Pekka Jousilahti, M.D., Ph.D., Pontiano Kaleebu, M.D., Ph.D., Maryam Kavousi, M.D., Ph.D., Frank Kee, M.D., Davood Khalili, M.D., Ph.D., Wolfgang Koenig, M.D., Anna Kontsevaya, M.D., Ph.D., Kari Kuulasmaa, Ph.D., Karl J. Lackner, M.D., David M. Leistner, M.D., Lars Lind, M.D., Ph.D., Allan Linneberg, M.D., Ph.D., Thiess Lorenz, M.A., Magnus Nakrem Lyngbakken, M.D., Ph.D., Reza Malekzadeh, M.D., Sofia Malyutina, M.D., Ph.D., Ellisiv B. Mathiesen, M.D., Ph.D., Olle Melander, M.D., Ph.D., Andres Metspalu, M.D., Ph.D., J. Jaime Miranda, M.D., Ph.D., Marie Moitry, M.D., Joseph Mugisha, Ph.D., Mahdi Nalini, M.D., Ph.D., Vijay Nambi, M.D., Ph.D., Toshiharu Ninomiya, M.D., Ph.D., Karen Oppermann, M.D., Ph.D., Eleonora d'Orsi, Ph.D., Andrzej Pająk, M.D., Ph.D., Luigi Palmieri, Ph.D., Demosthenes Panagiotakos, M.D., Ph.D., Arokiasamy Perianayagam, Ph.D., Annette Peters, Ph.D., Hossein Poustchi, M.D., Ph.D., Andrew M. Prentice, Ph.D., Eva Prescott, M.D., Ulf Risérus, M.D., Ph.D., Veikko Salomaa, M.D., Ph.D., Susana Sans, M.D., Ph.D., Satoko Sakata, M.D., Ph.D., Ben Schöttker, Ph.D., Aletta E. Schutte, Ph.D., Sadaf G. Sepanlou, Ph.D., Sanjib Kumar Sharma, M.D., Jonathan E. Shaw, M.D., Leon A. Simons, M.D., Stefan Söderberg, M.D., Ph.D., Abdonas Tamosiunas, M.D., Barbara Thorand, Ph.D., Hugh Tunstall-Pedoe, M.D., Raphael Twerenbold, M.D., Diego Vanuzzo, M.D., Giovanni Veronesi, Ph.D., Julia Waibel, M.Sc., S. Goya Wannamethee, Ph.D., Masafumi Watanabe, M.D., Ph.D., Philipp S. Wild, M.D., Yao Yao, M.D., Ph.D., Yi Zeng, M.D., Andreas Ziegler, Ph.D., and Stefan Blankenberg, M.D.

The authors' affiliations are as follows: the Center for Population Health Innovation, University Heart and Vascular Center Hamburg (C.M., F.M.O., T.L., R.T., J.W., A.Z., S.B.), University Medical Center Hamburg–Eppendorf (C.M., F.M.O., T.L., R.T., J.W., A.Z., S.B.), the German Center for Cardiovascular Research (DZHK) Partner Site Hamburg–Kiel–Lübeck (C.M., T.L., R.T., J.W., S.B.), the Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, and Network Aging Research, Heidelberg University, Heidelberg (H.B., B.S.), the Department of Internal Medicine B, University Medicine Greifswald, and DZHK Partner Site Greifswald, Greifswald (M.D.), DZHK Partner Site Munich Heart Alliance (W.K., A. Peters), the German Heart Center, Technical University of Munich (W.K.), and the Institute for Medical Information Processing, Biometry, and Epidemiology, Medical Faculty, Ludwig Maximilians Universität München (A. Peters, B.T.), Munich, the Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm (W.K.), the Institute of Clinical Chemistry and Laboratory Medicine (K.J.L.), Preventive Cardiology and Preventive Medicine (P.S.W.), and Clinical Epidemiology and Systems Medicine, Center for Thrombosis and Hemostasis (P.S.W.), University Medical Center of the Johannes Gutenberg–University Mainz, DZHK Partner Site Rhine–Main (K.J.L., P.S.W.), and Institute for Molecular Biology (P.S.W.), Mainz, the University Heart and Vascular Center Frankfurt, DZHK Partner Site Rhine–Main, Frankfurt (D.M.L.), the Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (A. Peters, B.T.), and German Center for Diabetes Research (DZD) Partner Site Munich–Neuherberg (A. Peters, B.T.), Neuherberg — all in Germany; the Department of Medicine (Cardiology), McMaster University, Hamilton, Canada (D.P.L.); the Experimental Medicine Research Unit, School of Medicine, National Autonomous University of Mexico, Mexico City (J.A.-D.), and Centro de Estudios en Diabetes, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Cuernavaca (C.G.) — both in Mexico; Université de Lille, INSERM, Centre Hospitalier University de Lille, Institut Pasteur de Lille, UMR1167–RID-AGE–Risk Factors and Molecular Determinants of Aging-Related Diseases, Epidemiology and Public Health Department, Lille (P.A.), the Department of Cardiology, INSERM UMR1295, Toulouse Rangueil University Hospital, Toulouse (J.F.), and the Department of Public Health, Strasbourg University Hospital, University of Strasbourg, Strasbourg (M.M.) — all in France; the Division of Clinical and Health Services Research, National Institute on Minority Health and Health Disparities (L.A.-S.), and the Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute (M.N.), National Institutes of Health, Bethesda, MD; the Department of Public Health and Primary Care, Ghent University, Ghent, Belgium (D.D.B.); the Department of Medicine, Baylor College of Medicine (C.M.B.), and Michael E. DeBakey Veterans Affairs Hospital and Baylor College of Medicine (V.N.), Houston, and the Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas (J.L.) — all in Texas; the CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru (A.B.-O., J.J.M.); the Department of Epidemiology and Public Health (M.B.) and the Research Department of Primary Care and Population Health (S.G.W.), University College London, London, the Centre for Public Health, Queens University Belfast, Belfast (F.K.), and the Cardiovascular Epidemiology Unit, Institute of Cardiovascular Research, University of Dundee, Dundee (H.T.-P.) — all in the United Kingdom; the Emory Global Diabetes Research Center and Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta (R.M.C.-L.); the School of Public Health, University of Queensland, Brisbane (A.D.), the Division of Cancer Epidemiology, Cancer Council Victoria (A.M.H.), the Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, the University of Melbourne (A.M.H.), and Baker Heart and Diabetes Institute (J.E.S.), Melbourne, the Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney (J.J.M.), and the George Institute for Global Health (A.E.S.), Sydney, and the School of Population Health (A.E.S.), University of New South Wales, Kensington (L.A.S.) — all in Australia; the Department of Cardiovascular and Endocrine–Metabolic Diseases and Aging, Istituto Superiore di Sanità, Rome (C.D., L.P.), the Research Center in Epidemiology and Preventive Medicine, Department of Medicine and Surgery, University of Insubria, Varese (M.M.F., L.I., G.V.), the Department of Epidemiology and Prevention, IRCCS Neuromed, Pozzilli (G. de Gaetano, L.I.), Clinica Medica, University of Milano–Bicocca, Milan (G. Grassi), and the MONICA (Monitoring Cardiovascular Diseases)–Friuli Study Group, Udine (D.V.) — all in Italy; the Department of Epidemiology, Cardiovascular Disease Prevention and Health Promotion, National Institute of Cardiology, and Lazarski University, Warsaw (W.D.), and the Department of Epidemiology and Population Studies, Institute of Public Health, Faculty of Health Sciences, Jagiellonian University Medical College, Krakow (A. Pajak) — all in Poland; the Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen (R.P.D.), and the Departments of Neurology (M.K.I.) and Epidemiology (M.K.I., M.K.), Erasmus University Medical Center Rotterdam, Rotterdam — both in the Netherlands; the Department of Clinical Sciences Malmö, Lund University, Malmö (G.E., O.M.), the Department of Public Health and Caring Sciences, Clinical Nutrition and Metabolism, Uppsala University (U.R.), and the Department of Medical Sciences, Uppsala University, Uppsala (L.L.), and the Department of Public Health and Clinical Medicine, University of Umea, Umea (S. Söderberg) — all in Sweden; the Department of Epidemiology, Tel Aviv University School of Public Health, Tel Aviv, Israel (U.G.); HUNT (Trøndelag Health Study)

Research Center, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger (K.H.), the K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Trondheim(K.H.), the Department of Cardiology, Division of Medicine, Akershus University Hospital, Lørenskog (M.N.L.), the K.G. Jebsen Center for Cardiac Biomarkers, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo (M.N.L.), and the Department of Clinical Medicine, University of Tromsø the Arctic University of Norway, Tromsø (E.B.M.) — all in Norway; the Department of Chronic Diseases, Institute for Clinical Effectiveness and Health Policy, Buenos Aires (V.I.); Medical Research Council Unit The Gambia at London School of Hygiene and Tropical Medicine, Banjul, Gambia (M.J., A.M.P.); the Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki (P.J., K.K., V.S.); Medical Research Council–Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit, Entebbe, Uganda (P.K., J.M.); the Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences (D.K.), the Liver and Pancreaticobiliary Disease Research Center (R.M.), the Digestive Oncology Research Center (R.M., H.P.), and the Digestive Disease Research Center (R.M., M.N., H.P., S.G.S.), Digestive Disease Research Institute, Shariati Hospital, Tehran University of Medical Sciences — all in Tehran, Iran; the National Research Center for Therapy and Preventive Medicine, Ministry of Healthcare of the Russian Federation, Moscow (A.K.), and the Research Institute of Internal and Preventive Medicine–Branch of the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk (S.M.) — both in Russia; the Department of Clinical Medicine, Faculty of Health and Medical Sciences (A.L.), and the Department of Cardiology, Bispebjerg Hospital (E.P.), University of Copenhagen, and the Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital (A.L.) — both in Copenhagen; the Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia (A.M.); the Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka (T.N., S. Sakata), and the Global Center of Excellence Program Study Group, Yamagata University School of Medicine, Yamagata (M.W.) — both in Japan; the School of Medicine, University of Passo Fundo, Passo Fundo (K.O.), and the Department of Public Health, Postgraduate Program in Public Health, Federal University of Santa Catarina, Florianopolis (E.O.) — both in Brazil; the School of Health Sciences and Education, Harokopio University, Athens, Greece (D.P.); the National Council of Applied Economic Research, Delhi (A. Perianayagam), and the International Institute for Population Sciences, Mumbai (A. Perianayagam) — both in India; the Catalan Department of Health, Barcelona (S. Sans); the Hypertension in Africa Research Team, South African Medical Research Council Unit for Hypertension and Cardiovascular Disease, North-West University, Potchefstroom (A.E.S.), and the School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Pietermaritzburg (A.Z.) — both in South Africa; the Department of Internal Medicine, B.P. Koirala Institute of Health Sciences, Dharan, Nepal (S.K.S.); the Laboratory of Population Studies, Institute of Cardiology, and the Department of Preventive Medicine, Faculty of Public Health, Lithuanian University of Health Sciences, Kaunas, Lithuania (A.T.); the China Center for Health Development Studies (Y.Y.) and Center for Healthy Aging and Development Studies, National School of Development (Y.Z.), Peking University, and the Key Laboratory of Epidemiology of Major Diseases, Peking University, Ministry of Education (Y.Y.) — both in Beijing; the Center for the Study of Aging and Human Development and Geriatrics Division, Medical School of Duke University, Durham, NC (Y.Z.); and Cardio-CARE, Davos (A.Z., S.B.), and the Swiss Institute of Bioinformatics, Lausanne (A.Z.) — both in Switzerland.

REFERENCES

1. Joseph P, Leong D, McKee M, et al. Reducing the global burden of cardiovascular disease, part 1: the epidemiology and risk factors. Circ Res 2017;121:677-94.

2. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364:937-52.

3. Yusuf S, Joseph P, Rangarajan S, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet 2020;395:795-808.

4. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021;42:2439-54.

5. Yadlowsky S, Hayward RA, Sussman JB, McClelland RL, Min YI, Basu S. Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. Ann Intern Med 2018;169:20-9.

6. World Health Organization. World health statistics 2011 (https://web.archive.org/web/20111116011154/http://www.who.int/whosis/whostat/2011/en/).

7. Jaspers NEM, Blaha MJ, Matsushita K, et al. Prediction of individualized lifetime benefit from lowering, blood pressure lowering, antithrombotic therapy, and smoking cessation in apparently healthy people. Eur Heart J 2020;41:1190-9.

8. Evans A, Salomaa V, Kulathinal S, et al. MORGAM (an international pooling of cardiovascular cohorts). Int J Epidemiol 2005;34:21-7.

9. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45:1-67.

10. Riley RD, Tierney JF, Stewart LA. Individual participant data meta-analysis: a handbook for healthcare research. Medford, MA: Wiley, 2021.

11. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. Am J Epidemiol 1997;145:72-80.

12. Laaksonen MA, Virtala E, Knekt P, Oja H, Härkänen T. SAS macros for calculation of population attributable fraction in a cohort study design. J Stat Softw 2011;43:1-25.

13. R Foundation for Statistical Computing. The R project for statistical computing (http://www.r-project.org/index.html).

14. Pencina MJ, D'Agostino RB Sr, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the Framingham heart study. Circulation 2009;119:3078-84.

15. Joseph P, Kutty VR, Mohan V, et al. Cardiovascular disease, mortality, and their associations with modifiable risk factors in a multi-national South Asia cohort: a PURE substudy. Eur Heart J 2022;43:2831-40.

16. Walli-Attaei M, Joseph P, Rosengren A, et al. Variations between women and men in risk factors, treatments, cardiovascular disease incidence, and death in 27 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet 2020;396:97-109.

17. Berry JD, Dyer A, Cai X, et al. Lifetime risks of cardiovascular disease. N Engl J Med 2012;366:321-9.

18. Ueshima H, Sekikawa A, Miura K, et al. Cardiovascular disease and risk factors in Asia: a selected review. Circulation 2008;118:2702-9.

19. Lawes CM, Vander Hoorn S, Rodgers A; International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. Lancet 2008;371:1513-8.

20. SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. N Engl J Med 2015;373:2103-16.

21. Brunner FJ, Waldeyer C, Ojeda F, et al. Application of non-HDL cholesterol for

population-based cardiovascular risk stratification: results from the Multinational Cardiovascular Risk Consortium. Lancet 2019;394:2173-83.
22. Johannesen CDL, Langsted A, Mortensen MB, Nordestgaard BG. Association between low density lipoprotein and all cause and cause specific mortality in Denmark: prospective cohort study. BMJ 2020;371:m4266.
23. Liu Y, Liu F, Zhang L, et al. Association between low density lipoprotein cholesterol and all-cause mortality: results from the NHANES 1999-2014. Sci Rep 2021;11:22111.
24. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. N Engl J Med 2017;376:1713-22.
25. Robinson JG, Farnier M, Krempf M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. N Engl J Med 2015;372:1489-99.
26. van Bruggen FH, Luijendijk HJ. Evolocumab's long-term mortality risk unclear due to shortened follow-up of FOURIER. Am J Cardiovasc Drugs 2022; 22:5-8.
27. GBD 2020 Alcohol Collaborators. Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020. Lancet 2022; 400:185-235.
28. Schraufnagel DE, Balmes JR, Cowl CT, et al. Air Pollution and noncommunicable diseases: a review by the Forum of International Respiratory Societies' Environmental Committee, part 2: air pollution and organ systems. Chest 2019;155: 417-26.
29. Basner M, Babisch W, Davis A, et al. Auditory and non-auditory effects of noise on health. Lancet 2014;383:1325-32.
30. Walli-Attaei M, Rosengren A, Rangarajan S, et al. Metabolic, behavioural, and psychosocial risk factors and cardiovascular disease in women compared with men in 21 high-income, middle-income, and low-income countries: an analysis of the PURE study. Lancet 2022;400:811-21.
31. Blankenberg S, Salomaa V, Makarova N, et al. Troponin I and cardiovascular risk prediction in the general population: the BiomarCaRE consortium. Eur Heart J 2016;37:2428-37.
32. Neumann JT, Twerenbold R, Ojeda F, et al. Application of high-sensitivity troponin in suspected myocardial infarction. N Engl J Med 2019;380:2529-40.
33. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020;76:2982-3021.
34. Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The global burden of cardiovascular diseases and risk: a compass for future health. J Am Coll Cardiol 2022;80:2361-71.
35. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1·8 million participants. Lancet 2014;383:970-83.

SVN Stroke & Vascular Neurology

Associations of polygenic risk scores with risks of stroke and its subtypes in Chinese

Songchun Yang ,[1,2] Zhijia Sun,[1] Dong Sun,[1] Canqing Yu,[1,3,4] Yu Guo,[5] Dianjianyi Sun,[1,3,4] Yuanjie Pang,[1,4] Pei Pei ,[3] Ling Yang,[6,7] Iona Y Millwood,[6,7] Robin G Walters,[6,7] Yiping Chen,[6,7] Huaidong Du,[6,7] Yan Lu,[8] Sushila Burgess,[7] Daniel Avery ,[7] Robert Clarke ,[7] Junshi Chen,[9] Zhengming Chen,[7] Liming Li,[1,3,4] Jun Lv ,[1,3,4,10] On behalf of the China Kadoorie Biobank Collaborative Group

To cite: Yang S, Sun Z, Sun D, *et al*. Associations of polygenic risk scores with risks of stroke and its subtypes in Chinese. *Stroke & Vascular Neurology* 2023;**0**. doi:10.1136/svn-2023-002428

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/svn-2023-002428).

Received 28 February 2023
Accepted 11 August 2023

For numbered affiliations see end of article.

Correspondence to
Dr Jun Lv; lvjun@bjmu.edu.cn

ABSTRACT

Background and purpose Previous studies, mostly focusing on the European population, have reported polygenic risk scores (PRSs) might achieve risk stratification of stroke. We aimed to examine the association strengths of PRSs with risks of stroke and its subtypes in the Chinese population.

Methods Participants with genome-wide genotypic data in China Kadoorie Biobank were split into a potential training set (n=22 191) and a population-based testing set (n=72 150). Four previously developed PRSs were included, and new PRSs for stroke and its subtypes were developed. The PRSs showing the strongest association with risks of stroke or its subtypes in the training set were further evaluated in the testing set. Cox proportional hazards regression models were used to estimate the association strengths of different PRSs with risks of stroke and its subtypes (ischaemic stroke (IS), intracerebral haemorrhage (ICH) and subarachnoid haemorrhage (SAH)).

Results In the testing set, during 872 919 person-years of follow-up, 8514 incident stroke events were documented. The PRSs of any stroke (AS) and IS were both positively associated with risks of AS, IS and ICH ($p<0.05$). The HR for per SD increment (HR_{SD}) of PRS_{AS} was 1.10 (95% CI 1.07 to 1.12), 1.10 (95% CI 1.07 to 1.12) and 1.13 (95% CI 1.07 to 1.20) for AS, IS and ICH, respectively. The corresponding HR_{SD} of PRS_{IS} was 1.08 (95% CI 1.06 to 1.11), 1.08 (95% CI 1.06 to 1.11) and 1.09 (95% CI 1.03 to 1.15). PRS_{ICH} was positively associated with the risk of ICH (HR_{SD}=1.07, 95% CI 1.01 to 1.14). PRS_{SAH} was not associated with risks of stroke and its subtypes. The addition of current PRSs offered little to no improvement in stroke risk prediction and risk stratification.

Conclusions In this Chinese population, the association strengths of current PRSs with risks of stroke and its subtypes were moderate, suggesting a limited value for improving risk prediction over traditional risk factors in the context of current genome-wide association study under-representing the East Asian population.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Polygenic risk scores (PRSs) might achieve risk stratification of stroke.
⇒ Evidence from the East Asian population (including Chinese) is lacking.

WHAT THIS STUDY ADDS

⇒ The association strengths of current PRSs with risks of stroke and its subtypes were moderate in the Chinese population.
⇒ PRS for ischaemic stroke was positively associated with the risk of intracerebral haemorrhage.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ In the Chinese population, current PRSs might have limited value for improving stroke risk prediction over traditional risk factors.
⇒ Further studies are warranted to assess whether new PRSs based on larger genome-wide association study or other developing methods have considerable potential to translate into population health benefits.

INTRODUCTION

Stroke is one of the leading causes of death and disease burdens globally.[1] Stroke includes two main subtypes, such as ischaemic stroke (IS) and haemorrhagic stroke (HS). The latter could further be divided into intracerebral haemorrhage (ICH) and subarachnoid haemorrhage (SAH). With the accumulation of genomic data worldwide, the genetic background of stroke and its subtypes is gradually being revealed. Polygenic risk score (PRS), a method used to combine minor genetic effects across the whole genome, has been increasingly used in stroke research. Several studies based on European populations have developed PRSs for any stroke (AS) or IS and suggested their potential to improve risk prediction and risk stratification.[2–9] The incidence of stroke in China, especially ICH, is higher than in Western countries.[1] Recently, a PRS for AS was developed based on the Chinese population and showed similar

association strength in predicting the risk of IS and HS.[10] However, IS and HS might have different aetiological mechanisms.[11–13] Different stroke subtypes also have their specific genetic loci.[14] No study has specifically developed PRSs for subtypes of stroke in the Chinese population.

The present study was based on a subcohort with genomic data from the China Kadoorie Biobank (CKB). We aimed to examine the association strengths of PRSs with risks of stroke and its subtypes in the Chinese population.

METHODS

Participants

CKB is an ongoing prospective study with 512 724 participants aged 30–79 enrolled from five urban and five rural regions in China between 2004 and 2008. Details of the study have been described elsewhere.[15]

Among all CKB participants, there are 100 639 participants with genome-wide genotypic data. Of them, 24 657 participants were selected based on a case–control design nested within the cohort with the primary aim of studying CVD ('case–control samples'), which formed four matched-case-control training sets (figure 1A, online supplemental methods, tables 1 and 2). The other 75 982 participants were randomly selected from the entire CKB cohort ('population-based samples'); after excluding participants with self-reported coronary artery disease or stroke or transient ischaemic attack at baseline (n=3832), the remaining participants were used as a 'testing set' (n=72 150) (figure 1A, online supplemental methods).

Study design

The current study can be divided into four parts (figure 1B). (1) Validation of previous PRSs. Four previously reported stroke-related PRSs were selected for validation.[2 4 5 10] (2) Development of new PRSs. Clumping and thresholding ('C+T') and LDpred[16] were used to develop new PRSs for stroke and its subtypes based on two genome-wide association studies with large sample

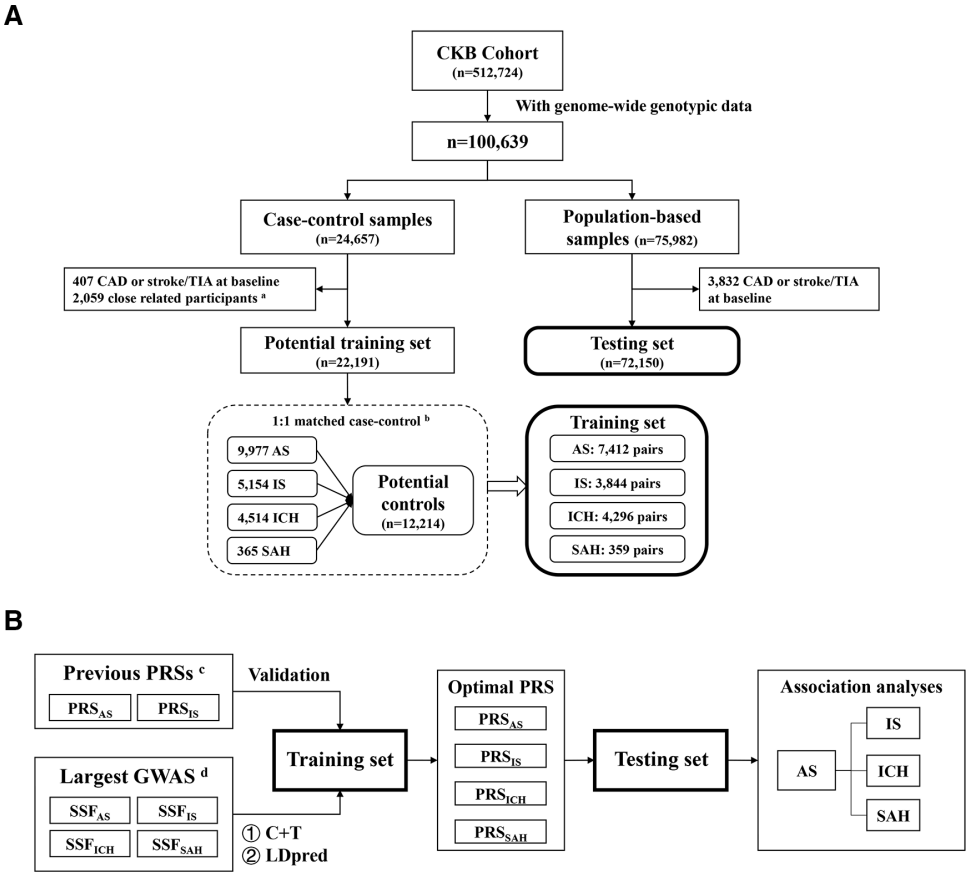


Figure 1 Overview of the present study. (A) Flow chart for the study population; (B) Study design. The current study can be divided into four parts: (1) validation of previous PRSs, (2) development of new PRSs, (3) identification of the optimal PRS for each outcome and (4) validation and evaluation of the optimal PRS for each outcome. [a]Participants who had a first or second-degree relative in the sample (kinship coefficient ϕ>0.125) were removed by using PLINK 1.9. [b]Please refer to online supplemental methods for detailed procedures of case-control matching. [c]See online supplemental methods and table 3 for details. [d]See online supplemental methods and table 4 for details. AS, any stroke; C+T, clumping and thresholding; CAD, coronary heart disease; CKB, China Kadoorie Biobank; GWAS, genome-wide association study; ICH, intracerebral haemorrhage; IS, ischaemic stroke; PRS, polygenic risk score; SAH, subarachnoid haemorrhage; SSF, summary statistics file; TIA, transient ischaemic attack.

sizes.[14 17] (3) Identification of the optimal PRS for each outcome. The performances of different PRSs in predicting each outcome were compared in the corresponding training sets. (4) Validation and evaluation of the optimal PRS for each outcome. We prospectively examined the associations between optimal PRSs and risks of stroke and its subtypes. We evaluated the impact of PRSs on the risk prediction improvement by adding the optimal PRS to traditional risk prediction models in the testing set.

Assessment of traditional stroke risk factors

The baseline questionnaire collected information on sociodemographic characteristics, lifestyle behaviours, dietary habits, and personal and family medical history.[15] Traditional stroke risk factors considered in the present study included sex, age, systolic and diastolic blood pressure (SBP and DBP), smoking, body mass index (BMI), waist circumference, hypertension, diabetes and family history of stroke. Details on the collection and definition of these variables have been described in our previous work.[18 19]

Genetic data

At baseline, a 10 mL random blood sample was collected from each participant. Genotyping and imputation in this study were centrally conducted, with details provided in our previous study.[19 20] Briefly, two custom-designed single nucleotide polymorphism (SNP) arrays (Affymetrix Axiom CKB array) were used for genotyping. Imputation was performed based on haplotypes derived from the 1000 Genomes Project Phase 3. There were 9.54 million genetic variants with high reliability (online supplemental figure 1).

Polygenic risk scores

We searched the PGS Catalogue,[21] PubMed and Embase. Four previous stroke PRSs were selected for validation analyses (online supplemental methods and table 3).[2 4 5 10] Meanwhile, we ran gwasfilter to filter genome-wide association studies (GWAS) from the GWAS Catalogue (https://www.ebi.ac.uk/gwas/).[22 23] Based on ethnicity, sample size and accessibility of the summary statistics file (SSF), we finally included one AS SSF, two SAH SSFs, two ICH SSFs and two IS SSFs from two large-scale GWAS (online supplemental methods and table 4).[14 17] Similar to our latest research,[19] we developed new PRSs by using two methods: clumping and thresholding ('C+T') and LDpred[16] (online supplemental methods).

Ascertainment of stroke outcomes

All participants were followed up for morbidity and mortality since their baseline enrolment. Incident events were identified by linking with local disease and death registries and the national health insurance database and supplemented by active follow-up.[15] In the testing set, only 653 (0.91%) were lost to follow-up before censoring on 31 December 2018. Trained staff blinded to baseline information coded all events using the International Classification of Diseases, 10th Revision (ICD-10). Incident stroke events during the follow-up were defined as I60–I64, including SAH (I60), ICH (I61), other nontraumatic intracranial haemorrhage (I62), IS (I63) and unspecified stroke (I64). In the testing set, the events coded as I62 and I64 accounted for only 0.9% (n=76) and 3.5% (n=302) of all incident stroke events.

Since 2014, medical records of incident stroke cases have been retrieved and reviewed by qualified cardiovascular specialists blinded to baseline information. According to a previous study,[24] by October 2018, the reporting accuracy was 91.7%, 90.4% and 82.7% for IS, ICH and SAH[24]; the corresponding diagnostic accuracy was 93.1% (including silent lacunar infarction), 98.2% and 98.1%, respectively.[24]

Identification of the optimal PRS in the training set

In each training set, we used the conditional logistic regression model to measure the association of each PRS with the risk of the corresponding stroke outcome, stratified by the case–control pair, with the top 10 principal components of ancestry (PCA) and array versions as the covariates. We defined the optimal PRS as the PRS with the highest OR per SD, as our previous study did.[19]

Validation and evaluation of the optimal PRS in the testing set

In the testing set, we used the Cox regression model to measure the association of optimal PRSs with risks of stroke and stroke subtypes. The model was stratified by sex and ten study regions, with age as the time scale and adjusting for the top 10 PCA and array versions. We further adjusted for SBP, BMI and family history of stroke in sensitivity analyses. We evaluated the proportional hazards assumptions by examining Schoenfeld residuals. Either non-existent or minimal deviations were observed. In subgroup analyses, the tests for multiplicative interaction were performed using likelihood ratio tests by comparing models with and without cross-product terms between the stratifying variable and PRS.

To evaluate the impact of PRS on risk prediction improvement, we defined the 'CKB-CVD models' as the traditional risk prediction models, as our previous study did.[19] The 'CKB-CVD models' distinguish risks of IS and haemorrhagic stroke and have good discrimination without relying on blood lipids.[18] We added the PRS to traditional models to get a 'PRS-enhanced model'. We assessed the discrimination performance by using Harrell's C.[25] We used the net reclassification improvement (NRI) and integrated discrimination improvement to evaluate model reclassification before and after the addition of PRS.[26]

The study adhered to the PRS Reporting Standards and statement Strengthening the reporting of observational studies in epidemiology for cohort studies simultaneously (online supplemental file 2).[27 28] Analyses were done with

Table 1 The optimal PRSs associated with risks of stroke and its subtypes in the training sets

Outcomes	Method	PRS source*	No of variants	OR_{SD} (95% CI)	P value	Note
Any stroke (N=7412 pairs)						
	Previous study	PGS002259	448	1.13 (1.09 to 1.16)	1.44×10^{-11}	
	C+T	GCST005838 (p=1×10^{-6}, r^2=0)	38	1.11 (1.07 to 1.14)	1.90×10^{-9}	
	LDpred	GCST005838 (ρ=0.01, Ref=1KGP-EAS)	1 017 531	1.14 (1.10 to 1.18)	3.38×10^{-14}	Optimal
Ischaemic stroke (N=3844 pairs)						
	Previous study	PGS000039	1 563 569	1.07 (1.01 to 1.12)	0.012	
	C+T	GCST90018864 (p=0.02, r^2=0.8)	32 158	1.18 (1.13 to 1.24)	3.55×10^{-11}	Optimal
	LDpred	GCST90018864 (ρ=0.01, Ref=1KGP-EUR)	1 017 672	1.17 (1.11 to 1.23)	1.46×10^{-9}	
Intracerebral haemorrhage (N=4296 pairs)						
	C+T	GCST90018870 (p=0.001, r^2=0.2)	1326	1.09 (1.04 to 1.14)	1.37×10^{-4}	
	LDpred	GCST90018870 (ρ=0.1, Ref=1KGP-EUR)	1 017 664	1.10 (1.05 to 1.15)	3.09×10^{-5}	Optimal
Subarachnoid haemorrhage (N=359 pairs)						
	C+T	GCST90018703 (p=0.4, r^2=0)	7899	1.25 (1.06 to 1.47)	9.21×10^{-3}	Optimal
	LDpred	GCST90018923 (ρ=0.01, Ref=1KGP-EUR)	1 017 665	1.15 (0.98 to 1.35)	0.096	

The current table only displays the optimal PRS obtained from different strategies (previous study, C+T and LDpred) for each disease outcome. The detailed results of all PRSs can be found in online supplemental table 7.
*'PGS' indicates the index in the PGS Catalogue. 'GCST' indicates the index in the GWAS Catalogue. The information in brackets is the parameter used for developing the PRS.
C+T, clumping and thresholding; EAS, East Asian; EUR, European; 1KGP, 1000 Genomes Project (Phase 3); PRS, polygenic risk score; Ref, reference population.

Stata (V.17.0, StataCorp) and R (V.4.0.3). All statistical tests were two sided with α=0.05.

RESULTS

Selection of the optimal PRSs in the training sets

In this study, four 1:1 matched training sets were defined to identify the optimal PRS for AS (7412 pairs), IS (3844 pairs), ICH (4296 pairs) and SAH (359 pairs) (figure 1, online supplemental methods). Among the training sets, 72.7%, 61.6%, 77.9% and 63.8% of the participants were from rural areas in China; 51.9%, 50.5%, 53.4% and 38.4% of the participants were men, respectively. Among the cases, the median age of disease onset (25th–75th percentile) was 65.3 (57.0–72.0), 64.1 (56.1–70.6), 65.9 (57.7–73.0) and 61.0 (53.8–69.2) years, respectively. Among all training sets, the proportion of the control group using the first version of the SNP array was lower than that of the case group (p<0.001) (online supplemental table 2). The performance of PRS for AS and IS developed in previous studies was not better than that of the newly developed PRS in the present study (table 1, online supplemental table 5). The optimal PRS for AS came from the LDpred method, and the optimal PRS for IS, ICH and SAH came from the C+T method. The OR_{SD} (95% CI) of the optimal PRSs was 1.14 (1.10 to 1.18) for AS, 1.18 (1.13 to 1.24) for IS, 1.10 (1.05 to 1.15) for ICH and 1.25 (1.06 to 1.47) for SAH (table 1, online supplemental table 5).

Associations of PRSs with stroke and its subtypes in the testing set

The testing set included 72 150 Chinese participants, of which 59.8% were women. The median age was 50.6 years in women and 51.9 years in men. During 872 919 person-years of follow-up (over 12 years on average), 8514 incident stroke events were documented, including 7507 IS, 1193 ICH and 132 SAH (table 2). The correlations among the optimal PRSs were weak (all correlation coefficients<0.2) (online supplemental figure 2).

The PRS_{AS} and PRS_{IS} were both positively associated with risks of AS, IS and ICH (p<0.05). The HR_{SD} (95% CIs) of PRS_{AS} was 1.10 (1.07 to 1.12), 1.10 (1.07 to 1.12) and 1.13 (1.07 to 1.20) for AS, IS and ICH, respectively. The corresponding HR_{SD} (95% CIs) of PRS_{IS} was 1.08 (1.06 to 1.11), 1.08 (1.06 to 1.11) and 1.09 (1.03 to 1.15) (figure 2, online supplemental table 6). PRS_{ICH} was positively associated with the risk of ICH in the whole testing set (HR_{SD}=1.07), though it was not statistically significant in women (p for sex interaction=0.056) (figure 2C). PRS_{SAH} was not associated with risks of any outcomes (figure 2). A strong association of PRS_{AS} with the risk of SAH (HR_{SD}=1.38, 95% CI 1.03 to 1.87) was observed in men but not in women (p for sex interaction=0.055) (figure 2D).

In sensitivity analyses, the associations of PRSs with risks of stroke and its subtypes did not change significantly after additional adjustment for SBP, BMI and family history of stroke (online supplemental table 6). In subgroup analyses, there was no strong evidence supporting a different association strength across subgroups for IS and ICH after

Table 2 Characteristics of the testing set

	Women	Men
No of participants	43170	28980
Baseline characteristics		
Age, years	50.6 (42.5–58.3)	51.9 (43.2–60.3)
Rural areas	22449 (52.0)	15772 (54.4)
Array 1	5948 (13.8)	4503 (15.5)
Primary school and below	23605 (54.7)	11882 (41.0)
Daily smokers	915 (2.1)	16317 (56.3)
Body mass index, kg/m^2	23.6 (21.4–26.0)	23.3 (21.1–25.7)
Waist circumference, cm	78.0 (72.0–84.5)	81.5 (74.5–88.5)
Hypertension	14062 (32.6)	10653 (36.8)
Diabetes	2477 (5.7)	1553 (5.4)
Family history of stroke	7619 (17.6)	5075 (17.5)
Follow-up		
Follow-up time, years	12.6 (11.7–13.4)	12.4 (11.4–13.3)
Total person-years*	529498	343421
Incident events†		
Any stroke	4763 (11.0)	3751 (12.9)
Ischaemic stroke	4254 (9.9)	3253 (11.2)
Intracerebral haemorrhage	600 (1.4)	593 (2.0)
Subarachnoid haemorrhage	87 (0.2)	45 (0.2)

Data are presented as n (%) or median (25th–75th percentile) unless otherwise specified.
*Person-years were calculated as the time from the baseline date to the first of the following: death, lost to follow-up or the global censoring date (31 December 2018).
†Only the first event was counted.

considering multiple testing (p for interaction>0.05/8) (online supplemental figures 3 and 4).

Addition of the optimal PRS to traditional risk prediction models

Based on the traditional models defined in this study, the addition of the PRS did not improve or only slightly improve the discrimination performance of the models. For IS, the addition of PRS_{AS} increased Harrell's C by 0.0010 in men (p=0.002). For haemorrhagic stroke, the addition of PRSs did not influence Harrell's C significantly (p>0.05) (figure 3). The addition of the PRS offered little to no improvement in stroke risk stratification. For example, the categorical NRIs at the 10% high-risk threshold for ischaemic and haemorrhagic stroke were all not significant in both sexes (p>0.05) (online supplemental table 7).

DISCUSSION

Based on the largest biobank in the Chinese population, only moderate associations were observed between PRSs and risks of stroke and its subtypes in this Chinese population, with an HR_{SD} of about 1.10. The addition of current PRSs offered little to no improvement in stroke risk prediction and risk stratification. We also found that the PRSs developed from GWAS summary statistics of IS were positively associated with the risk of ICH.

In the present study, the associations of PRSs with risks of stroke and its subtypes were moderate, suggesting a limited value for improving risk prediction over traditional risk factors. The HR_{SD} for PRS was usually greater than 1.20 in previous studies of the general population. A PRS for IS (PGS000039) that was developed with the metaGRS method and combined PRSs of 5 stroke subtypes and 14 stroke-related traits had an HR_{SD} of 1.26 (95% CI 1.22 to 1.31) in the European population.[5] Another PRS for stroke (PGS002259) was also developed using the metaGRS method in a Chinese population, with the HR_{SD} for stroke being 1.28 (95% CI 1.21 to 1.36).[10] However, these two PRSs showed much weaker associations with the risk of stroke or IS in the present study than in previous studies. Since both PRSs were developed using the elastic-net logistic regression, a machine learning approach, the potential overfitting may undermine their generalisation performance.

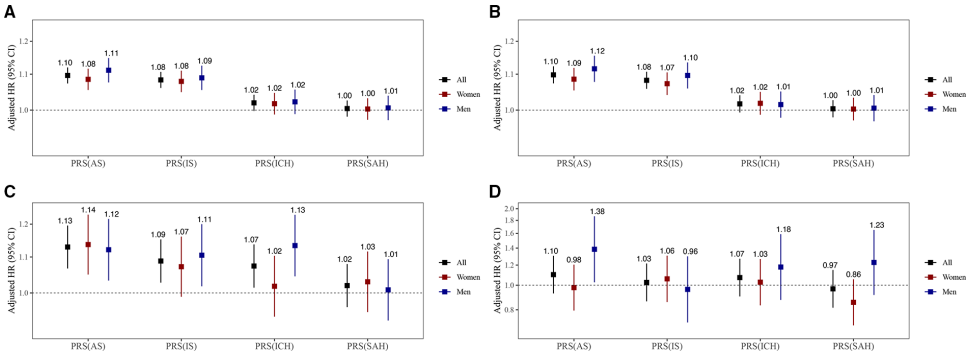


Figure 2 Associations of PRSs with risks of stroke and its subtypes. (A) AS, (B) IS, (C) ICH, (D) SAH. The PRSs reported here are the optimal PRSs for stroke and its subtypes in the training sets (see table 1), which were standardised (0 mean, unit SD) in the testing set. Cox models were stratified by sex and 10 study regions and adjusted for the top 10 principal components of ancestry and array versions, with age as the time scale. The number above the closed square represents the HR. The number of stroke events in women and men has been reported in table 2. The vertical lines indicate 95% CIs. AS, any stroke; ICH, intracerebral hemorrhage; IS, ischaemic stroke; PRS, polygenic risk score; SAH, subarachnoid haemorrhage.

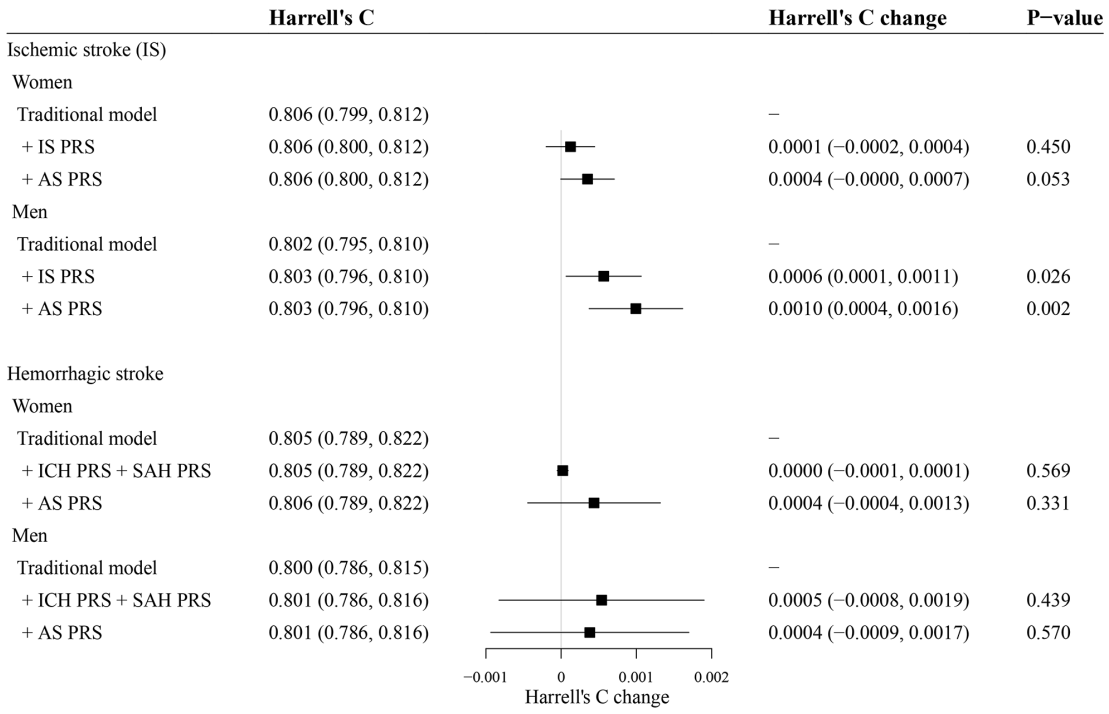


Figure 3 C statistics evaluating the performance of PRS. The traditional risk prediction models (traditional models) were defined as sex-specific Cox models stratified by 10 study regions, with time on study as the time scale, including models for ischaemic stroke (ICD-10: I63) and models for haemorrhagic stroke (ICD-10: I60–I62).[18] Predictors included in traditional models were the same as the 'CKB-CVD models', including age, systolic and diastolic blood pressure, use of antihypertensives, current daily smoking, self-reported diabetes and waist circumference. Interactions between age and the other six predictors were also included. The 95% CIs of Harrell's C and Harrell's C changes were calculated by 100 bootstrap replications using the BCa method in Stata. CKB, China Kadoorie Biobank; CVD, Cardiovascular disease; ICD, International Classification of Disease; PRS, Polygenic risk score.

The incidence rate of ICH is much higher in Chinese than in European populations. However, non-European populations are under-represented in GWAS, which serves as the basis for PRS development. The largest GWAS for ICH included only 3400 ICH cases, with most of them from European populations.[17] The present study attempted to develop PRS for ICH based on summary statistics from this GWAS. The weak associations observed in the present study are either explained by the difference in genetic background between ethnic groups or suggest that this GWAS may be underpowered. The stronger association estimate between PRS and HS risk reported in the previous study was likely due to the inclusion of PRSs for risk factors of HS (such as blood pressure) in the metaGRS method.[10] It is worth mentioning that, in the present study, the PRSs directly developed from GWAS summary statistics of IS were also positively associated with the risk of ICH. Although there are differences in aetiology and risk factor profile between IS and ICH,[11–13] they might also have some partially shared aetiological mechanisms like the cerebral small-vessel disease.[29]

This study has the following strengths. The large sample size and a large number of stroke events (including IS and ICH) enabled us to separate powerful training sets and the testing set and to conduct subgroup analyses. The lost to follow-up rate was less than 1% at an average follow-up period of over 12 years in CKB. The main subtypes of stroke (ie, IS, ICH and SAH) were well classified, and the reporting and diagnostic accuracy of stroke events were high.[24] The genotyping and imputation of genetic data in this study were centrally conducted through a standard quality control process. Genetic variants with high reliability covered the whole genome well.

However, several limitations merit consideration. First, we did not further consider the subtypes of IS (eg, large-atherosclerotic stroke, cardioembolic stroke and small vessel stroke) as over 75% of the incident IS events were coded as unspecified IS (ICD-10: I63.9), which precluded us from conducting more detailed analyses. Previous studies have suggested that there are differences in genetic loci of different IS subtypes.[14 30] Subsequent studies can explore whether distinguishing IS subtypes can further improve the predictive ability of PRS for IS. Second, compared with IS and ICH, the number of SAH events was relatively small. Therefore, it is difficult to exclude chance factors for the positive results observed in the present study. Further studies with more SAH events are warranted to examine our findings. Third, the genetic variants with ambiguous SNP (ie, A/T, C/G) and those that were not found in CKB or had low imputation quality scores were removed during the standard quality control process of PRSs. This might weaken the associations of previous PRSs with stroke and its subtypes. Fourth, because information on blood lipids was not available for

the current study population, we were unable to compare the impacts of blood lipids and PRS on traditional stroke risk prediction model improvement. However, the addition of blood lipids may enhance the traditional non-laboratory-based models, as previous studies have shown.[31 32] Therefore, adding PRS to a 'lipid-enhanced model' might lead to a more minor improvement than what we have observed in the present study.

CONCLUSIONS

In this Chinese population, the associations of optimal PRSs with risks of stroke and its subtypes were moderate, suggesting a limited value for improving risk prediction over traditional risk factors in the context of current GWAS under-representing the East Asian population. As GWAS of stroke and its subtypes progress among East Asians, further studies are warranted to assess whether new PRSs have considerable potential to translate into precision public health and population health benefits and, if so, to determine the appropriate context for their use.

Author affiliations
[1]Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing, China
[2]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, Hunan, China
[3]Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing, China
[4]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
[5]Fuwai Hospital Chinese Academy of Medical Sciences, Beijing, China
[6]Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK
[7]Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK
[8]NCDs Prevention and Control Department, Suzhou CDC, Suzhou, Jiangsu, China
[9]China National Center for Food Safety Risk Assessment, Beijing, China
[10]State Key Laboratory of Vascular Homeostasis and Remodeling, Peking University, Beijing, China

Acknowledgements The most important acknowledgment is to the participants in the study and the members of the survey teams in each of the 10 regional centres, as well as to the project development and management teams based in Beijing, Oxford and the 10 regional centres.

Collaborators International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Yu Guo, Liming Li (PI), Jun Lv, Richard Peto, Robin Walters. International Co-ordinating Centre, Oxford: Daniel Avery, Derrick Bennett, Ruth Boxall, Sue Burgess, Ka Hung Chan, Yumei Chang, Yiping Chen, Zhengming Chen, Johnathan Clarke; Robert Clarke, Huaidong Du, Ahmed Edris Mohamed, Zammy Fairhurst-Hunter, Hannah Fry, Simon Gilbert, Alex Hacker, Mike Hill, Michael Holmes, Pek Kei Im, Andri Iona, Maria Kakkoura, Christiana Kartsonaki, Rene Kerosi, Kuang Lin, Mohsen Mazidi, Iona Millwood, Sam Morris, Qunhua Nie, Alfred Pozarickij, Paul Ryder, Saredo Said, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Becky Stevens, Iain Turnbull, Robin Walters, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang, Pang Yao. National Co-ordinating Centre, Beijing: Yu Guo, Xiao Han, Can Hou, Jun Lv, Pei Pei, Chao Liu, Canqing Yu, Qingmei Xia. 10 Regional Co-ordinating Centres: Qingdao CDC: Zengchang Pang, Ruqin Gao, Shanpeng Li, Haiping Duan, Shaojie Wang, Yongmei Liu, Ranran Du, Yajing Zang, Liang Cheng, Xiaocao Tian, Hua Zhang, Yaoming Zhai, Feng Ning, Xiaohui Sun, Feifei Li. Licang CDC: Silu Lv, Junzheng Wang, Wei Hou. Heilongjiang Provincial CDC: Wei Sun, Shichun Yan, Xiaoming Cui. Nangang CDC: Chi Wang, Zhenyuan Wu,Yanjie Li, Quan Kang. Hainan Provincial CDC: Huiming Luo, Tingting Ou. Meilan CDC: Xiangyang Zheng, Zhendong Guo, Shukuan Wu, Yilei Li, Huimei Li. Jiangsu Provincial CDC: Ming Wu, Yonglin Zhou, Jinyi Zhou, Ran Tao, Jie Yang, Jian Su. Suzhou CDC: Fang Liu, Jun Zhang, Yihe Hu, Yan Lu, Liangcai Ma, Aiyu Tang, Shuo Zhang, Jianrong Jin, Jingchao Liu. Guangxi Provincial CDC: Mei Lin, Zhenzhen Lu. Liuzhou CDC: Lifang Zhou, Changping Xie, Jian Lan,Tingping Zhu,Yun Liu, Liuping Wei, Liyuan Zhou, Ningyu Chen, Yulu Qin, Sisi Wang. Sichuan Provincial CDC: Xianping Wu, Ningmei Zhang, Xiaofang Chen, Xiaoyu Chang. Pengzhou CDC: Mingqiang Yuan, Xia Wu, Xiaofang Chen, Wei Jiang, Jiaqiu Liu, Qiang Sun. Gansu Provincial CDC: Faqing Chen, Xiaolan Ren, Caixia Dong. Maiji CDC: Hui Zhang, Enke Mao, Xiaoping Wang, Tao Wang, Xi zhang. Henan Provincial CDC: Kai Kang, Shixian Feng, Huizi Tian, Lei Fan. Huixian CDC: XiaoLin Li, Huarong Sun, Pan He, Xukui Zhang. Zhejiang Provincial CDC: Min Yu, Ruying Hu, Hao Wang. Tongxiang CDC: Xiaoyi Zhang, Yuan Cao, Kaixu Xie, Lingli Chen, Dun Shen. Hunan Provincial CDC: Xiaojun Li, Donghui Jin, Li Yin, Huilin Liu, Zhongxi Fu. Liuyang CDC: Xin Xu, Hao Zhang, Jianwei Chen,Yuan Peng, Libo Zhang, Chan Qu.

Contributors JL conceived and designed the study. LL, ZC and JC: members of the China Kadoorie Biobank Steering Committee, designed and supervised the whole study, obtained funding, and, together with CY, YG, DiS, YP, PP, LY, YC, HD, YL, SB, DA, IYM and RGW: acquired the data. SY, ZS and DoS analysed the data. SY drafted the manuscript. CY, YP, DiS and RC helped to interpret the results. JL contributed to the critical revision of the manuscript for important intellectual content. All authors reviewed and approved the final manuscript. JL is the guarantor.

Funding This work was supported by the National Natural Science Foundation of China (82192904, 82192901, 82192900). The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants from the UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), grants (2016YFC0900500) from the National Key R&D Program of China, National Natural Science Foundation of China (81390540, 91846303, 81941018) and Chinese Ministry of Science and Technology (2011BAI09B01).

Disclaimer The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and CKB had ethical approvals from the Ethical Review Committee of the Chinese Center for Disease Control and Prevention (Beijing, China) (approval notice: 005/2004) and the Oxford Tropical Research Ethics Committee, University of Oxford (UK) (reference: 025-04). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Details of how to access China Kadoorie Biobank data and details of the data release schedule are available from www.ckbiobank.org/site/Data+Access.

ORCID iDs
Songchun Yang http://orcid.org/0000-0002-0021-4178
Pei Pei http://orcid.org/0000-0002-5741-6563
Daniel Avery http://orcid.org/0000-0002-9823-9575
Robert Clarke http://orcid.org/0000-0002-9802-8241
Jun Lv http://orcid.org/0000-0001-7916-3870

REFERENCES

1 Feigin VL, Stark BA, Johnson CO, *et al*. Stroke collaborators. global, regional, and national burden of stroke and its risk factors, 1990-

2019: a systematic analysis for the global burden of disease study 2019. *Lancet Neurol* 2021;20:795–820.
2 Ibrahim-Verbaas CA, Fornage M, Bis JC, *et al*. Predicting stroke through genetic risk functions: the CHARGE risk score project. *Stroke* 2014;45:403–12.
3 Malik R, Bevan S, Nalls MA, *et al*. Multilocus genetic risk score associates with ischemic stroke in case-control and prospective cohort studies. *Stroke* 2014;45:394–402.
4 Rutten-Jacobs LC, Larsson SC, Malik R, *et al*. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK biobank participants. *BMJ* 2018;363:k4168.
5 Abraham G, Malik R, Yonova-Doing E, *et al*. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun* 2019;10:5819.
6 Li J, Chaudhary DP, Khan A, *et al*. Polygenic risk scores augment stroke subtyping. *Neurol Genet* 2021;7:e560.
7 Marston NA, Patel PN, Kamanu FK, *et al*. Clinical application of a novel genetic risk score for ischemic stroke in patients with cardiometabolic disease. *Circulation* 2021;143:470–8.
8 O'Sullivan JW, Shcherbina A, Justesen JM, *et al*. Combining clinical and polygenic risk improves stroke prediction among individuals with atrial fibrillation. *Circ Genom Precis Med* 2021;14:e003168.
9 Sun L, Pennells L, Kaptoge S, *et al*. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. *PLoS Med* 2021;18:e1003498.
10 Lu X, Niu X, Shen C, *et al*. Development and validation of a polygenic risk score for stroke in the Chinese population. *Neurology* 2021;97:e619–28.
11 Chen Z, Iona A, Parish S, *et al*. Adiposity and risk of ischaemic and haemorrhagic stroke in 0·5 million Chinese men and women: a prospective cohort study. *Lancet Glob Health* 2018;6:e630–40.
12 Sun L, Clarke R, Bennett D, *et al*. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nat Med* 2019;25:569–74.
13 Gu X, Li Y, Chen S, *et al*. Association of lipids with ischemic and hemorrhagic stroke: a prospective cohort study among 267 500 Chinese. *Stroke* 2019;50:3376–84.
14 Malik R, Chauhan G, Traylor M, *et al*. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018;50:524–37.
15 Chen Z, Chen J, Collins R, *et al*. China Kadoorie biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;40:1652–66.
16 Vilhjálmsson BJ, Yang J, Finucane HK, *et al*. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 2015;97:576–92.
17 Sakaue S, Kanai M, Tanigawa Y, *et al*. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* 2021;53:1415–24.
18 Yang S, Han Y, Yu C, *et al*. Development of a model to predict 10-year risk of ischemic and hemorrhagic stroke and ischemic heart disease using the China Kadoorie biobank. *Neurology* 2022;98:e2307–17.
19 Yang S, Sun D, Sun Z, *et al*. Minimal improvement in coronary artery disease risk prediction in Chinese population using polygenic risk scores: evidence from the China Kadoorie biobank. *Chin Med J (Engl)* 2023.
20 Zhu Z, Li J, Si J, *et al*. A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic etiology with obesity. *Eur Respir J* 2021;58:2100199.
21 Lambert SA, Gil L, Jupp S, *et al*. The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* 2021;53:420–5.
22 Yang S, Li C, Hu Y, *et al*. Gwasfilter: an R script to filter genome-wide Association study. *Chin J Epidemiol* 2021;42:1876–81.
23 Buniello A, MacArthur JAL, Cerezo M, *et al*. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
24 Turnbull I, Clarke R, Wright N, *et al*. Diagnostic accuracy of major stroke types in Chinese adults: a clinical adjudication study involving 40,000 stroke cases. *Lancet Reg Health West Pac* 2022;21:100415.
25 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
26 Pencina MJ, D'Agostino RB, D'Agostino RB, *et al*. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
27 Wand H, Lambert SA, Tamburro C, *et al*. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 2021;591:211–9.
28 Elm E von, Altman DG, Egger M, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
29 Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol* 2019;18:684–96.
30 Linden AB, Clarke R, Hammami I, *et al*. Genetic associations of adult height with risk of cardioembolic and other subtypes of ischemic stroke: a mendelian randomization study in multiple ancestries. *PLoS Med* 2022;19:e1003967.
31 Ueda P, Woodward M, Lu Y, *et al*. Laboratory-based and office-based risk scores and charts to predict 10-year risk of cardiovascular disease in 182 countries: a pooled analysis of prospective cohorts and health surveys. *Lancet Diabetes Endocrinol* 2017;5:196–213.
32 Kaptoge S, Pennells L, De Bacquer D, *et al*. The WHO CVD risk chart working group. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health* 2019;7:e1332–45.

Twin Research and Human Genetics (2023), 1–8
doi:10.1017/thg.2023.25

CAMBRIDGE UNIVERSITY PRESS

Article

Genetic and Environmental Influences on Blood Pressure and Serum Lipids Across Age-Groups

Ke Miao[1,2], Yutong Wang[1,2], Weihua Cao[1,2], Jun Lv[1,2], Canqing Yu[1,2], Tao Huang[1,2], Dianjianyi Sun[1,2], Chunxiao Liao[1,2], Yuanjie Pang[1,2], Runhua Hu[1,2], Zengchang Pang[3], Min Yu[4], Hua Wang[5], Xianping Wu[6], Yu Liu[7], Wenjing Gao[1,2] and Liming Li[1,2]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China, [2]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China, [3]Qingdao Center for Disease Control and Prevention, Qingdao, China, [4]Zhejiang Center for Disease Control and Prevention, Hangzhou, China, [5]Jiangsu Center for Disease Control and Prevention, Nanjing, China, [6]Sichuan Center for Disease Control and Prevention, Chengdu, China and [7]Heilongjiang Center for Disease Control and Prevention, Harbin, China

Abstract

Aging plays a crucial role in the mechanisms of the impacts of genetic and environmental factors on blood pressure and serum lipids. However, to our knowledge, how the influence of genetic and environmental factors on the correlation between blood pressure and serum lipids changes with age remains to be determined. In this study, data from the Chinese National Twin Registry (CNTR) were used. Resting blood pressure, including systolic and diastolic blood pressure (SBP and DBP), and fasting serum lipids, including total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TGs) were measured in 2378 participants (1189 twin pairs). Univariate and bivariate structural equation models examined the genetic and environmental influences on blood pressure and serum lipids among three age groups. All phenotypes showed moderate to high heritability (0.37–0.59) and moderate unique environmental variance (0.30–0.44). The heritability of all phenotypes showed a decreasing trend with age. Among all phenotypes, SBP and DBP showed a significant monotonic decreasing trend. For phenotype-phenotype pairs, the phenotypic correlation (Rph) of each pair ranged from −0.04 to 0.23, and the additive genetic correlation (Ra) ranged from 0.00 to 0.36. For TC&SBP, TC&DBP, TG&SBP and TGs&DBP, both the Rph and Ra declined with age, and the Ra difference between the young group and the older adult group is statistically significant ($p < .05$). The unique environmental correlation (Re) of each pair did not follow any pattern with age and remained relatively stable with age. In summary, we observed that the heritability of blood pressure was affected by age. Moreover, blood pressure and serum lipids shared common genetic backgrounds, and age had an impact on the phenotypic correlation and genetic correlations.

Keywords: Heritability; blood pressure; serum lipids; genetic; twin study

(Received 19 March 2023; revise received 23 May 2023; accepted 23 May 2023)

Hypertension, especially high systolic blood pressure (SBP), is widely recognized as a risk factor for cardiovascular disease, with 10.8 million deaths worldwide in 2019 (GBD 2019 Risk Factors Collaborators, 2020). Dyslipidemia is defined as abnormal serum lipid levels, including low levels of high-density lipoprotein cholesterol (HDL-C) and elevated levels of serum total cholesterol (TC), triglycerides (TGs) and low-density lipoprotein cholesterol (LDL-C).

Both hypertension and dyslipidemia are risk factors for damage to the vessel wall damage. Dyslipidemia directly affects blood pressure (Ruixing et al., 2009). First, dyslipidemia reduces the elasticity of the blood vessel wall and causes an increase in SBP. It causes excessive lipid molecules to stimulate the elastic components of the arterial wall, resulting in the production of large amounts of extracellular matrix and inflammatory mediators, such as interleukins and chemokines. These inflammatory mediators and lipids eventually form atheromatous plaques locally with platelets in the arterial wall. Finally, in terms of hemodynamics, the increase in lipid molecules in the blood will increase the viscosity of the blood, slow down the flow of blood in the blood vessels, and make it difficult for peripheral blood to flow back into the heart, resulting in an increase in both SBP and diastolic blood pressure (DBP). However, no definitive evidence has indicated that hypertension can directly cause dyslipidemia (Hurtubise et al., 2016). The elasticity of the blood vessel wall in middle-aged and older adults is weakened, SBP is prone to increase. And the low metabolic rate of middle-aged and elderly people may lead to abnormality of hepatic lipid metabolism, resulting in dyslipidemia. In addition, some hypertension drugs, such as thiazide diuretics and β-blockers, have adverse effects on lipid metabolism, which may further promote the occurrence of dyslipidemia (Deshmukh et al., 2008).

Both blood pressure and serum lipids are affected by genetic and environmental factors. In general, blood pressure increases

Corresponding author: Wenjing Gao; Email: pkuepigwj@126.com; Liming Li; Email: lmlee@vip.163.com

Cite this article: Miao K, Wang Y, Cao W, Lv J, Yu C, Huang T, Sun D, Liao C, Pang Y, Hu R, Pang Z, Yu M, Wang H, Wu X, Liu Y, Gao W, Li L. Genetic and Environmental Influences on Blood Pressure and Serum Lipids Across Age-Groups. *Twin Research and Human Genetics* https://doi.org/10.1017/thg.2023.25

with age. At the same time, the pattern of serum lipids changes dramatically with age, especially in the following four life periods: infancy, adolescence, female menopause, and old age (Snieder et al., 1999). These specific age trends indicate that the effect of genetic and environmental factors changes with age (Snieder et al., 1999). In addition, evidence from epidemiological studies has found a positive correlation between blood pressure and serum lipids (Lepira et al., 2005; Tang et al., 2022; Zhang et al., 2007), and the correlation also decreases with age (Bønaa & Thelle, 1991). In previous twin studies, it was found that there was a moderate or low genetic correlation between blood pressure and serum lipids (Benyamin et al., 2007; Duan et al., 2011; Panizzon et al., 2015; Zhang et al., 2009). However, most previous studies did not consider the effect of age on genetic correlation, and it is unclear how the influence of genetic and environmental factors on the correlation between blood pressure and serum lipids changes with age.

Therefore, we aimed to quantify genetic and environmental contributions to blood pressure and serum lipids. The degree to which the genetic and environmental contributions are shared for blood pressure and serum lipids was explored using twins in the Chinese National Twin Registry (CNTR). Furthermore, we studied the changes in different age groups.

Method

Study Population

The participants were from the CNTR, the largest twin registration system in China, which has been detailed elsewhere (Gao et al., 2019). The study data from thematic surveys of the CNTR between 2013 and 2017 included a questionnaire survey (demographic information and medication history), physical examination (blood pressure) and blood sampling (biochemical tests). This study received ethical approval from the Biomedical Ethics Committee at Peking University, Beijing, China. All participants provided written informed consent.

Twins who are 18 years old and over were included. Exclusion standards were as follows: (1) twins with missing key variables, such as age and zygosity; (2) twin pairs who were raised separately for more than one year before the age of 5; (3) opposite-sex twin pairs; (4) twins who had been taking hypolipidemic agents in the previous month; and (5) pregnant women and their cotwins. Ultimately, this study divided 2378 participants (1189 twin pairs) into three age groups based on age: young ($\leq$ 5 years old), middle-aged (46 to 55 years old), and older adults (>55 years old); see Figure 1.

Measurements of Blood Pressure

Blood pressure of the right arm was measured twice in a sitting position using an electronic sphygmomanometer after at least 5 minutes of rest. A third blood pressure measurement was taken if the difference between the first two blood pressures was greater than 10 mmHg. The average of the two measurements with the closest values was taken as the mean blood pressure. Blood pressure needed to be adjusted if antihypertensive drugs were taken, and 10 mmHg and 15 mmHg were added to the measured DBP and SBP respectively (Tobin et al., 2005).

Measurements of Serum Lipids

The peripheral blood was collected after at least 12 hours of fasting. The blood samples were stored in a portable blood refrigerator box at 4 °C. Then, a desktop frozen centrifugal machine was used to separate the serum and blood cells for 20 minutes at 2500 rpm. Separated serum was used to detect serum lipids according to the same testing process and standards.

Twin Zygosity

The CNTR used the 'Peas in the Pod Questionnaire' (PPQ) to judge the zygosity of twins, which asks about the degree of similarity of twins. Compared with genetic data, the accuracy of this method can reach 87% (Wang et al., 2015).

Data Analysis

All data analyses were conducted using R (4.0.2) software, and structural equation models (SEM) were performed using the OpenMx (2.18.1) package of R. For serum TC, TG, HDL-C and LDL-C showing a nonnormal distribution, a natural logarithm transformation was carried out. For each trait, individuals with blood pressure or serum lipid value that exceeded ±3 standard deviation (*SD*) from the mean and their co-twins were removed. The models were adjusted for age and sex; $p < .05$ indicated statistical significance.

Structural Equation Modeling

The SEM was utilized to evaluate the genetic and environmental factors influencing a trait. This approach partitions the variation into four components: additive genetic component (A), unique environmental component (E), dominant genetic component (D) and common environmental component (C). Since only twins reared together are included, the effects of C or D cannot be estimated simultaneously in the same model due to the confounding of these components (Rijsdijk & Sham, 2002). As it is dependent on the pattern of correlations between MZ and DZ twins, an ACE or ADE model was chosen. C is calculated when the MZ correlation is less than twice the DZ correlation, while D is calculated if the MZ correlation is more than the DZ correlation. Then A, C or D were gradually dropped to fit the submodels. Using likelihood ratio tests, we compared ACE or ADE models and the submodels with the saturated models. The criteria for the suitable model was the $p > .05$ and the lowest Akaike information criterion (AIC) value.

Univariate Structural Equation Modeling

We constructed the univariate SEM for blood pressure and serum lipids in the three age groups. The proportion of the variance of A in the overall variation was defined as heritability.

We fitted homogeneity models among the three age groups to examine whether the variance components could be equal. A likelihood ratio test was performed between the homogeneity models and the suitable models in the total population and if $p < .05$, we could judge that the homogeneity model was unacceptable. Based on the homogeneity models, heritability estimates in the three age groups and the 95% confidence intervals (CI), we observed whether heritability changed with age.

Bivariate Structural Equation Modeling

Based on the suitable models of univariate SEM, bivariate SEM was fitted in the three age groups. The bivariate Cholesky model was used to calculate the genetic correlation coefficient (Ra) and

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

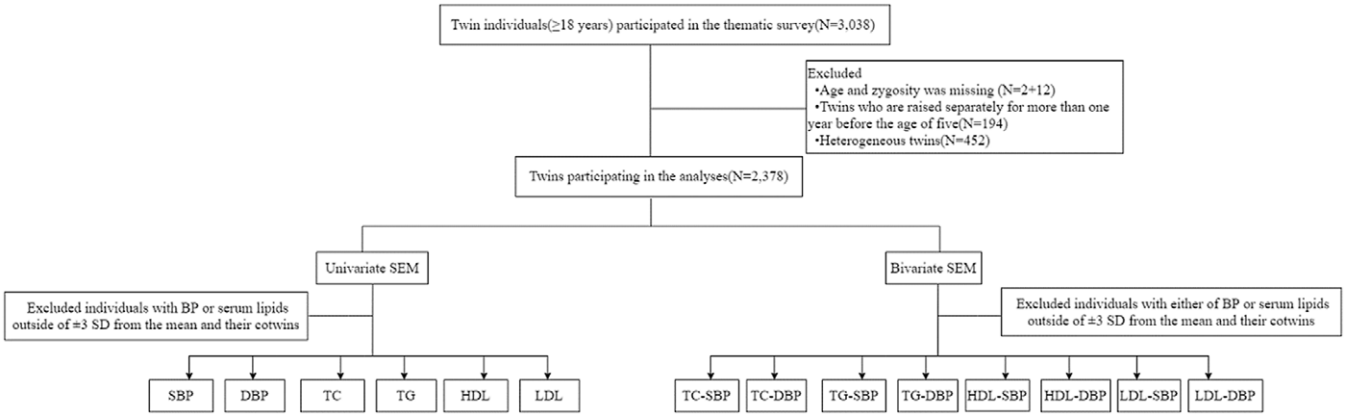


Figure 1. Flow chart of the study population and data analysis.
Note: SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol. SEM, structural equation model; SD, standard deviation

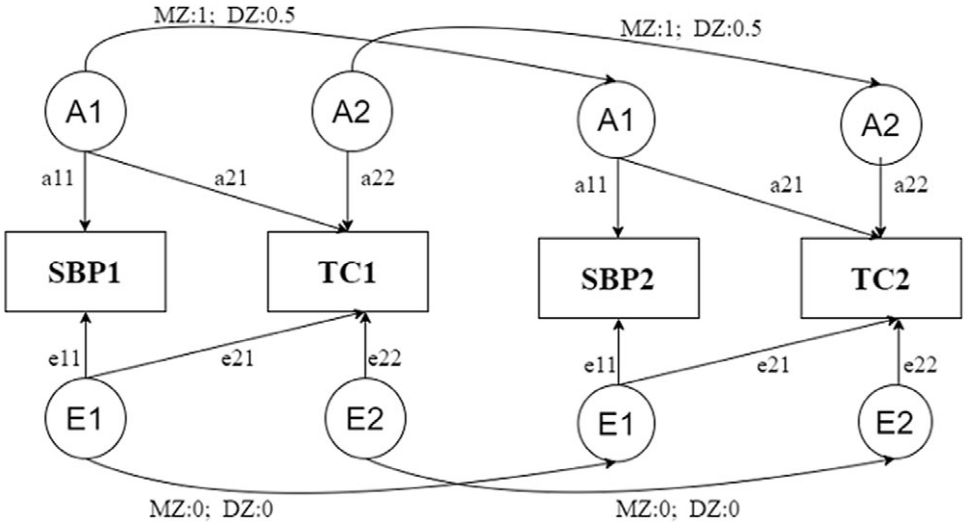


Figure 2. Bivariate Cholesky model for SBP and TC as examples.
Note: The observed phenotypes of twin 1 and twin 2 are in boxes, and the latent variables are in circles. SBP, systolic blood pressure; TC, total cholesterol; a11, the effect of the additive genetic component on SBP; e11, the effect of the unique environmental component on SBP; a21, the combined effect of the additive genetic component on SBP; a22, the effect of additive genetic component on TC; e22, the effect of the unique environmental component on TC.

environmental correlations coefficient (Re) between BP and serum lipids indicators (Figure 2)

Like univariate SEM, we constructed homogeneity models pairwise for each of the three age groups with equal variance components between two age groups. We compare the homogeneity model with the suitable model for each age group using a likelihood ratio test. If $p < .05$, the homogeneity model was considered unacceptable. Examining the model fitting effect, the selection criteria of the suitable models and the criteria for determining whether different age groups are the same refer to the univariate SEM.

Results

Sample Characteristics

A total of 2378 (1189 pairs) adult twins were included in this study, of which 840 pairs (70.6%) were MZ. The average age of all participants was 48.2 ± 12.1 years old, and 68.5% were male. SBP and DBP increased with age among the three age groups, while no age-dependent trend was observed in serum lipid indicators (Table 1).

Heritability of Blood Pressure and Serum Lipids

In the total population and the three age groups, the model-fitting process is shown in Supplementary Tables S1–S5. In the total population, except for TGs, the suitable models for other phenotypes were ACE models, and the suitable model for TGs was AE model (Supplementary Table 4). The heritability was 0.39 for SBP, 0.39 for DBP, 0.45 for TC, 0.59 for TGs, 0.47 for HDL-C and 0.37 for LDL-C (Table 2).

Among the three age groups, we chose AE models as the suitable models for phenotypes. The homogeneity model for only SBP and DBP could be accepted ($p < .001$; Supplementary Table 4), indicating that heritability is not equal among the three age groups. Since unique environmental variation increased, the heritability of SBP and DBP declined in the older adult group compared to the younger group. For TC and TGs, heritability point values at different age groups showed a downward trend as genetic variation declined and environmental variation increased. Still, the heritability difference was insignificant ($p > .05$). Genetic and environmental variation of HDL-C and LDL-C was similar among the three groups, so the heritability remains relatively stable with age (Figure 3, Supplementary Figure S1).

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

Table 1. Characteristics of the study participants ($N = 2378$)

Characteristic	Overall	≤45 years old	46–55 years old	>55 years old
No.	2378	934	848	596
Age (years)	48.2 (12.1)	36.2(6.6)	50.4 (2.8)	63.7 (5.9)
Sex/zygosity(%)				
MZF	560 (23.5)	242(25.9)	200 (23.5)	118 (19.8)
MZM	1120 (47.1)	364(39.0)	410 (48.4)	346 (58.0)
DZF	188 (8.0)	86(9.2)	76 (9.0)	26 (4.4)
DZM	510 (21.4)	242(25.9)	162 (19.1)	106 (17.8)
Blood pressure				
SBP(mmHg)	134.4 (22.5)	125.2(18.1)	135.8 (21.3)	146.9 (23.8)
DBP(mmHg)	82.6 (13.9)	79.0(13.3)	84.5 (13.8)	85.6 (13.6)
Serum lipid				
TC(mmol/L)	4.8 (4.2,5.4)	4.8(4.1,5.3)	4.9 (4.3,5.5)	4.8 (4.2,5.4)
TGs(mmol/L)	1.4 (0.9,2.1)	1.4(0.9,2.3)	1.4 (1.0,2.2)	1.3 (0.9,1.9)
LDL-C(mmol/L)	2.5 (2.0,3.0)	2.4(2.0,2.9)	2.6 (2.1,3.1)	2.5 (2.0,3.0)
HDL-C(mmol/L)	1.3 (1.1,1.5)	1.3(1.1,1.5)	1.3 (1.1,1.6)	1.4 (1.1,1.6)

Note: SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MZF, monozygotic female; MZM, monozygotic male; DZF, dizygotic female; MZM: monozygotic male. Data are reported as the mean (standard deviation) for blood pressure indicators and the median (interquartile range) for serum lipid indicators.

Table 2. Parameter estimates (95% CI) from the univariate model of blood pressure and serum lipids in the total population

Phenotype	A (95% CI)	C (95% CI)	E (95% CI)	h2 (95% CI)	e2 (95% CI)
SBP	0.31 (0.16, 0.49)	0.13 (0.00, 0.29)	0.36 (0.33, 0.40)	0.39 (0.20, 0.59)	0.44 (0.40, 0.49)
DBP	0.35 (0.19, 0.54)	0.17 (0.00, 0.33)	0.38 (0.35, 0.42)	0.39 (0.21, 0.59)	0.42 (0.38, 0.47)
TC	0.45 (0.28, 0.64)	0.18 (0.00, 0.33)	0.36 (0.33, 0.40)	0.45 (0.29, 0.64)	0.37 (0.33, 0.41)
TGs	0.59 (0.52, 0.66)	0.00 (0.00, 0.00)	0.42 (0.38, 0.46)	0.59 (0.54, 0.63)	0.41 (0.37, 0.46)
HDL-C	0.45 (0.30, 0.63)	0.22 (0.05, 0.38)	0.29 (0.26, 0.32)	0.47 (0.31, 0.65)	0.30 (0.27, 0.33)
LDL-C	0.37 (0.21, 0.56)	0.25 (0.08, 0.41)	0.37 (0.33, 0.41)	0.37 (0.21, 0.56)	0.37 (0.33, 0.41)

Note: SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol. A, additive genetic component; E, unique environmental component. h2, heritability, that is, the ratio of A to the total variance; e2, the proportion of E to the total variance.

Genetic and Environmental Correlation between Blood Pressure and Serum Lipids Indicators

There were eight phenotype-phenotype pairs of blood pressure and serum lipid indicators. For each of them we chose the AE model as the binary Cholesky decomposition model according to the suitable model of univariate SEM. The model-fitting process is shown in Supplementary Tables S6–S9. Therefore, the phenotypic correlation (Rph) was decomposed into additive genetic correlation (Ra) and unique environmental correlation (Re).

In total populations, the Rphs were 0.15 for TC&SBP, 0.17 for TGs&SBP, 0.11 for LDL-C&SBP, −0.04 for HDL-C&SBP, 0.19 for TC&DBP, 0.23 for TGs&DBP, 0.16 for LDL-C&DBP and −0.08 for HDL-C&DBP. For these phenotypes-phenotypes pairs, the Ra ranged from 0.00 to 0.36. 0–85% of the phenotypic correlations were determined by genetic factors, and the remaining 15–100% were explained by environmental factors (Table 3).

Among the age groups, Rph and Ra of each phenotype-phenotype pair showed the same change trend. For TC&SBP, TC&DBP, TGs&SBP and TGs&DBP, both the Rph and Ra declined with age, and the difference of Ra between the young group and the older adult group was statistically significant ($p < .05$) according to the 95% CI and the unacceptable homogeneity models (Figures 4 and 5, Supplementary Table S9). However, for LDL-C&DBP, HDL-C&SBP, and HDL-C&DBP, the lowest Ra and Rph appeared in the middle age group. The rE did not follow any pattern with age, and the differences in Re among the three age groups for all phenotype-phenotype pairs were not statistically significant according to the point value of Re and their 95% CI (Figure 5).

Discussion

To our knowledge, age plays a crucial role in genetic and environmental influences on blood pressure and serum lipids. As expected, the heritability and genetic correlations of most

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

Table 3. Parameter estimates (95% CI) from the bivariate model of blood pressure and serum lipid in total population

Phenotype1	Phenotype2	Rph (95% CI)	Ra (95% CI)	Re (95% CI)	Pa (95% CI)	Pe (95% CI)
SBP	**TC**	0.15 (0.10, 0.20)	0.09 (0.00, 0.41)	0.19 (0.01, 0.34)	0.22 (0.00, 0.96)	0.78 (0.04, 1.00)
	TGs	0.17 (0.12, 0.21)	0.28 (0.16, 0.46)	0.09 (0.05, 0.15)	0.70 (0.50, 0.85)	0.30 (0.15, 0.50)
	LDL-C	0.11 (0.06, 0.15)	0.00 (0.00, 0.33)	0.16 (0.00, 0.25)	0.00 (0.00, 1.00)	1.00 (0.00, 1.00)
	HDL-C	−0.04 (−0.08, 0.00)	−0.07 (−0.22, 0.00)	−0.02 (−0.11, 0.00)	0.73 (0.00, 1.00)	0.27 (0.00, 1.00)
DBP	**TC**	0.19 (0.14, 0.23)	0.22 (0.00, 0.50)	0.16 (0.01, 0.37)	0.50 (0.00, 0.97)	0.50 (0.03, 1.00)
	TGs	0.23 (0.19, 0.27)	0.36 (0.24, 0.54)	0.14 (0.08, 0.19)	0.69 (0.55, 0.81)	0.31 (0.19, 0.45)
	LDL-C	0.16 (0.11, 0.21)	0.04 (0.00, 0.38)	0.23 (0.02, 0.35)	0.09 (0.00, 0.91)	0.91 (0.09, 1.00)
	HDL-C	−0.08 (−0.12, −0.03)	−0.17 (−0.32, 0.00)	−0.02 (−0.16, 0.00)	0.85 (0.00, 1.00)	0.15 (0.00, 0.58)

Note: Ra, genetic correlation between 2 phenotypes; Re, unique environmental correlation between 2 phenotypes; Rph, phenotypic correlation between two phenotypes; Pa, percentage of genetic contributions to the correlation between two phenotypes; Pe, percentage of environmental contributions to the correlation between two phenotypes; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

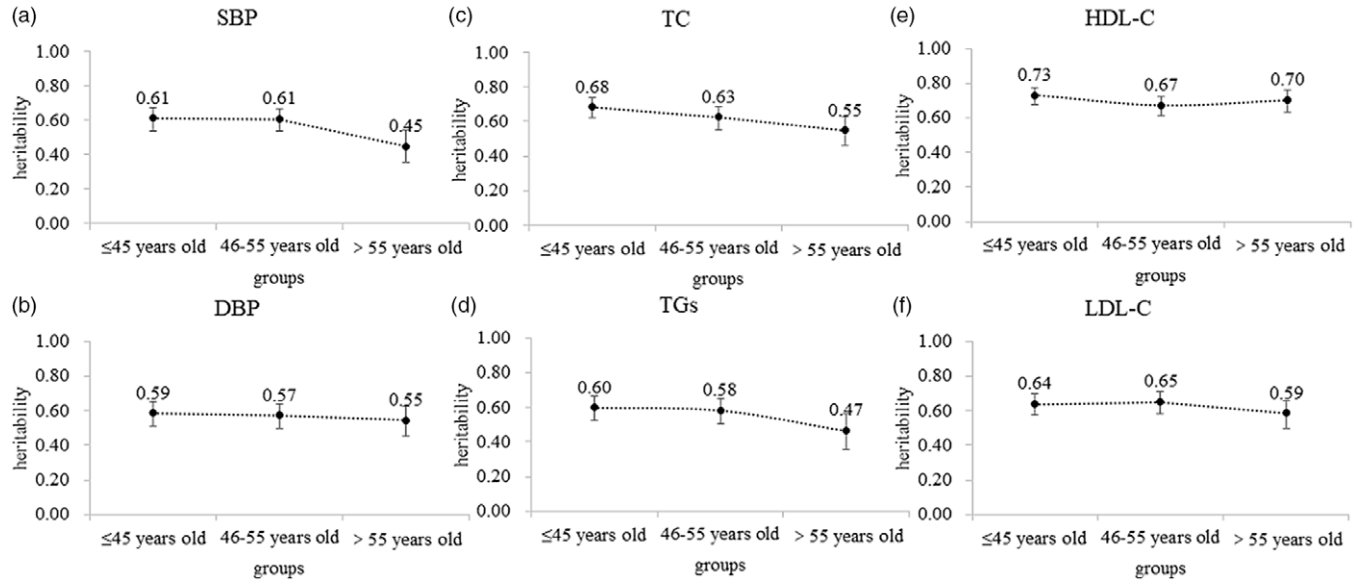


Figure 3. Heritability of blood pressure and serum lipids by three age groups.
Note: SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TGs, triglycerides; LDL-C,low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

phenotype and phenotype-phenotype pairs decreased with age. However, some heritability or genetic correlations still reached extreme values in the middle-aged group.

Univariate analysis showed that the overall heritability of blood pressure and serum lipid indicators was estimated to be moderate to high, suggesting that these phenotypic variations were mainly due to genetic effects, which is consistent with previous findings (Jermendy et al., 2011; Liao et al., 2017; Liao et al., 2015; Snieder et al., 1999). The environmental factors that lead to phenotypic variation are mainly the unique environment rather than the common environment, which may be because most of the participants in this study were middle-aged and elderly, and the shared environmental exposure of twins may no longer play a significant role (Benyamin et al., 2007). We found that the heritability of SBP and DBP declined with age. The unacceptable homogeneity models support this hypothesis. Differences in the heritability across age groups may reflect the cumulative effect of the unique environment, such as smoking, drinking and exercising (Heller et al., 1993; Province et al., 1989). Alternatively, these results may reflect cohort differences. In line with previous studies (Snieder et al., 1999), the heritability of TC, TG, HDL-C and LDL-C remains relatively stable with age. However, other studies found that the heritability of serum lipids fluctuates with age, possibly due to the age-dependent expression of lipid-related genes (Heller et al., 1993; Simino et al., 2014). In addition, women enter menopause in this age group (46−55 years old; Marlatt et al., 2022). In this period, LDL-C levels rise rapidly, while HDL-C levels fall (Marlatt et al., 2022; Snieder et al., 1999) when a lack of ovarian hormones affects lipid metabolism. It may be due to the complex changes in genetic and environmental factors that lead to stability in heritability. The role of unique environmental factors such as lifestyle becomes increasingly important as age increases. Thus, unique environmental factors should be given more attention at a young age to prevent or slow disease progression.

Most previous studies have analyzed blood pressure and serum lipid indicators separately. Few have investigated genetic and

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

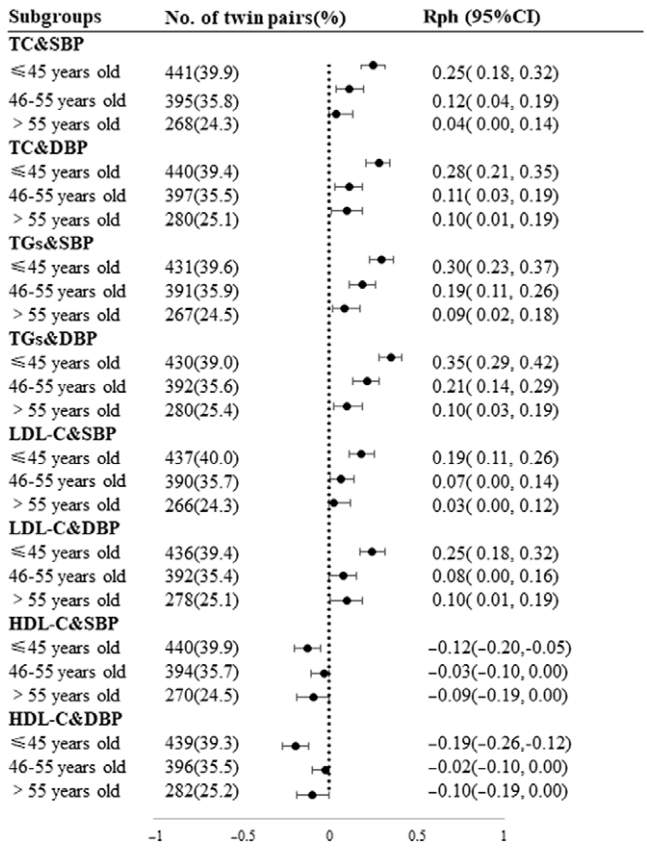


Figure 4. Phenotypic correlation (95% CIs) from the best-fitting bivariate AE model of blood pressure and serum lipid in three age groups.
Note: Rph, phenotypic correlation between two phenotypes; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

environmental correlations in metabolic syndrome subphenotypes, including blood pressure, TG and HDL-C, instead of the four serum lipid indicators (Benyamin et al., 2007; Duan et al., 2011; Panizzon et al., 2015; Zhang et al., 2009). We found a weak genetic correlation (0.00−0.36) between blood pressure and serum lipids through bivariate analysis, consistent with previous twin studies (Benyamin et al., 2007; Duan et al., 2011; Panizzon et al., 2015; Zhang et al., 2009). Genetic correlation provides epidemiological evidence for exploring the overlapping genes between blood pressure and serum lipids. The trend may indicate the change in genetic effects of overlapping genes over time. A genetic correlation of 1.00 or −1.00 suggests that the genetic effects on the two traits overlap entirely, while a genetic correlation of zero indicates that completely different genes affect the two traits. Apart from genetic correlation, the percentage of genetic contributions to the correlation between two phenotypes (Pa) reflects the magnitude of genetic influences on phenotypes. This study found that genetics contributed significantly (0−85%) to the correlation between blood pressure and serum lipids, suggesting that genes affecting blood pressure overlap with genes affecting serum lipids. A large genomewide association study (GWAS; Willer et al., 2013) found that 20 and 29 loci associated with lipid levels were associated with SBP and DBP respectively, of which the BRAP gene (Kim et al., 2016) and the SLC39A8 gene (Yao et al., 2015) were duplicated in other studies. Another GWAS on Koreans found that *BRAP, ACAD10,* and *ALDH2* were genes shared by SBP and DBP with TGs, HDL-C and LDL-C respectively (Kim et al., 2016). Two other GWAS on the pleiotropy of coronary heart disease risk loci found that the *SH2B3* gene was the locus associated with SBP, DBP, HDL-C, and LDL-C (CARDIoGRAMplusC4D Consortium et al., 2013; Webb et al., 2017). These genomic studies support our findings that there are overlapping genes between blood pressure and serum lipids.

In the bivariate analysis of age groups, as expected, the phenotypic correlation between serum lipids and blood pressure generally declined with age, and the phenotypic correlation was highest in the young group. Considering that dyslipidemia in China has a higher prevalence and lower awareness, treatment and control rates than hypertension (Pan et al., 2016), we should pay more attention to the serum lipid levels of hypertensive patients to detect dyslipidemia early. Similarly, the genetic correlation of TC, TG and blood pressure phenotype-phenotype pairs declined with age. This shows that the genetic effect decreases, and the environmental effect increases with age. Compared with the results of a study using young twins with an average age of 21.31 years old (McCaffery et al., 1999), the proportions of genetic correlations to total phenotypic correlations (Pa) of TC&SBP, TC&DBP, TGs&SBP and TGs&DBP in the youth twin study (Pa = 93%, 77%, 90% and 100% respectively) were greater than those in our study (Pa = 22%, 50%, 70% and 69% respectively), indirectly indicating that age has a moderating effect on the genetic correlations of these phenotype-phenotype pairs. It is physiologically plausible that age modulates the magnitude of genetic correlations. Aging leads to cumulative exposure to lifestyle choices, such as smoking, alcohol consumption, diet and exercise, which may affect the regulation of blood pressure (Hart & Charkoudian, 2014), lipid metabolism (Liu & Li, 2015) and the expression of lipid-related genes (Simino et al., 2014).

Both univariate and bivariate results indicate that unique environmental factors play an increasingly important role in the heritability of traits or the phenotypic correlations of phenotype pairs with age. A study comparing the differences in environmental factors between dyslipidemia and hypertension in the Guangxi Zhuang population in China found that common risk factors for hypertension and dyslipidemia included age, total energy and total fat intake (Ruixing et al., 2009). In addition, smoking (Handa et al., 1990), alcohol consumption (Handa et al., 1990; Wakabayashi, 2009), and BMI (Wakabayashi, 2009) are also common risk factors for dyslipidemia and hypertension. These environmental risk factors specifically accumulate with age. Therefore, common risk factors can be intervened early to prevent and control dyslipidemia and hypertension simultaneously.

Our study focuses on the heritability and genetic correlation between blood pressure and serum lipids, and we also attempted to explore the effect of age on heritability and genetic correlation. However, our study also has limitations. First, although our sample size is sufficient to detect the change trends with age, a larger sample size is required for gender-stratified analysis. Second, this study is cross-sectional and cannot directly use correlation to make causal inferences.

Conclusion

In conclusion, our study shows the importance of genetic and independent environmental factors in the correlation of blood pressure with serum lipids. It shows the influence of age on genetic and independent environmental contributions. Genetic factors play a more significant role in the correlation between blood pressure and serum lipids in young age, whereas cumulative

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

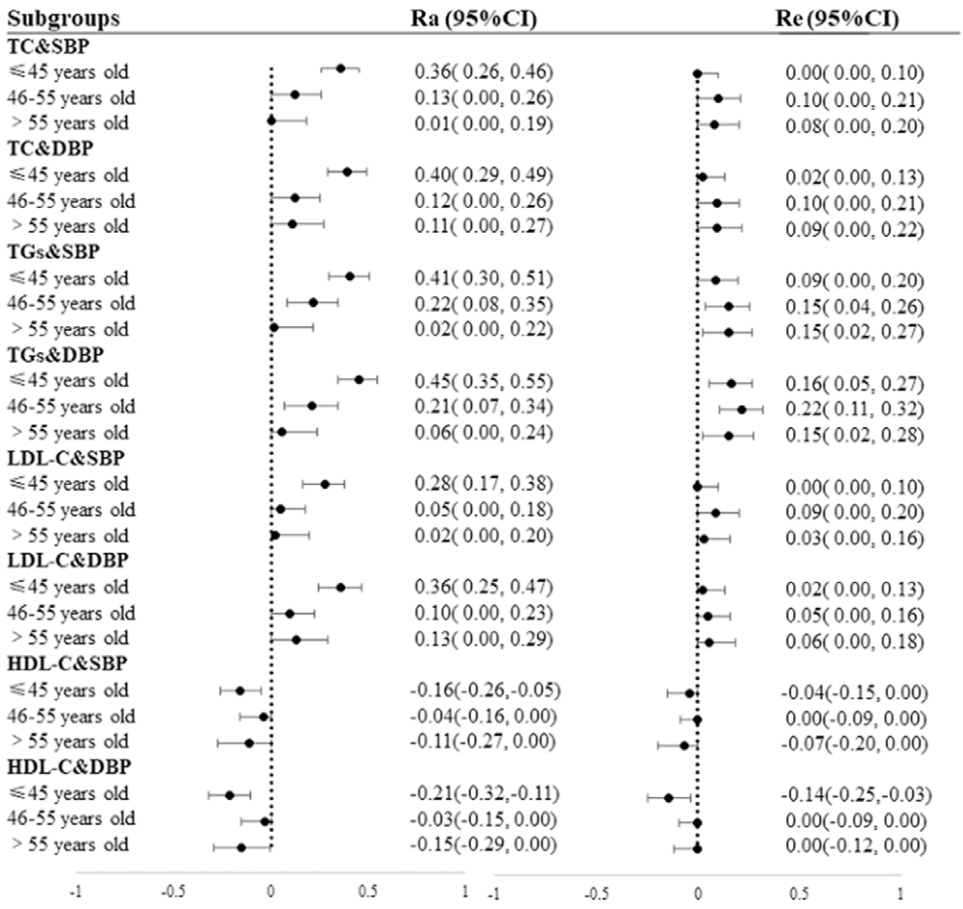


Figure 5. Genetic and environmental correlation (95% CI) from the best-fitting bivariate AE model of blood pressure and serum lipid in three age groups.
Note: Ra, genetic correlation between two phenotypes; Re, unique environmental correlation between two phenotypes; SBP, systolic blood pressure; DBP, diastolic blood pressure; TC, total cholesterol; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

unique environmental effects play a more substantial role in old age. Studies of common genes or epigenetic loci between blood pressure and blood lipids should consider using younger participants. Patients with hypertension or dyslipidemia should pay more attention to their own blood pressure and serum lipids and take intervention measures for common risk factors. Furthermore, more prospective studies are needed to validate our findings, and further studies are required to identify specific genes that affect blood pressure and serum lipids.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/thg.2023.25.

Data availability statement. The datasets generated during and analyzed during the current study are not publicly available but are available from the corresponding author at reasonable request.

Acknowledgments. We thank all the participants and project staff who participated in the Chinese National Twin Registry for their contributions.

Author contributions. Wenjing Gao: conceptualization, methodology, writing - review & editing; Zengchang Pang, Min Yu, Hua Wang, Xianping Wu, and Yu Liu: Investigation; Yutong Wang, Weihua Cao, Jun Lv, Canqing Yu, Shengfeng Wang, Tao Huang, Dianjianyi Sun, Chunxiao Liao, Yuanjie Pang and Runhua Hu: Writing — review and editing; Liming Li: supervision, project administration, funding acquisition; Ke Miao: software, writing original draft

Financial support. This study was funded by the National Natural Science Foundation of China (82073633, 81973126, 81573223), the Special Fund for Health Scientific Research in the Public Welfare (201502006, 201002007), and Peking University Outstanding Discipline Construction Project of Epidemiology and Biostatistics.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Ethics statement. This study was granted by the Biomedical Ethics Committee at Peking University, Beijing, China (approval number: IRB00001052-13022, IRB00001052-14021). All participants provided written informed consent when investigated. All methods were carried out in accordance with the relevant guidelines and regulations.

References

Benyamin, B., Sørensen, T. I. A., Schousboe, K., Fenger, M., Visscher, P. M., & Kyvik, K. O. (2007). Are there common genetic and environmental factors behind the endophenotypes associated with the metabolic syndrome? *Diabetologia*, *50*, 1880–1888. https://doi.org/10.1007/s00125-007-0758-1

Bønaa, K. H., & Thelle, D. S. (1991). Association between blood pressure and serum lipids in a population. The Tromsø Study. *Circulation*, *83*, 1305–1314. https://doi.org/10.1161/01.cir.83.4.1305

CARDIoGRAMplusC4D Consortium; Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R., Ingelsson, E., Saleheen, D., Erdmann, J., Goldstein, B. A., Stirrups, K., König, I. R., Cazier, J. B., Johansson, A., Hall, A. S., Lee, J. Y., Willer, C. J., Chambers, J. C., Esko, T., . . . Samani, N. J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics*, *45*, 25–33. https://doi.org/10.1038/ng.2480

Deshmukh, M., Lee, H. W., McFarlane, S. I., & Whaley-Connell, A. (2008). Antihypertensive medications and their effects on lipid metabolism. *Current Diabetes Reports*, *8*, 214–220. https://doi.org/10.1007/s11892-008-0037-7

Duan, H., Pang, Z., Zhang, D., Li, S., Kruse, T. A., Kyvik, K. O., Christensen, K., & Tan, Q. (2011). Genetic and environmental dissections of sub-phenotypes of metabolic syndrome in the Chinese population: A twin-based

heritability study. *Obesity Facts*, *4*, 99–104. https://doi.org/10.1159/000327735

Gao, W., Cao, W., Lv, J., Yu, C., Wu, T., Wang, S., Meng, L., Wang, D., Wang, Z., Pang, Z., Yu, M., Wang, H., Wu, X., Dong, Z., Wu, F., Jiang, G., Wang, X., Liu, Y., Deng, J., … **Li, L.** (2019). The Chinese National Twin Registry: A 'gold mine' for scientific research. *Journal of Internal Medicine*, *286*, 299–308. https://doi.org/10.1111/joim.12926

GBD 2019 Risk Factors Collaborators. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet*, *396*, 1223–1249. https://doi.org/10.1016/s0140-6736(20)30752-2

Handa, K., Tanaka, H., Shindo, M., Kono, S., Sasaki, J., & Arakawa, K. (1990). Relationship of cigarette smoking to blood pressure and serum lipids. *Atherosclerosis*, *84*, 189–193. https://doi.org/10.1016/0021-9150(90)90090-6

Hart, E. C., & Charkoudian, N. (2014). Sympathetic neural regulation of blood pressure: influences of sex and aging. *Physiology (Bethesda)*, *29*, 8–15. https://doi.org/10.1152/physiol.00031.2013

Heller, D. A., de Faire, U., Pedersen, N. L., Dahlén, G., & McClearn, G. E. (1993). Genetic and environmental influences on serum lipid levels in twins. *New England Journal of Medicine*, *328*, 1150–1156. https://doi.org/10.1056/nejm199304223281603

Hurtubise, J., McLellan, K., Durr, K., Onasanya, O., Nwabuko, D., & Ndisang, J. F. (2016). The different facets of dyslipidemia and hypertension in atherosclerosis. *Current Atherosclerosis Reports*, *18*, 82. https://doi.org/10.1007/s11883-016-0632-z

Jermendy, G., Horváth, T., Littvay, L., Steinbach, R., Jermendy, A. L., Tárnoki, A. D., Tárnoki, D. L., Métneki, J., & Osztovits, J. (2011). Effect of genetic and environmental influences on cardiometabolic risk factors: A twin study. *Cardiovascular Diabetology*, *10*, 96. https://doi.org/10.1186/1475-2840-10-96

Kim, Y. K., Hwang, M. Y., Kim, Y. J., Moon, S., Han, S., & Kim, B.-J. (2016). Evaluation of pleiotropic effects among common genetic loci identified for cardio-metabolic traits in a Korean population. *Cardiovascular Diabetology*, *15*, Article 20. https://doi.org/10.1186/s12933-016-0337-1

Lepira, F. B., M'Buyamba-Kabangu, J. R., Kayembe, K. P., & Nseka, M. N. (2005). Correlates of serum lipids and lipoproteins in Congolese patients with arterial hypertension. *Cardiovascular Journal of South Africa*, *16*, 249–255.

Liao, C., Gao, W., Cao, W., Lv, J., Yu, C., Wang, S., Zhao, Q., Pang, Z., Cong, L., Wang, H., Wu, X., & Li, L. (2017). Associations between obesity indicators and blood pressure in Chinese adult twins. *Twin Research and Human Genetics*, *20*, 28–35. https://doi.org/10.1017/thg.2016.95

Liao, C., Gao, W., Cao, W., Lv, J., Yu, C., Wang, S., Zhou, B., Pang, Z., Cong, L., Wang, H., Wu, X., & Li, L. (2015). Associations of body composition measurements with serum lipid, glucose and insulin profile: A Chinese twin study. *PLoS One*, *10*, e0140595. https://doi.org/10.1371/journal.pone.0140595

Liu, H. H., & Li, J. J. (2015). Aging and dyslipidemia: a review of potential mechanisms. *Ageing Research Reviews*, *19*, 43–52. https://doi.org/10.1016/j.arr.2014.12.001

Marlatt, K. L., Pitynski-Miller, D. R., Gavin, K. M., Moreau, K. L., Melanson, E. L., Santoro, N., & Kohrt, W. M. (2022). Body composition and cardiometabolic health across the menopause transition. *Obesity (Silver Spring)*, *30*, 14–27. https://doi.org/10.1002/oby.23289

McCaffery, J. M., Pogue-Geile, M. F., Debski, T. T., & Manuck, S. B. (1999). Genetic and environmental causes of covariation among blood pressure, body mass and serum lipids during young adulthood: A twin study. *Journal of Hypertension*, *17*, 1677–1685. doi: 10.1097/00004872-199917120-00004

Pan, L., Yang, Z., Wu, Y., Yin, R. X., Liao, Y., Wang, J., Gao, B., & Zhang, L. (2016). The prevalence, awareness, treatment and control of dyslipidemia among adults in China. *Atherosclerosis*, *248*, 2–9. https://doi.org/10.1016/j.atherosclerosis.2016.02.006

Panizzon, M. S., Hauger, R. L., Sailors, M., Lyons, M. J., Jacobson, K. C., Murray McKenzie, R., Rana, B., Vasilopoulos, T., Vuoksimaa, E., Xian, H., Kremen, W. S., & Franz, C. E. (2015). A new look at the genetic and environmental coherence of metabolic syndrome components. *Obesity (Silver Spring)*, *23*, 2499–2507. https://doi.org/10.1002/oby.21257

Province, M. A., Tishler, P., & Rao, D. C. (1989). Repeated-measures model for the investigation of temporal trends using longitudinal family studies: application to systolic blood pressure. *Genetic Epidemiology*, *6*, 333–347. https://doi.org/10.1002/gepi.1370060204

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, *3*, 119–133. https://doi.org/10.1093/bib/3.2.119

Ruixing, Y., Jinzhen, W., Weixiong, L., Yuming, C., Dezhai, Y., & Shangling, P. (2009). The environmental and genetic evidence for the association of hyperlipidemia and hypertension. *Journal of Hypertension*, *27*, 251–258. https://doi.org/10.1097/HJH.0b013e32831bc74d

Simino, J., Kume, R., Kraja, A. T., Turner, S. T., Hanis, C. L., Sheu, W., Chen, I., Jaquish, C., Cooper, R. S., Chakravarti, A., Quertermous, T., Boerwinkle, E., Hunt, S. C., & Rao, D. C. (2014). Linkage analysis incorporating gene-age interactions identifies seven novel lipid loci: The Family Blood Pressure Program. *Atherosclerosis*, *235*, 84–93. https://doi.org/10.1016/j.atherosclerosis.2014.04.008

Snieder, H., van Doornen, L. J., & Boomsma, D. I. (1999). Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *19*, 2826–2834. https://doi.org/10.1161/01.atv.19.12.2826

Tang, N., Ma, J., Tao, R., Chen, Z., Yang, Y., He, Q., Lv, Y., Lan, Z., & Zhou, J. (2022). The effects of the interaction between BMI and dyslipidemia on hypertension in adults. *Scientific Reports*, *12*, 927. https://doi.org/10.1038/s41598-022-04968-8

Tobin, M. D., Sheehan, N. A., Scurrah, K. J., & Burton, P. R. (2005). Adjusting for treatment effects in studies of quantitative traits: Antihypertensive therapy and systolic blood pressure. *Statistics in Medicine*, *24*, 2911–2935. https://doi.org/10.1002/sim.2165

Wakabayashi, I. (2009). Influence of body weight on the relationships of alcohol drinking with blood pressure and serum lipids in women. *Preventive Medicine*, *49*, 374–379. https://doi.org/10.1016/j.ypmed.2009.07.015

Wang, B., Gao, W., Yu, C., Cao, W., Lv, J., Wang, S., Pang, Z., Cong, L., Wang, H., Wu, X., & Li, L. (2015). Determination of zygosity in adult Chinese twins using the 450k methylation array versus questionnaire data. *PLoS One*, *10*, e0123992. https://doi.org/10.1371/journal.pone.0123992

Webb, T. R., Erdmann, J., Stirrups, K. E., Stitziel, N. O., Masca, N. G., Jansen, H., Kanoni, S., Nelson, C. P., Ferrario, P. G., König, I. R., Eicher, J. D., Johnson, A. D., Hamby, S. E., Betsholtz, C., Ruusalepp, A., Franzén, O., Schadt, E. E., Björkegren, J. L., Weeke, P. E., … **Kathiresan, S., Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators**. (2017). Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *Journal of the American College of Cardiology*, *69*, 823–836. https://doi.org/10.1016/j.jacc.2016.11.056

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang, H. Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T., … **Abecasis, G. R.; Global Lipids Genetics Consortium**. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, *45*, 1274–1283. https://doi.org/10.1038/ng.2797

Yao, C., Chen, B. H., Joehanes, R., Otlu, B., Zhang, X., Liu, C., Huan, T., Tastan, O., Cupples, L. A., Meigs, J. B., Fox, C. S., Freedman, J. E., Courchesne, P., O'Donnell, C. J., Munson, P. J., Keles, S., & Levy, D. (2015). Integromic analysis of genetic variation and gene expression identifies networks for cardiovascular disease phenotypes. *Circulation*, *131*, 536–549. https://doi.org/10.1161/circulationaha.114.010696

Zhang, S., Liu, X., Yu, Y., Hong, X., Christoffel, K. K., Wang, B., Tsai, H. J., Li, Z., Liu, X., Tang, G., Xing, H., Brickman, W. J., Zimmerman, D., Xu, X., & Wang, X. (2009). Genetic and environmental contributions to phenotypic components of metabolic syndrome: a population-based twin study. *Obesity (Silver Spring)*, *17*, 1581–1587. https://doi.org/10.1038/oby.2009.125

Zhang, X., Sun, Z., Zheng, L., Li, J., Liu, S., Xu, C., Li, J., Zhao, F., Hu, D., & Sun, Y. (2007). Prevalence of dyslipidemia and associated factors among the hypertensive rural chinese population. *Archives of Medical Research*, *38*, 432–439. https://doi.org/10.1016/j.arcmed.2006.12.005

https://doi.org/10.1017/thg.2023.25 Published online by Cambridge University Press

·基础理论与方法·

基于多中心数据库的观察性关联分析中残余混杂的控制与评估方法

郭金鑫[1]　赵厚宇[1]　詹思延[1,2,3]

[1]北京大学公共卫生学院流行病与卫生统计学系/重大疾病流行病学教育部重点实验室，北京 100191；[2]北京大学第三医院临床流行病学研究中心，北京 100191；[3]北京大学人工智能研究院智慧公众健康研究中心，北京 100871

通信作者：詹思延，Email：siyan-zhan@bjmu.edu.cn

【摘要】 基于健康医疗大数据的观察性研究越来越受到关注，残余混杂的控制与评估是其中亟须解决的关键问题，本文总结了多中心场景下开展关联分析的残余混杂统计学调整和敏感性分析方法。基于个体水平数据，可由各分中心使用断点回归等多种方法调整残余混杂，然后加权合并得到效应估计值；基于Meta水平数据，可采用贝叶斯Meta分析的方法获得调整后的合并效应值，也可开展残余混杂的敏感性分析，计算E值、$\hat{\rho}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$。上述方法应根据适用条件及优缺点进行合理选择，如利用分中心个体数据进行残余混杂调整，通常要求严格的研究设计，并面临较高的协调成本；贝叶斯Meta分析基于部分强假设；E值等敏感性分析结果仍需经过专业的判断，以评估残余混杂风险大小。因此，利用多中心数据库开展观察性关联分析时，残余混杂的控制与评估方法仍待进一步发展和完善。

【关键词】 多中心数据库；　观察性研究；　残余混杂；　统计方法

基金项目：国家自然科学基金（81973146）

Methods for controlling and evaluating residual confounding in the association analysis of observational study with a multicenter database

Guo Jinxin[1], Zhao Houyu[1], Zhan Siyan[1,2,3]

[1]Key Laboratory of Epidemiology of Major Diseases, Ministry of Education/Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; [2]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing 100191, China; [3]Center for Intelligent Public Health, Institute for Artificial Intelligence, Peking University, Beijing 100871, China

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

【Abstract】 The observational research based on big data in healthcare has attracted increasing attention, with the control and evaluation of residual confounding being the critical issue that needs to be solved urgently. This review summarized the methods for statistical adjustment and sensitivity analysis of residual confounding in the association analysis with a multicenter database. Based on individual-level data, the residual confounding can be adjusted in each subcenter using methods such as regression discontinuity design, while the pooled estimate can be obtained as a weighted average. Based on the center-level results, the Bayesian Meta-analysis method can adjust the pooled estimates. The sensitivity analysis of residual confounding can also be carried out using

DOI：10.3760/cma.j.cn112338-20230216-00083

收稿日期　2023-02-16　　**本文编辑**　万玉立

引用格式：郭金鑫，赵厚宇，詹思延. 基于多中心数据库的观察性关联分析中残余混杂的控制与评估方法[J]. 中华流行病学杂志，2023，44(8)：1296-1301. DOI：10.3760/cma.j.cn112338-20230216-00083.

Guo JX, Zhao HY, Zhan SY. Methods for controlling and evaluating residual confounding in the association analysis of observational study with a multicenter database[J]. Chin J Epidemiol, 2023, 44(8): 1296-1301. DOI: 10.3760/cma. j. cn112338-20230216-00083.

center-level data to calculate the *E*-value, $\hat{p}(q)$, $\hat{T}(r,q)$ and $\hat{G}(r,q)$. The abovementioned methods should be selected reasonably according to the requirements for practical applications, advantages, and disadvantages. For example, the use of subcenter individual data for residual confounding adjustment usually needs strict study design and frequent coordination; the Bayesian Meta-analysis is based on some strong assumptions; the interpretation of the results in the sensitivity analysis, such as *E*-value requires professional judgment to assess the risk of residual confounding. Therefore, the methods for controlling and evaluating residual confounding in association analysis based on multicenter databases still need further development and improvement.

【Key words】 Multicenter database; Observational study; Residual confounding; Statistical method

Fund program: National Natural Science Foundation of China (81973146)

真实世界证据在医学领域受到日益广泛的重视，基于健康医疗大数据的观察性研究逐渐成为热点。由于观察性研究的研究对象不是随机分组，即使通过多因素分析调整了部分已知混杂，关联效应结果仍可能受到未知、未测量或未完全控制的混杂因素的干扰，即残余混杂[1]。残余混杂的存在是流行病学因果推断面临的重大挑战之一，既往多项研究提出了基于单中心、集中式数据库的残余混杂控制与评估方法[2]。但健康医疗大数据常具有分散储存、多中心共享的特点，为保护个人隐私和信息安全，需要在不交换本地样本数据的前提下开展深度挖掘，这为残余混杂的处理带来更多方法学问题。本文对多中心数据场景下关联分析的残余混杂统计学调整或敏感性分析方法进行总结，阐述其基本原理、优缺点和应用，旨在为我国开展相关研究提供参考。

一、各分中心调整残余混杂后的加权合并

在多中心关联分析中，可基于文献报道的适用于单中心的残余混杂控制方法[2]，在每个分中心获得校正后估计值，然后协调中心通过加权平均的方式计算总效应，以实现对残余混杂的控制[3]。其中权重设置可使用倒方差法，即以各分中心效应值标准误平方的倒数作为权重，Mantel-Haenszel 法和 Peto 比值比法也是合并二分类变量相关效应值的常用方法。但值得注意的是，应用该分析思路需首先考虑各中心个体水平数据的可及性，以及不同残余混杂统计学调整方法的适用性。

以断点回归法为例，其适用场景为存在一个可观测的连续型变量决定暴露因素的分配，该配置变量大于某一临界值的个体均接受暴露（或暴露比例显著增加），小于临界值时均未暴露（或暴露比例显著降低）。当配置变量 X 由小于临界值增加到大于临界值时，结局变量 Y 出现跳跃，即结局变量 Y 和配置变量 X 的线性关系在临界值处存在一个断点，如果个体在该临界值附近时的其他影响因素无差别，那么造成结局变量 Y 在临界值处跳跃的唯一原因就是配置变量所指代的暴露的效应，由此估计暴露与结局的因果关联。

如浙江省从 2020 年开始为≥70 岁老年人群免费接种流感疫苗，为评估疫苗保护效果，Liu 等[4]使用断点回归设计开展了一项观察性研究：将年龄作为配置变量，以 70 岁为临界值。由于免费接种政策的执行，宁波市鄞州区 70 岁人群较 69 岁人群的流感疫苗接种率显著增加，在两者社会人口学特征相似的前提下，该研究类似于不完全依从的随机对照试验，69 岁和 70 岁人群住院率和死亡率的差异可归因于流感疫苗接种的差异。

断点回归法具有因果推断清晰、避免潜在混杂干扰的优点[5]，但只能解释临界值附近的局部平均效应，并需满足 2 个关键假设，①连续性：配置变量在临界值附近是连续的，且与结局变量的函数关系在临界值附近是连续的；②局部随机性：在临界值附近的个体是否接受暴露是随机的，潜在的混杂因素在临界值附近局域内均衡分布。

除断点回归法外，单中心研究的残余混杂控制方法还包括工具变量、双重差分、多重插补、扰动变量、本底事件率比校正、高维倾向性评分、经验分布校正等[6-7]，通常具有严格的研究设计要求。其中，工具变量法需选择符合条件的变量，指代研究的暴露因素。通过测量工具变量与暴露因素、工具变量与研究结局之间的关联，进而推断暴露因素与研究结局之间的关联。工具变量需满足 3 个关键假设：①相关性：与研究中的暴露因素相关；②排他性：仅通过暴露因素影响结局；③独立性：与影响“暴露-结局”关联的混杂因素无关。双重差分模型需寻找合适的暴露和对照组，分别对两组在暴露时点前后的结局进行第一次差分得到两组的变化量，消除个体不随时间变化的异质性。再对这两组变化

量进行第二次差分，以消除随时间变化的增量，得到暴露造成的净效应。该方法的假设条件：①暴露的实施对对照组不产生任何影响；②研究对象进入暴露组或对照组是完全随机的，暴露之外的因素对两组的影响相同；③暴露组和对照组的某些重要特征分布稳定，不随时间变化。

由于上述方法具有较为苛刻的适用条件，且在维护数据安全的前提下进行多中心整合存在较高的协调成本，这对各分中心调整残余混杂后加权合并构成了一定限制。

二、各分中心未调整残余混杂的模型校正

针对观察性研究中各分中心未控制残余混杂的情形，有研究提出了基于各分中心暴露效应点估计值及方差（Meta水平数据）的模型校正方法，规定了不同中心的偏倚分布特征，调整得到合并后效应估计值。如McCandless[8]提出了一种贝叶斯Meta分析方法，基于以下假设：①对于不同研究中心$j \in \{1, \cdots, k\}$，均存在一个二分类的残余混杂因素，与暴露因素无交互作用，且与其他已测量混杂因素相互独立；②对于$j \in \{1, \cdots, k\}$，残余混杂与研究结局的关联强度RR_j在对数转换后均服从先验分布$N(\mu_{RR}, \tau^2_{RR})$，其中超参数(μ_{RR}, τ_{RR})由研究者根据外部信息设定；③对于$j \in \{1, \cdots, k\}$，对照组中残余混杂因素的发生率p_{0j}，以及暴露组中该残余混杂的发生率p_{1j}，在logit转换后分别服从先验分布$N(\mu_{p_0}, \tau^2_{p_0})$、$N(\mu_{p_1}, \tau^2_{p_1})$，超参数$(\mu_{p_0}, \tau_{p_0})$、$(\mu_{p_1}, \tau_{p_1})$由研究者根据外部信息设定；④对于$j \in \{1, \cdots, k\}$，偏倚参数$(RR_j, p_{1j}, p_{0j})$在不同中心相互独立，且与各中心暴露与结局真实因果效应的对数取值θ_j无关；⑤对于$j \in \{1, \cdots, k\}$，θ_j服从先验分布$N(\mu, \tau^2)$，超参数μ服从超先验分布$N(0, 10^3)$，τ服从$U(0, 10^3)$。

基于偏倚参数(RR_j, p_{1j}, p_{0j})和分中心未调整残余混杂时暴露效应估计值的对数取值α_j，可得各中心调整后估计值的对数取值β_j：

$$\beta_j \approx \alpha_j - \Omega\left(RR_j, p_{1j}, p_{0j}\right)$$

其中：

$$\Omega\left(RR_j, p_{1j}, p_{0j}\right) = \log\frac{RR_j \times p_{1j} + (1 - p_{1j})}{RR_j \times p_{0j} + (1 - p_{0j})}$$

根据贝叶斯定理，参数后验分布正比于参数先验分布密度函数与似然函数的乘积，结合偏倚参数对效应估计进行的算术校正，进一步得到Meta分析的偏倚校正似然和后验。对于$j \in (1:k)$，当给定偏倚参数(RR_j, p_{1j}, p_{0j})，各分中心未调整残余混杂效应估计值的对数取值α_j及其标准误σ_j，则暴露与结局因果效应的对数取值θ_j的后验分布为正态分布，其均值为$\frac{\tau^2\beta_j + \sigma_j^2\mu}{\sigma_j^2 + \tau^2}$，即$\theta_j$的先验均值$\mu$与各中心调整残余混杂后效应估计值$\beta_j$的加权平均数。利用Metropolis-Hastings抽样算法对后验分布进行模拟逼近直至收敛，可计算得到各参数的均值、标准差、95%*CI*和中位数等信息：

$$\theta_j \left| RR_j, p_{1j}, p_{0j}, \alpha_j, \sigma_j \sim N\left(\frac{\tau^2\beta_j + \sigma_j^2\mu}{\sigma_j^2 + \tau^2}, \frac{\sigma_j^2\tau^2}{\sigma_j^2 + \tau^2}\right)\right.$$

贝叶斯统计本身考虑了混杂严重程度的不确定性，已在部分研究得到应用。例如，在一项探索他汀类药物使用与骨折发生关联的Meta分析中，纳入的17项观察性研究校正了年龄、性别、基础性疾病史等部分协变量，结果提示使用他汀药物对骨折发生具有保护作用（*HR*=0.73，95%*CI*：0.62~0.86），流感疫苗接种等健康保护行为可能是潜在的混杂因素，尚未纳入考虑，为进一步校正残余混杂的影响，研究者根据文献报道设置了偏倚参数(RR_j, p_{1j}, p_{0j})，即流感疫苗接种与骨折的关联强度RR_j、未使用他汀类药物人群的疫苗接种率p_{0j}、使用他汀类药物人群的疫苗接种率p_{1j}，参数取值分别为μ_{RR}=−0.62，τ_{RR}=0.27，μ_{p_0}=−0.48，τ_{p_0}=0.01，μ_{p_1}=−0.29，τ_{p_1}=0.04，使用上述算法进一步验证了他汀类药物的保护效应（*HR*=0.75，95%*CI*：0.63~0.89）[9]。

该方法的局限性：①实际情况可能经常违背部分强假设，如要求残余混杂为二分类变量，且与其他混杂因素相互独立；②选择合适的先验分布较为困难，包括对(μ_{RR}, τ_{RR})等超参数的设定，先验选择错误则有可能导致错误的结果。

三、加权合并后残余混杂的敏感性分析

1959年Cornfield等[10]提出运用敏感性分析评估观察性研究中的混杂偏倚。与回归分析得到调整后的效应估计值不同，敏感性分析无法真正控制混杂，但可以回答残余混杂能否全部或者部分解释研究中观察到的暴露效应，从而验证结果的稳健性。已有多名学者报道了单中心关联研究中残余混杂影响的敏感性分析方法[11-13]，其中包括边界因子法，该方法后续被推广至Meta分析[14]，为基于多中心数据库研究中的残余混杂评估提供了选择。

1. *E*值：边界因子法提出了一项关于因果关联证据（evidence）的测量指标——*E*值，主要用于回答下述问题：一项研究中残余混杂必须有多强才能

否定目前观察到的暴露与结局之间的关联？*E*值就是当观察到的暴露-结局关联可以全部被残余混杂解释时，该混杂因素与暴露、结局之间所需的最小关联强度。如果实际混杂强度小于*E*值所提示的强度阈值，则主要研究结果不能被残余混杂推翻为“无关联”。通过评估研究中混杂强度达到*E*值的可能性大小，该方法可以了解结果的稳健性。

*E*值计算方法：假设某研究的二分类暴露因素与研究结局的实际关联aRR_{ED}=1，残余混杂与暴露的关联强度为RR_{EU}，残余混杂与结局的关联强度为RR_{UD}，则未调整残余混杂时暴露与结局的关联$RR_{ED}=\frac{RR_{EU}\times RR_{UD}}{RR_{EU}+RR_{UD}-1}$。在$RR_{ED}$已知的情况下（即研究得到的暴露效应估计值），$RR_{EU}$与$RR_{UD}$相等时两者之积最小，此时：

$$E=RR_{EU}=RR_{UD}=RR_{ED}+\sqrt{RR_{ED}\times(RR_{ED}-1)}$$

上式适用于$RR_{ED}>1$时*E*值的计算，当$RR_{ED}<1$时应取倒数代入。研究者还总结了用于多分类结局、连续型结局、生存结局等效应指标的*E*值计算方法[15]。

在实际应用中，如有研究报道减肥手术可减少2型糖尿病患者5年内大血管事件的发生（*HR*=0.60，95%*CI*：0.42~0.86），取效应估计*HR*=0.60计算*E*值为2.72，取95%*CI*上限（0.86）计算*E*值为1.60。这提示如果研究中存在残余混杂，该混杂因素与减肥手术、大血管事件的关联强度需分别达到*HR*=2.72，效应估计值才能完全被残余混杂解释。在该研究中，一些强关联的大血管疾病危险因素还包括高血压（*HR*=1.09，95%*CI*：0.85~1.41）、血脂异常（*HR*=1.88，95%*CI*：1.34~2.63）和当前吸烟者（*HR*=1.48，95%*CI*：1.17~1.87）。由此推断，研究中存在残余混杂且关联强度远大于已知危险因素的可能性较低，结果较为可靠[16]。

在此基础上，研究者将边界因子法扩展至Meta分析[14]。对于一项基于多中心数据库的观察性研究，在加权合并各分中心效应估计值后，*E*值可根据合并效应值进行计算，方法同上。此时*E*值的含义是为了将合并效应估计推翻为“无关联”，所需各分中心残余混杂与暴露、结局间的平均关联强度。

*E*值的主要优点在于对研究中的残余混杂不做任何假设，如不需混杂因素为二分类变量等。*E*值计算过程简单易行，且适用于*RR*、*OR*、*HR*等不同类型评估指标。但缺点为*E*值通过定义残余混杂与暴露、结局的关联强度相同来获得，代表所需混杂的最小关联强度，在实际中该定义可能会被违背。*E*值的解释也需研究人员对取值大小合理性进行判断。

2. $\hat{p}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$：在基于多中心数据库的研究中，研究者还提出了统计量$\hat{p}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$用于残余混杂的敏感性分析，估计方法包括参数法和非参数法，可借助R软件包“EValue”或在线计算工具（https://www. evalue-calculator. com/Meta/）实现。

在参数法中[14]，设定暴露与结局事件之间真实效应强度阈值（如*RR*=1.1）的对数取值q，当暴露与结局之间真实效应值对数取值大于q的时候，则认为暴露与结局事件有足够大的关联关系。$p(q)$表示在所有研究中心内，暴露与结局之间真实效应值大于边界值q的中心数占比。假设：①对于不同研究中心$j\in(1,\cdots,k)$，存在残余混杂时暴露与结局之间的效应估计值为$\beta_{c,j}$，其对数取值表示为M_j^c，M_j^c服从正态分布$N(\mu^c,V^c)$；②对于$j\in(1,\cdots,k)$，暴露与结局之间的真实效应值为$\beta_{t,j}$，其对数取值表示为M_j^t，M_j^t服从正态分布$N(\mu^t,V^t)$；③对于$j\in(1,\cdots,k)$，引入一个偏倚因子$B_j^*=\log B_j$，B_j^*服从正态分布$N(\mu_{B^*},\delta_{B^*}^2)$，且$M_j^t+B_j^*=M_j^c$。

以$\hat{y}_R^c$表示由各中心效应估计值获得的合并效应值，τ_c^2为各中心间异质性的估计值，可推导出$p(q)$的一致估计量：

$$\hat{p}(q)=1-\Phi\left(\frac{q+\mu_{B^*}-\hat{y}_R^c}{\sqrt{\tau_c^2-\delta_{B^*}^2}}\right),\tau_c^2>\delta_{B^*}^2$$

式中，Φ为标准正态累积分布函数，在实际应用中将设定敏感性参数μ_{B^*}和$\delta_{B^*}^2$不同的取值，计算真实效应大于边界值的研究中心占比。如果对于较大的μ_{B^*}，仍有很多中心的真实效应大于边界值，这表明即使残余混杂的影响很大，在绝大部分研究中心内暴露与结局之间的关联仍然成立。

在上式的基础上，设定一个$p(q)$的阈值r，即真实效应大于边界值的研究中心占比小于r时，认为对于特定的μ_{B^*}，没有足够的研究支持暴露与结局之间具有足够强的关联关系。可进一步推导出使$p(q)$下降至阈值r以下所需偏倚因子$\hat{T}(r,q)$的大小：

$$\hat{T}(r,q)=\exp[\Phi^{-1}(1-r)\sqrt{\tau_c^2}-q+\hat{y}_R^c]$$

基于$\hat{T}(r,q)$，参照E值推导过程取$RR_{EU}=RR_{UD}$，可进一步得出使$p(q)$下降至阈值r以下所需的最小残余混杂强度$\hat{G}(r,q)$：

$$\hat{G}(r,q)=\hat{T}(r,q)+\sqrt{[\hat{T}(r,q)]^2-\hat{T}(r,q)}$$

由上可知，参数法需假设偏倚因子在各分中心间呈对数正态分布。Mathur和Vanderweele[17]也提出了非参数法作为补充，通过设定各分中心同质性的偏倚强度，获得$\hat{p}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$的估计。

在用于多中心研究的残余混杂评估时，$\hat{p}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$可以从不同的角度描述暴露与结局之间关联的稳健程度，有学者呼吁将其作为同类研究的常规报告项目[18]。如在一项评估2型糖尿病与肿瘤发病关联的Meta分析中，纳入了全球144项研究，主分析结果提示2型糖尿病人群较一般人群的肝癌发病风险升高（*RR*=2.23，95%*CI*：1.99~2.49）[19]。为评估残余混杂对结果的影响，该研究应用R软件包“EValue”计算了$\hat{p}(q)$和$\hat{G}(r,q)$。

定义关联强度阈值*RR*=1.1（q=ln1.1），即各项研究中2型糖尿病与肝癌的真实效应强度*RR*≥1.1才被视为足够强的关联，假设纳入的每项研究均存在残余混杂，且与2型糖尿病、肝癌的关联强度分别达到*RR*=1.5，根据上述参数取值可得$\hat{p}(q)$=98%（95%*CI*：95%~100%），含义：即使Meta分析纳入的各项研究均有一定强度的残余混杂（*RR*=1.5），经调整后真实关联强度≥1.1的研究占总数的比例仍达到98%。

另定义占比阈值r=70%，关联强度阈值仍为*RR*=1.1。即当真实关联强度*RR*≥1.1的研究数占比达到70%时，可将Meta分析中2型糖尿病与肿瘤发病的关联视为因果关联。此时计算$\hat{G}(r,q)$=2.9，含义：若要使真实关联强度大于阈值的研究数在纳入总数的占比降至70%，需各项研究的残余混杂与2型糖尿病、肝癌的关联强度均达到*RR*=2.9，该情形发生的可能性较小，提示2型糖尿病与肝癌间因果关联的结论较为可靠。

本文所述用于多中心数据库的观察性关联分析中残余混杂的控制与评估方法见表1。

四、讨论与总结

“健康中国2030”规划已提出加强健康医疗大数据的应用，数据引领决策也已成为医药卫生领域的发展趋势。基于多中心数据库的大型观察性研究在调整已测量混杂的同时，将越来越多地面临残余混杂控制与评估的方法学挑战，本文对相关分析思路及代表性方法进行了系统总结。

基于个体水平数据，在满足必要的假设前提情况下，可由各分中心使用工具变量、双重差分、断点回归等方法校正偏倚，然后将效应值发送至协调中心进行加权合并。这种研究思路可充分利用数据信息，提高统计效能，且与传统多因素校正回归等方法相比可实现对残余混杂更加有效的控制。但不足之处在于，即使是针对单中心研究的残余混杂控制方法，相关选择也并不丰富，且通常存在严格的研究设计要求及统计假设条件，如应用工具变量和断点回归法均需找到合适的工具变量或配置变量，双重差分法则需选择合适的对照保证组间可比性，在实际应用中可能会受到较大的限制，应严格把握使用前提。此外，为保护个人信息安全，如何协调各分中心在不交换本地数据前提下开展深入的统计分析和模型调整，这对项目管理提出了更高的要求。

因此，基于Meta水平数据的残余混杂控制方法成为了重要补充。有学者提出的贝叶斯Meta分析算法，可基于各分中心暴露效应点估计及方差获得调整残余混杂后的合并值，尽管部分强假设在实际研究中常不能完全满足，一些超参数的取值也难以设定，导致调整后的估计值可能存在偏倚，但仍可借助参数系列取值的方法验证效应估计的稳健性。此外，作为评估残余混杂影响的最后一道防线，敏感性分析虽然不能直接测量混杂效应强度或得到校正后效应估计值，但通过计算E值、$\hat{p}(q)$、$\hat{T}(r,q)$和$\hat{G}(r,q)$，能够以较少的统计假设实现对研究中残余混杂风险的评估，具有对数据分布特征要求较低、适用场景更广的特点。但缺点为计算结果大小的合理性需经过专业的判断，导致分析应用过程复杂，且存在一定的主观性。

值得一提的是，在研究设计和数据收集阶段的质量控制是消除混杂偏倚最为关键的环节，无法由后期的统计分析过程所替代。在此基础上，作为因果推断的重要研究领域，基于多中心数据库关联分析中残余混杂的控制与评估方法仍面临诸多问题，有待统计学者进一步发展和完善。

利益冲突　所有作者声明无利益冲突

作者贡献声明　郭金鑫：研究设计、文献查阅、文章撰写；赵厚宇：研究设计、论文修改、工作支持；詹思延：研究设计/指导、论文修改

表 1　基于多中心数据库的观察性关联分析中残余混杂的控制与评估方法

项目	数据要求[a]	适用条件	优势	局限性
统计学调整				
基于断点回归、工具变量、双重差分等方法，在各分中心调整残余混杂后加权合并	个体水平	因不同方法而异，以断点回归法为例： • 研究中存在合适的配置变量，影响暴露因素的分配 • 配置变量在临界值附近是连续的，且与研究结局的函数关系在临界值附近连续 • 在临界值附近的个体是否接受暴露是随机的	• 充分利用数据信息，提高分析效能，实现残余混杂的有效控制	• 需符合不同方法的研究设计要求和统计假设条件，应用范围受到限制 • 需在保护个人隐私和信息安全的前提下获取个体水平数据，研究协调管理要求高 • 不同方法可能存在各自的局限性，如断点回归法只能解释临界值附近的局部平均效应，较难推广到整体中
贝叶斯 Meta 分析	Meta 水平	满足下列假设： • 残余混杂为二分类变量，与暴露因素无交互作用，且与其他已测量混杂因素相互独立 • 不同中心间残余混杂与研究结局的关联强度呈对数正态分布 • 不同中心间残余混杂因素在暴露组和对照组的发生率经 logit 转换后均呈正态分布 • 不同中心间残余混杂与结局关联强度及残余混杂发生率均相互独立，且与暴露与结局真实因果效应无关	• 适用于个体数据无法获得的研究场景 • 贝叶斯统计本身考虑了混杂严重程度的不确定性	• 研究假设常被违背，残余混杂难以完全控制 • 模型中的先验分布难以选择，且对研究结果影响较大
敏感性分析				
E 值	Meta 水平	基于 *RR*、*OR*、*HR* 等效应指标估计值均可计算	• 适用于个体数据无法获得的研究场景 • 对研究中的残余混杂不做限制性假设	• *E* 值结果仍需专业判断，以评估残余混杂风险大小 • *E* 值为残余混杂与暴露、结局的关联强度相同时的取值，代表一种最小关联强度，实际可能被违背
$\hat{p}(q)$、$\hat{T}(r,q)$ 和 $\hat{G}(r,q)$	Meta 水平	满足下列假设： • 参数法：不同中心间偏倚因子效应强度呈对数正态分布；不同中心偏倚因子均与暴露和结局间的真实因果效应相互独立 • 非参数法：各分中心偏倚强度同质	• 适用于个体数据无法获得的研究场景 • 对每个中心内的残余混杂不做限制性假设； • 从各分中心效应估计分布差异的角度评估了研究结果稳健性	• 对不同中心间的残余混杂分布或同质性做了限制性假设 • $\hat{p}(q)$、$\hat{T}(r,q)$ 和 $\hat{G}(r,q)$ 结果仍需专业判断，以评估残余混杂风险大小

注：[a]Meta 水平数据指能够用于开展 Meta 分析的各分中心暴露效应点估计值及方差

参 考 文 献

[1] Verbeek JH, Whaley P, Morgan RL, et al. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies:a GRADE concept paper[J]. Environ Int, 2021, 157: 106868. DOI:10.1016/j.envint.2021.106868.

[2] Uddin MJ, Groenwold RHH, Ali MS, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview[J]. Int J Clin Pharm, 2016, 38(3):714-723. DOI:10.1007/s11096-016-0299-0.

[3] Mathur MB, Vanderweele TJ. Methods to address confounding and other biases in Meta-analyses: review and recommendations[J]. Annu Rev Public Health, 2022, 43:19-35. DOI:10.1146/annurev-publhealth-051920-114020.

[4] Liu GX, Liu ZK, Zhao HY, et al. The effectiveness of influenza vaccine among elderly Chinese: a regression discontinuity design based on Yinzhou regional health information platform[J]. Hum Vaccin Immunother, 2022, 18(6):2115751. DOI:10.1080/21645515.2022.2115751.

[5] Maciejewski ML, Basu A. Regression discontinuity design [J]. JAMA, 2020, 324(4): 381-382. DOI: 10.1001/jama.2020.3822.

[6] Zhang X, Stamey JD, Mathur MB. Assessing the impact of unmeasured confounders for credible and reliable real-world evidence[J]. Pharmacoepidemiol Drug Saf, 2020, 29(10):1219-1227. DOI:10.1002/pds.5117.

[7] Zhang X, Faries DE, Li H, et al. Addressing unmeasured confounding in comparative observational research[J]. Pharmacoepidemiol Drug Saf, 2018, 27(4):373-382. DOI: 10.1002/pds.4394.

[8] McCandless LC. Meta-analysis of observational studies with unmeasured confounders[J]. Int J Biostat, 2012, 8(2): 5. DOI:10.2202/1557-4679.1350.

[9] McCandless LC. Statin use and fracture risk: can we quantify the healthy-user effect? [J]. Epidemiology, 2013, 24(5):743-752. DOI:10.1097/EDE.0b013e31829eef0a.

[10] Cornfield J, Haenszel W, Hammond E, et al. Smoking and lung cancer: Recent evidence and a discussion of some questions[J]. J Natl Cancer Inst, 1959, 22(1):173-203.

[11] Bonvini M, Kennedy EH. Sensitivity analysis via the proportion of unmeasured confounding[J]. J Am Stat Assoc, 2022, 117(539): 1540-1550. DOI: 10.1080/01621459.2020.1864382.

[12] Peña JM. Simple yet sharp sensitivity analysis for unmeasured confounding[J]. J Causal Inference, 2022, 10(1):1-17. DOI:10.1515/jci-2021-0041.

[13] Ding P, Vanderweele TJ. Sensitivity analysis without assumptions[J]. Epidemiology, 2016, 27(3):368-377. DOI: 10.1097/ede.0000000000000457.

[14] Mathur MB, Vanderweele TJ. Sensitivity analysis for unmeasured confounding in Meta-analyses[J]. J Am Stat Assoc, 2020, 115(529):163-172. DOI:10.1080/01621459.2018.1529598.

[15] Vanderweele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value[J]. Ann Intern Med, 2017, 167(4):268-274. DOI:10.7326/m16-2607.

[16] Haneuse S, Vanderweele TJ, Arterburn D. Using the e-value to assess the potential effect of unmeasured confounding in observational studies[J]. JAMA, 2019, 321(6):602-603. DOI:10.1001/jama.2018.21554.

[17] Mathur MB, Vanderweele TJ. Robust metrics and sensitivity analyses for Meta-analyses of heterogeneous effects[J]. Epidemiology, 2020, 31(3): 356-358. DOI: 10.1097/ede.0000000000001180.

[18] Mathur MB, Vanderweele TJ. How to report E-values for Meta-analyses: recommended improvements and additions to the new GRADE approach[J]. Environ Int, 2022, 160:107032. DOI:10.1016/j.envint.2021.107032.

[19] Ling SP, Brown K, Miksza JK, et al. Association of type 2 diabetes with cancer: a Meta-analysis with bias analysis for unmeasured confounding in 151 cohorts comprising 32 million people[J]. Diabetes Care, 2020, 43(9): 2313-2322. DOI:10.2337/dc20-0204.

·综 述·

全球新冠疫苗研发与接种策略进展

吴侯[1] 刘珏[1] 刘民[1] 梁万年[2]
[1]北京大学公共卫生学院流行病与卫生统计学系,重大疾病流行病学教育部重点实验室(北京大学),北京 100191;[2]清华大学万科公共卫生与健康学院,北京 100084
通信作者:刘民,E-mail:liumin@bjmu.edu.cn;梁万年,E-mail:liangwn@tsinghua.edu.cn
DOI:10.16462/j.cnki.zhjbkz.2023.08.014

【摘要】 WHO 于 2023 年 5 月 5 日宣布新型冠状病毒感染(corona virus disease 2019, COVID-19)疫情不再构成国际关注的突发公共卫生事件。全球疫情相对稳定,但是各国仍不能放松对COVID-19的警惕。接种新冠疫苗仍是有效的预防手段。本文主要围绕当前全球新冠疫苗研发进展、不同国家疫苗接种策略调整情况等进行对比分析,并结合 WHO 推荐对新冠疫苗接种策略的最新指导意见,分析探讨全球新冠疫苗研发及各国疫苗接种策略对我国的启示,提出适合我国国情的针对性疫苗接种建议。

【关键词】 新型冠状病毒感染;疫苗;接种策略;世界卫生组织
【中图分类号】R373.1 【文献标识码】A 【文章编号】1674-3679(2023)08-0955-08
基金项目: 科技创新 2030—新一代人工智能重大项目(2021ZD0114104,2021ZD0114105,2021ZD0114101);国家自然科学基金(71934002,72122001)

Progress in global COVID-19 vaccine development and vaccination strategies *WU Yu*[1], *LIU Jue*[1], *LIU Min*[1], *LIANG Wannian*[2]
[1]*Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China*; [2]*Vanke School of Public Health, Tsinghua University, Beijing 100084, China*
Corresponding authors: LIU Min, E-mail: liumin@bjmu.edu.cn; LIANG Wannian, E-mail: liangwn@tsinghua.edu.cn

【Abstract】 On May 5, 2023, the World Health Organization declared that the outbreak of COVID-19 no longer constitutes a public health emergency of international concern. The global epidemic is relatively stable, but countries should not relax their vigilance against the novel coronavirus. Vaccination against COVID-19 remains an effective means of prevention. This study mainly makes a comparative analysis of the current progress of global COVID-19 vaccine research and development and the adjustment of vaccination strategies in different countries, combined with the latest guidance recommended by the World Health Organization on COVID-19 vaccine vaccination strategies, analyzes and discusses the inspiration of global COVID-19 vaccine research and development and national vaccination strategies for China, and puts forward targeted vaccination recommendations suitable for China's national conditions.

【Keywords】 COVID-19; vaccine; vaccination strategies; World Health Organization
Fund programs: National Key Research and Development Project of China (2021ZD0114104, 2021ZD0114105, 2021ZD0114101); National Natural Science Foundation of China (71934002, 72122001)
(*Chin J Dis Control Prev* 2023,**27**(8):955-962)

WHO 于 2023 年 5 月 5 日宣布新型冠状病毒感染(corona virus disease 2019, COVID-19)疫情不再构成国际关注的突发公共卫生事件(public health emergency of international concern, PHEIC),但同时强调,这并不意味着COVID-19不再是一个全球卫生威胁[1]。截至 2023 年 5 月 31 日,全球COVID-19病例数超过 7.67 亿,死亡人数超过 693 万[2]。尽管风险评估显示全球COVID-19的风险仍然很高,但已有

证据表明,由于人群对于新冠病毒感染的免疫水平提高以及全民疫苗接种,当前COVID-19对人群的健康危害有所降低[1]。对于老年人、慢性病患者、孕妇、医护人员和免疫力低下的人群而言,接种新冠疫苗仍是有效的预防手段。本文对当前全球新冠疫苗接种情况、疫苗研发进展,以及全球不同国家疫苗接种策略调整情况进行综述。

1 全球新冠疫苗接种情况

据牛津大学 Our World in Data 统计,截至2023 年 6 月,全球有 70.10%的人口至少接种了 1 剂COVID-19 疫苗,有 64.37%的人口进行了全程疫苗接种。在低收入国家中,30.10% 的人口至少接种了 1 剂 COVID-19疫苗,全球共接种超过 134 亿剂COVID-19 疫苗,目前疫苗接种剂量仍以 207 246 剂/日的速度增加。累计疫苗接种剂量前 3 位的国家是中国(34.9 亿)、印度(22.0 亿)、美国(6.7 亿)[3]。

从接种人口比例来看,全球约 1/3 国家的疫苗全程接种率达到 75%,其中卡塔尔、阿拉伯联合酋长国和文莱是全球新冠疫苗接种率最高的国家,疫苗覆盖率接近 100%。我国疫苗全程接种率为 91.9%。虽然全球新冠疫苗接种率正在上升,但全球仍有近 1/3 的人口尚未接种疫苗,尤其是中低收入地区。非洲地区的新冠疫苗至少 1 剂次疫苗接种率为全球最低(36.83%)。

2 全球新冠疫苗研究进展

据 WHO 报告,截至 2023 年 3 月底,全球已有 382 款候选新冠疫苗处于研发进程中,其中有 199 款处于临床前开发阶段,183 款处于临床试验阶段[4]。根据制备技术,新冠疫苗分为灭活病毒疫苗、减毒活疫苗、mRNA 疫苗、DNA 疫苗、病毒载体疫苗、病毒样颗粒疫苗和蛋白亚单位疫苗[5]。在所有临床疫苗技术路线中,蛋白亚单位疫苗占比最高(32.2%),其次为病毒载体疫苗(13.7%)。我国已经覆盖灭活疫苗、腺病毒载体疫苗、重组蛋白疫苗、减毒流感病毒载体疫苗和核酸疫苗 5 条技术路线。在候选疫苗的剂量选择中,2 剂疫苗占比最高,为 55.2%[4]。见表 1、表 2。

为了有效评估COVID-19疫苗的质量、安全性和有效性,WHO 使用紧急使用清单对疫苗第 2 阶段末期和第 3 阶段临床试验数据以及关于安全性、有效性、质量和风险管理计划的大量额外数据进行严格

表 1 WHO 报告的全球COVID-19候选疫苗

Table 1 Global COVID-19 vaccine candidates reported by the WHO

平台 Platform	临床疫苗技术路线 Candidates in clinical phase	候选疫苗数量 Candidate vaccines (no.)	候选疫苗(占比/%) Candidate vaccines (proportion/%)
PS	蛋白质亚基 Protein subunit	59	32.2
VVnr	病毒载体(非复制) Viral Vector (non-replicating)	25	13.7
DNA	DNA	17	9.3
IV	灭活病毒 Inactivated Virus	22	12.2
RNA	RNA	43	23.5
VVr	病毒载体(复制) Virus Vector (replicating)	4	2.2
VLP	类病毒颗粒 Virus Like Particle	7	3.8
VVr + APC	VVr + 抗原呈递细胞 VVr + Antigen Presenting Cell	2	1.1
LAV	减毒活病毒 Live Attenuated Virus	2	1.1
VVnr + APC	VVnr +抗原呈递细胞 VVnr + Antigen Presenting Cell	1	0.5
BacAg-SpV	细菌抗原-孢子表达载体 Bacterial antigen-spore expression vector	1	0.5

表 2 WHO 报告的全球COVID-19候选疫苗的接种剂量和接种时间

Table 2 The dose and timing of global COVID-19 vaccine candidates reported by the WHO

接种剂量和接种时间 Number of doses & schedule	候选疫苗数量 Number of candidate vaccines	候选疫苗(占比/%) Candidate vaccines (proportion/%)
1 剂 dose	47	25.7
d 0	47	
2 剂 doses	101	55.2
d 0 + 14	8	
d 0 + 21	37	
d 0 + 28	56	
3 剂 doses	2	1.14
d 0 + 28 + 56	2	
尚无数据 No data	33	18.0

审查。评估过程中会对突发事件造成的威胁以及使用该产品可能带来的好处与任何潜在风险进行权衡。作为紧急使用清单程序的一部分,疫苗生产企业必须承诺继续生成数据,以使疫苗得到完全许可和 WHO 预认证。WHO 预认证程序将以滚动方式评估疫苗试验和部署工作产生的更多临床数据,以确保疫苗始终符合必要的质量、安全性和有效性标准,从而扩大疫苗供应[6]。目前已有 15 款疫苗被 WHO 列入紧急使用清单,分别来自辉瑞/BioNTech 公司、阿斯利康公司/牛津大学、巴拉特生物技术公司、强生公司、莫德纳公司、北京生物制品研究所、北京科兴中维生物技术有限公司、印度血清研究所、美国诺瓦瓦克斯公司、中国康希诺生物制剂公司、韩国 SK 生物科技公司[7]。见表3。

我国疫苗研发进展始终处于全球第一梯队,目前共计有 5 条技术路线,17 款产品正在进行或已完成Ⅲ期临床试验[8]。3 款疫苗被纳入紧急使用清单,包括 BBIBP-Corv 疫苗(中国生物北京生物制品研究所)、克尔莱福疫苗(北京科兴中维生物技术有限公司)和克威莎疫苗(康希诺生物股份公司)[7]。

3　WHO 推荐对新冠疫苗策略的最新指导意见

目前,奥密克戎变异株已成为全球流行的主要新冠病毒株,尤其是 XBB 亚型[9]。但其粗病死率和死亡率都有所下降,这提示新冠病毒的传播已日趋稳定,感染力强而致病力低[10]。

WHO 于 3 月 20 — 23 日召开会议,修订了 COVID-19疫苗接种的路线图[11]。此次修订既反映出奥密克戎变异株的影响,也考虑到病毒感染和疫苗接种在人群中形成的高水平免疫。修订后的路线图再次强调了为风险人群接种疫苗的重要性。同时,各国应该根据实际情况来决定是否继续为低风险人群接种疫苗,尤其是不能影响常规疫苗接种。此外,路线图还重点关注如何维持卫生系统的复原力,并重新考虑了为低风险人群接种新冠疫苗的成本效益,同时修订了有关续打加强针及其接种间隔的建议。

表 3　WHO 紧急使用清单中的COVID-19疫苗

Table 3　COVID-19 Vaccines with WHO Emergency Use Listing

疫苗名称 Vaccine name	研发机构 Research and development institution
COMIRNATY®COVID-19 mRNA 疫苗(核苷修饰) COMIRNATY®COVID-19 mRNA Vaccine (nucleoside modified)	辉瑞/BioNTech 公司 Pfizer / BioNTech
COMIRNATY®原始株/Omicron BA.1 (15/15 mcg) COVID-19 mRNA 疫苗(核苷修饰) COMIRNATY®Original/Omicron BA.1 (15/15 mcg) COVID-19 mRNA Vaccine (nucleoside modified)	辉瑞/BioNTech 公司 Pfizer / BioNTech
COMIRNATY®原始株/Omicron BA.4-5 (15/15 mcg) COVID-19 mRNA 疫苗(核苷修饰) COMIRNATY®Original/Omicron BA.4-5(15/15 mcg) COVID-19 mRNA Vaccine (nucleoside modified)	辉瑞/BioNTech 公司 Pfizer / BioNTech
COMIRNATY®原始株/Omicron BA.4-5 (5/5mcg) COVID-19 mRNA 疫苗(核苷修饰) COMIRNATY®Original/Omicron BA.4-5 (5/5 mcg) COVID-19 mRNA Vaccine (nucleoside modified)	辉瑞/BioNTech 公司 Pfizer / BioNTech
VAXZEVRIA COVID-19疫苗(ChAdOx1-S[重组]) VAXZEVRIA COVID-19 Vaccine (ChAdOx1-S [recombinant])	阿斯利康公司/牛津大学 Oxford/AstraZeneca
COVISHIELD™COVID-19疫苗(ChAdOx1-S[重组]) COVISHIELD™COVID-19 Vaccine (ChAdOx1-S [recombinant])	印度血清研究所 Serum Institute of India Pvt. Ltd
COVID-19疫苗 (Ad26.COV2-S [重组]) COVID-19 Vaccine (Ad26.COV2-S [recombinant])	强生公司 Janssen-Cilag International NV
SPIKEVAX COVID-19 mRNA 疫苗(核苷修饰) SPIKEVAX COVID-19 mRNA Vaccine (nucleoside modified)	莫德纳公司 Moderna Biotech
COVID-19灭活疫苗(Vero 细胞) Inactivated COVID-19 Vaccine (Vero Cell)	北京生物制品研究所 Beijing Institute of Biological Products Co., Ltd. (BIBP)
CoronaVac COVID-19灭活疫苗(Vero 细胞) CoronaVac COVID-19 Vaccine (Vero Cell), Inactivated	北京科兴中维生物技术有限公司 Sinovac Life Sciences Co., Ltd
COVAXIN®COVID-19疫苗(全病毒颗粒灭活冠状病毒疫苗) COVAXIN®COVID-19 vaccine (Whole Virion Inactivated Corona Virus vaccine)	巴拉特生物技术公司 Bharat Biotech International Ltd
COVOVAX™COVID-19疫苗(SARS-CoV-2 rS 蛋白纳米颗粒[重组]) COVOVAX™COVID-19 vaccine (SARS-CoV-2 rS Protein Nanoparticle [recombinant])	印度血清研究所 Serum Institute of India Pvt. Ltd
NUVAXOVID™COVID-19疫苗(SARS-CoV-2 rs[重组佐剂]) NUVAXOVID™COVID-19 vaccine (SARS-CoV-2 rS [Recombinant, adjuvanted])	美国诺瓦瓦克斯公司 Novavax CZ a.s.
CONVIDECIA COVID-19疫苗(Ad5-nCoV-S [重组]) CONVIDECIA COVID-19 Vaccine (Ad5-nCoV-S [Recombinant])	中国康希诺生物制剂公司 CanSino Biologics Inc.
SKYCovione™(GBP510) COVID-19疫苗(重组蛋白亚基) SKYCovione™(GBP510) COVID-19 Vaccine (Recombinant protein subunit)	韩国 SK 生物科技公司 SK Bioscience Co., Ltd,

表 4　WHO 免疫战略咨询专家组更新的COVID-19疫苗接种指南

Table 4　The Strategic Advisory Group of Experts in WHO updates COVID-19 vaccination guidance

人群 Population	定义 Definition	疫苗接种建议 Vaccination recommendations
高优先使用人群 High priority group	老年人、患有严重合并症(例如糖尿病和心脏病)的年轻人、免疫缺陷症患者(如艾滋病毒携带者和移植受体,包括≥6 个月的儿童)、孕妇和一线卫生工作者 Older adults; younger adults with significant comorbidities (e.g. diabetes and heart disease); people with immunocompromising conditions (e.g. people living with HIV and transplant recipients, including children aged 6 months and older); pregnant persons; and frontline health workers.	在最后一次加强针注射后 6 个月或 12 个月再注射 1 剂加强针,时间范围取决于年龄和免疫力低下以及当前的流行病学状况等因素。 SAGE recommends an additional booster of either 6 or 12 months after the last dose, with the timeframe depending on factors such as age and immunocompromising conditions
中等优先使用人群 Medium priority group	50~<60 岁以下未患有合并症的健康成年人,以及患有合并症的儿童和青少年 Healthy adults- usually under the age of 50-<60 without comorbidities and children and adolescents with comorbidities	建议全程接种新冠疫苗并接种 1 剂加强针。尽管额外的加强针对这一群体是安全的,但鉴于公共卫生回报相对较低,因此并不常规推荐这一做法。 SAGE recommends primary series and first booster doses for the medium priority group. Although additional boosters are safe for this group, SAGE does not routinely recommend them, given the comparatively low public health returns.
低优先使用人群 Low priority group	年龄为 6 个月到 17 岁的健康儿童和青少年 Healthy children and adolescents aged 6 months to 17 years	建议全程接种疫苗以及加强针。但是鉴于其疾病负担较低,因此各国应当基于各国实际情况,如疾病负担、成本效益、其他卫生或方案优先事项以及机会成本,重新考虑是否为该群体接种疫苗。 Primary and booster doses are safe and effective in children and adolescents. However, considering the low burden of disease, SAGE urges countries considering vaccination of this age group to base their decisions on contextual factors, such as the disease burden, cost effectiveness, and other health or programmatic priorities and opportunity costs.

修订后的路线图确定了 3 个等级的新冠疫苗优先使用人群,即:高优先使用人群、中等优先使用人群和低优先使用人群。这些优先级主要基于重症和死亡的风险,并考虑疫苗性能、成本效益、规划因素和社区接受程度[11]。3 个等级的新冠疫苗优先使用人群及疫苗接种建议。见表 4。

此外,已经制定额外加强针政策的国家应该基于国家疾病负担、成本效益以及机会成本来评估不断变化的接种需求。除路线图外,专家组还更新了有关二价新冠疫苗的建议,各国可以考虑使用针对奥密克戎 BA.5 的二价 mRNA 疫苗作为基础免疫疫苗[11]。

WHO COVID-19 疫苗成分技术咨询小组于 2023 年5 月 18 日称,新的新冠疫苗制剂应产生对 XBB.1.5或 XBB.1.16 变种毒株的抗体反应,实现针对 XBB 谱系的中和抗体反应的其他疫苗制剂或疫苗制剂开发平台也可以列入考虑。该小组还指出,目前在人群中很难检测到新冠原始株和早期流行株(例如:Alpha、Beta、Gamma、Delta),接种针对原始毒株疫苗对当前的流行变异株所产生的中和抗体非常低或检测不到,因此未来的疫苗不会再针对新冠原始毒株[12]。

WHO 提供了关于全球疫苗接种状况的最新情况,并考虑了对可能终止国际关注的突发公共卫生事件的影响。WHO 向所有缔约国发布了 7 条临时建议:(1)保持国家能力增长并为未来事件做好准备,避免出现恐慌和忽视的循环;(2)将新冠疫苗接种纳入生命全程疫苗接种计划;(3)汇集来自不同呼吸道病原体监测数据源的信息,以实现全面把控;(4)准备好在国家监管框架内授权的医疗对策,以确保长期供应和可用性;(5)继续与社区及其领导者合作,以实现强大、有弹性和包容性的风险沟通和社区参与,以及信息流行病学管理计划;(6)根据风险评估,继续取消与新冠国际旅行相关的卫生措施,并且不要求任何新冠疫苗接种证明作为国际旅行的先决条件;(7)继续支持研究以改进减少传播和具有广泛适用性的疫苗;了解新冠后疾病的全谱、发病率和影响,以及新冠在免疫功能低下人群中的演变;并制定相关的综合护理途径[1]。

4　全球各国和地区新冠疫苗接种策略调整

美国:截至 2023 年 6 月,美国约有 85.80%的人口接种了新冠疫苗,约 73.50%的人口全程接种了新冠疫苗。美国疾病预防控制中心于 2023 年 6 月 5 日更新了疫苗接种策略,建议≥6 岁的人群都应当接种 1 针最新的辉瑞/BioNTech 疫苗或莫德纳疫苗,>65 岁人群应当接种第 2 剂,中度或重度免疫功能低下的人群应当额外接种最新的辉瑞/BioNTech 疫苗或莫德纳疫苗,<6 岁儿童可能需要多剂最新的 COVID-19 疫苗,至少包括 1 剂最新的辉瑞/BioNTech疫苗或莫德纳疫苗,这取决于他们的年龄和之前疫苗接种情况[13]。2023 年 1 月 26 日,美国食品药品监督管理局建议每年更新一次新冠疫苗以应对流行株的变异,并在春季为疫苗研发选择流行变异株,在秋季进行重点人群疫苗接种。疫苗接种

计划将被最大程度简化,无论是初级接种还是加强针,二价奥密克戎/原始毒株 mRNA 新冠疫苗都将彻底取代原始型疫苗[14]。美国拜登政府于 2023 年 5 月1 日表示,将于下周结束剩余的联邦疫苗要求措施,不再强制要求公众接种新冠疫苗[15]。拜登政府于 2023 年 5 月 9 日签署公告,宣布取消赴美国际航班旅客的新冠疫苗接种要求,同时,美国国土安全部将不再要求通过陆路入境口岸,以及渡轮码头进入美国的非美国旅客完全接种新冠疫苗,不再需要根据要求提供相关的疫苗接种证明,新规将于 5 月12 日生效。这与公共卫生紧急状态的结束时间相一致[16]。

英国:截至 2023 年 6 月,英国约有 86.3%的人口接种了新冠疫苗,约 71.2%的人口全程接种了新冠疫苗。英国国民医疗服务体系(national health service, NHS)建议 5 岁以上人群均可以接种两剂新冠疫苗,<5 岁儿童仅能在由于健康问题或免疫系统功能低下而导致感染风险升高时或与免疫功能低下的人生活在一起时才建议接种两剂新冠疫苗。75 岁以上的老年人、居住在护理院的老年人、>5 岁的免疫功能低下的人群可以接种第 3 针加强针[17]。在 WHO 宣布 COVID-19疫情不再构成PHEIC后,NHS 修订了疫苗接种策略。上述疫苗接种策略将持续至 2023 年 6 月 30 日,在 6 月 30 日后,居民将无法预约接种 COVID-19疫苗,除非感染新冠病毒的风险增加,并且在多数情况下,居民只能等到秋季才可以接种新冠疫苗。在特殊情况下,如果居民出现了新的症状并且开始接受治疗,在临床医生的建议下,患者可以快速接种疫苗[17]。2023 年 1 月 25 日,英国疫苗接种和免疫联合委员会提出,应计划在 2023 年秋季为那些感染新冠后出现重症的高风险人群提供加强剂疫苗接种。同时,建议年龄较大和有免疫抑制的人应计划在春季再接种 1 剂加强疫苗[18]。

日本:2023 年 4 月 29 日起,日本修订入境要求,进入日本的旅客不再需要疫苗接种证明或检测(阴性)证明[19]。2023 年 6 月 26 日,日本专家表示日本全国的新冠感染病例正在小幅上升,第 9 波疫情可能已经开始,需要采取措施减少新冠死亡人数,包括进行第 6 次疫苗接种,尤其是针对老年人。根据政府公布的数据来看,截至 2023 年 6 月,日本约有 81.0%的人口都完成了第 1 针接种,68.7%的人口接种了 3 剂疫苗。日本厚生劳动省 2023 年 6 月 16 日决定引进针对 Omicron 变异株 XBB 亚型的一价疫苗,用于 9 月起实施的新冠疫苗秋季接种[20]。

中国:国务院应对新型冠状病毒感染疫情联防联控机制综合组 2023 年 4 月 10 日通过国家卫生健康委员会网站发布的《应对近期新冠病毒感染疫情疫苗接种工作方案》指出,当前疫苗接种主要是针对两大目标人群,一类是未感染且未完成既定免疫程序的人群;一类是已感染且未完成基础免疫的人群。而对于已感染且未完成基础免疫的人群,可在感染 3 个月后接种 1 剂次疫苗[21]。

2023 年 5 月 18 日,北京市疾病预防控制中心发布周报显示,2023 年第 19 周(5 月 8 日至 5 月 14 日),新冠感染数再度超过流感,已连续 3 周成为法定传染病排名第一的病种[22]。最近感染的新冠病毒毒株都是 XBB1.15 或者 XBB1.9 等变异株,在未来 6 个月内,新的变异株和前者在遗传上、抗原性上很接近,要比去年底的 BA 5.2/BF.7 和 XBB1.15/XBB1.9 的遗传距离更近。因此,最近感染的人体内中和抗体的保护通常会维持 6~10 个月,在此时间间隔内接种疫苗是没有必要的[23]。截至目前,相关部门尚未对国内新冠疫苗接种政策做出调整。我国于 2020 年 12 月 15 日启动的重点人群新冠疫苗接种,各地参照核酸检测“应检尽检”的模式,由各级政府组织和安排费用保障,个人不承担疫苗的成本和接种费用[24]。在疫情结束紧急状态后新冠疫苗是否可能被纳入国家免疫规划,仍需要继续检测新冠病毒的情况,以支持决策者对未来的新冠病毒疫苗接种政策提出明智的建议。

根据香港特别行政区政府卫生署发表的最新新冠疫苗接种服务安排,为了维持足够容量同时确保成本效益,2023 年 4 月 20 日起非高危群组的市民如需接种第 4 针,需参照季节性流感疫苗的安排,到私营市场自费接种,属于高危组别(≥50 岁的居民;18~<50 岁患有长期疾病的成年人;≥6 个月免疫力弱的人士;孕妇;医护人员)的香港市民暂不受影响。不管既往接种多少剂疫苗,于接种上一剂疫苗或感染新冠超过 6 个月后,居民仍可预约免费接种。同时,在接种点安排新冠和流感疫苗同时接种。在香港,老年群体中季节性流感疫苗的接种率已达到 60%左右的水平,将来新冠疫苗也是要走这一路线,即聚焦于老年人和脆弱群体的免疫水平持续性强化[24]。

澳门特别行政区政府卫生局于 2023 年 4 月 5 日调整新冠疫苗接种策略。卫生局参考 WHO 及各地卫生部门的最新建议,结合澳门流行病学特点,将按不同年龄组和健康状况的人群作风险分类,较高风险人群额外再接种 1 剂二价疫苗(即二价 mRNA 新冠疫苗),接种策略调整方向如下:(1)6 个月至 17 岁无高风险情况者:建议接种 3 剂(灭活疫苗或 mRNA 疫苗),不可接种3剂以上;(2)18~<60岁无高风险情

表 5 新冠疫苗的目标人群、时间间隔和疫苗选择

Table 5 Target population, time interval and vaccine selection for COVID-19vaccines

目标人群 Target population	时间间隔 Time interval	疫苗选择 Vaccine selection
未感染 Uninfected		
未完成基础免疫人群 Incomplete basic immunization population	按现行规定执行 In accordance with the current regulations	新冠病毒灭活疫苗(国药中生北京公司、北京科兴公司、国药中生武汉公司、深圳康泰公司、医科院生物所);重组新冠病毒疫苗(智飞龙科马[中国仓鼠卵巢(Chinese Hamster Ovary ,CHO)细胞];腺病毒载体新冠病毒疫苗(康希诺肌注式)。 Inactivated COVID-19 vaccine (Sinopharmate Beijing Company, Sinopharmate Beijing Company, Sinopharmate Wuhan Company, Shenzhen Kangtai Company, Institute of Biology, Academy of Medical Sciences); Recombinant COVID-19 vaccine (Chinese Hamster Ovary (CHO) cells); Adenovirus vector COVID-19 vaccine (Concino intramuscular injection).
18 岁以上未完成第 1 剂次加强免疫接种的人群 People over 18 years of age who have not completed the first dose of booster immunization	完成基础免疫 3 个月后实施第 1 剂次加强免疫 The first dose of booster immunization was administered 3 months after completion of basic immunization	神州细胞重组新冠病毒 4 价 S 三聚体蛋白疫苗、石药集团新冠病毒 mRNA 疫苗;重组新冠病毒疫苗[智飞龙科马(CHO 细胞)、珠海丽珠(CHO 细胞)、成都威斯克(sf9 细胞)、浙江三叶草(CHO 细胞)、神州细胞 2 价 S 三聚体];腺病毒载体新冠病毒疫苗(康希诺吸入用);流感病毒载体新冠病毒疫苗(北京万泰)。 Shenzhou cell recombinant novel coronavirus 4-valent S trimer protein vaccine, Stone Pharmaceutical Group novel coronavirus mRNA vaccine; Recombinant COVID-19 vaccine [Zhifeilong Coma (CHO cells), Zhuhai Lizhu (CHO cells), Chengdu Wesker (sf9 cells), Zhejiang clover (CHO cells), Shenzhou cell 2-valent S trimer]; Adenovirus vector novel coronavirus vaccine (Consinol inhalation); Influenza mildew vector novel coronavirus vaccine (Beijing Wantai).
18 岁以上未完成第 2 剂次加强免疫接种的人群(感染高风险人群、60 岁以上老年人群、具有较严重基础性疾病人群和免疫力低下人群) People over 18 years of age who have not completed the second dose of booster immunization (people at high risk of infection, people over 60 years of age, people with more serious underlying diseases and people with low immunity)	完成第 1 剂次加强免疫 6 个月后实施第 2 剂次加强免疫 The second booster dose was administered 6 months after the first booster dose was completed	
已感染 Infected	未完成基础免疫人群 Incomplete basic immunization population	感染 3 个月后 Three months after infection

况者:建议接种 3 剂,已完成 3 剂或 4 剂但未曾接种二价疫苗者,可额外再接种 1 剂二价疫苗;(3)12~<60 岁具有高风险情况、≥60 岁人士,建议接种 4 剂,已接种 4 剂但未曾接种二价疫苗者,可额外接种 1 剂二价疫苗;(4)孕妇:怀孕期间,最后 1 剂接种已超过 6 个月者,应额外接种 1 剂二价疫苗;(5)前线医务人员:除按其所属年龄组和健康状况的建议接种外,已完成相应接种剂数但未曾接种二价疫苗者,可在上一剂接种 6 个月后额外接种 1 剂二价疫苗[25]。

5 启示与建议

继续做好重点人群疫苗接种工作。在做好疫苗安全和质量评估的前提下,继续为老年人等重点人群提供定期、免费的加强针接种服务。

加大对疫苗有效性的宣传力度。及时公布疫苗有效性及新型疫苗的研发进展,减少公众疫苗犹豫。

加强变异株监测,加快针对变异株的新型疫苗研发。在疫苗审评审批方面与国际先进同行接轨,鼓励和支持创新。同时,应继续加强新冠变异株监测,评估变异株流行态势变化并开展对疫苗有效性影响的及时动态研判。

结合我国国情合理规划疫苗接种策略。受 PHEIC 结束的影响,全球新冠疫苗的需求和投资热情均会在短期内出现明显回落,在过渡期内,WHO 建议卫生管理部门应当发挥资源的协调作用,通过合理进行疫苗免疫规划,保障各国对疫苗的需求。

利益冲突 无

参 考 文 献

[1] WHO. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic [EB/OL]. (2023-05-05) [2023-06-07]. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-

coronavirus-disease-(covid-19)-pandemic.

[2] WHO. WHO Coronavirus (COVID-19) Dashboard [EB/OL]. (2023-06-01) [2023-06-02]. https://covid19.who.int/.

[3] Our world in data. Coronavirus (COVID-19) Vaccinations [EB/OL]. (2023-06-01) [2023-06-02]. https://ourworldindata.org/covid-vaccinations.

[4] WHO. COVID-19 vaccine tracker and landscape [EB/OL]. (2023-03-30) [2023-06-02]. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines.

[5] 张旋旋,佘广彪,刘晓雅,等. 新冠病毒蛋白亚单位疫苗研究进展[J]. 中国医药导刊, 2022,24(5):434-438.
Zhang XX, She GB, Liu XY, et al. Research progress of SARS-CoV-2 protein subunit vaccine [J]. Chin J Med Guide, 2022, 24(5):434-438.

[6] WHO. WHO lists two additional COVID-19 vaccines for Emergency use and COVAX roll-out [EB/OL]. (2021-02-15) [2023-06-02]. https://www.who.int/news/item/15-02-2021-who-lists-two-additional-covid-19-vaccines-for-emergency-use-and-covax-roll-out.

[7] WHO. COVID-19 Vaccines with WHO Emergency Use Listing [EB/OL]. (2022-06-16) [2023-07-03]. https://extranet.who.int/pqweb/vaccines/vaccinescovid-19-vaccine-eul-issued.

[8] 预防界. 应对变异、广谱多价,专家共话新冠病毒疫苗研发新进展[EB/OL]. (2023-04-19) [2023-06-02]. https://new.qq.com/rain/a/20230419A012SF00.
The prevention community. In response to variation, broad spectrum and multi-price, experts shared new progress in the development of COVID-19 vaccine [EB/OL]. (2023-04-19) [2023-06-02]. https://new.qq.com/rain/a/20230419A012SF00.

[9] 屠博文,王楠,董泽丰,等. 新冠病毒 Omicron XBB.1.5 和 XBB.1.9 及其亚分支变异株流行情况及相关认知进展[J]. 江苏预防医学, 2023,34(02):119-122. DOI:10.13668/j.issn.1006-9070.2023.02.001.
Tu BW, Wang N, Dong ZF, et al. Prevalence of Omicron XBB.1.5 and XBB.1.9 and their subbranch variants and related cognitive progress of novel coronavirus [J]. Jiangsu preventive medicine, 2023,34(02):119-122. DOI:10.13668/j.issn.1006-9070.2023.02.001.

[10] Karyakarte RP, Das R, Dudhate S, et al. Clinical characteristics and outcomes of laboratory-confirmed SARS-CoV-2 cases infected with omicron subvariants and the XBB recombinant variant [J]. Cureus, 2023,15(2):e35261. DOI:10.7759/cureus.35261.

[11] WHO. SAGE updates COVID-19 vaccination guidance [EB/OL]. (2023-04-19) [2023-06-02]. https://www.who.int/news/item/28-03-2023-sage-updates-covid-19-vaccination-guidance.

[12] WHO. Statement on the antigen composition of COVID-19 vaccines [EB/OL]. (2023-05-18) [2023-06-02]. https://www.who.int/news/item/18-05-2023-statement-on-the-antigen-composition-of-covid-19-vaccines.

[13] WHO. Stay Up to Date with COVID-19 Vaccines [EB/OL]. (2023-06-01) [2023-06-02]. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html.

[14] ABC News. FDA proposes most Americans receive one annual COVID vaccine, similar to flu [EB/OL]. (2023-01-26) [2023-06-02]. https://abcnews.go.com/Health/fda-proposes-americans-receive-annual-covid-vaccine-similar/story?id=96617332.

[15] White House. The Biden-Harris Administration Will End COVID-19 Vaccination Requirements for Federal Employees, Contractors, International Travelers, Head Start Educators, and CMS-Certified Facilities [EB/OL]. (2023-05-01) [2023-06-02]. https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/01/the-biden-administration-will-end-covid-19-vaccination-requirements-for-federal-employees-contractors-international-travelers-head-start-educators-and-cms-certified-facilities/.

[16] White House. A Proclamation on Revoking the Air Travel COVID-19 Vaccination Requirement [EB/OL]. (2023-05-09) [2023-06-02]. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/05/09/a-proclamation-on-revoking-the-air-travel-covid-19-vaccination-requirement/.

[17] NHS. About COVID-19 vaccination [EB/OL]. (2023-03-21) [2023-06-02]. https://www.nhs.uk/conditions/covid-19/covid-19-vaccination/about-covid-19-vaccination/.

[18] UK Health Security Agency. JCVI advises an autumn COVID-19 vaccine booster [EB/OL]. (2023-01-25) [2023-06-02]. https://www.gov.uk/government/news/jcvi-advises-an-autumn-covid-19-vaccine-booster.

[19] 日本驻华大使馆. 新冠病毒疫情(前往日本所需手续)(4 月 28 日更新)[EB/OL]. (2023-04-28) [2023-06-02]. https://www.cn.emb-japan.go.jp/itpr_zh/00_000485_00186.html.
Japanese Embassy in China. COVID-19 (Procedures Required for Travel to Japan) (Updated on April 28) [EB/OL]. (2023-04-28) [2023-06-02]. https://www.cn.emb-japan.go.jp/itpr_zh/00_000485_00186.html.

[20] 界面新闻. 日本将引进针对 XBB 变异株的新冠疫苗,用于 9 月起秋季接种[EB/OL]. (2023-06-17) [2023-07-02]. https://www.jiemian.com/article/9595613.html.
Interface news. Japan will introduce a novel coronavirus vaccine against the XBB variant for autumn vaccination starting in September [EB/OL]. (2023-06-17) [2023-07-02]. https://www.jiemian.com/article/9595613.html.

[21] 国家卫生健康委员会. 关于印发应对近期新冠病毒感染疫情疫苗接种工作方案的通知[EB/OL]. (2023-04-10) [2023-06-02]. https://www.gov.cn/lianbo/2023-04/10/content_5750733.htm.
National Health Commission. Notice on the issuance of the work plan for vaccination in response to the recent outbreak of novel coronavirus infection [EB/OL]. (2023-04-10) [2023-06-02]. https://www.gov.cn/lianbo/2023-04/10/content_5750733.htm.

[22] 北京市疾病预防控制中心. 2023 年第 19 周疫情周报[EB/OL]. (2023-05-18) [2023-06-02]. https://www.bjcdc.org/cdcmodule/jkdt/yqbb/2023/82468.shtml.
Beijing Center for Disease Control and Prevention. Weekly Epidemic Report of the 19th week of 2023 [EB/OL]. (2023-05-18) [2023-06-02]. https://www.bjcdc.org/cdcmodule/jkdt/yqbb/2023/82468.shtml.

[23] 澎湃新闻. 中疾控:我国新冠病毒主要流行株已变为 XBB 系

列变异株，致病力无明显变化［EB/OL］.（2023-05-08）［2023-06-02］. https://www.thepaper.cn/newsDetail_forward_22998689.
Pengpai News. Chinese Center for Disease Control and Prevention: The main circulating strain of the novel coronavirus in China has changed into the XBB series variant strain, and the pathogenicity has not changed significantly［EB/OL］.（2023-05-08）［2023-06-02］. https://www.thepaper.cn/newsDetail_forward_22998689.

［24］第一财经. 继续免费还是部分自费？专家建议优化中国新冠疫苗政策［EB/OL］.（2023-05-08）［2023-06-02］. https://finance.eastmoney.com/a/202305082714859920.html.
China Business News. Remain free or partially self-financed? Experts suggest improving China's COVID-19 vaccine policy［EB/OL］.（2023-05-08）［2023-06-02］. https://finance.eastmoney.com/a/202305082714859920.html.

［25］澳门特别行政区政府新闻局. 参考世卫及各地卫生部门最新建议 本澳将调整新冠疫苗接种策略［EB/OL］.（2023-04-05）［2023-06-02］. https://www.gcs.gov.mo/detail/zh-hans/N23DE2adiv.
Information Bureau of the Government of the Macao Special Administrative Region. In light of the latest advice from WHO and local health authorities, Australia will adjust its COVID-19 vaccination strategy［EB/OL］.（2023-04-05）［2023-06-02］. https://www.gcs.gov.mo/detail/zh-hans/N23DE2adiv.

（收稿日期：2023-06-08）
（修回日期：2023-07-10）
本文编辑：江玲琼（中文）
李宝珠（英文）

（上接第 926 页）

［8］欧阳平，李小溪，冷芬，等. 机器学习算法在体检人群糖尿病风险预测中的应用［J］. 中华疾病控制杂志，2021，25（7）：849-853，868. DOI：10.16462/j.cnki.zhjbkz.2021.07.020.
Ouyang P, Li XX, Leng F, et al. Application of machine learning algorithm in diabetes risk prediction of physical examination population［J］. Chin J Dis Control Prev, 2021, 25(7): 849-853, 868. DOI: 10.16462/j.cnki.zhjbkz.2021.07.020.

［9］朱丽，张蓉，张淑莲，等. 中国不同胎龄新生儿出生体重曲线研制［J］. 中华儿科杂志，2015，53（2）：97-103. DOI：10.3760/cma.j.issn.0578-1310.2015.02.007.
Zhu L, Zhang R, Zhang SL, et al. Chinese neonatal birth weight curve for different gestational age［J］. Chin J Pediatr, 2015, 53(2): 97-103. DOI: 10.3760/cma.j.issn.0578-1310.2015.02.007.

［10］中国肥胖问题工作组数据汇总分析协作组. 我国成人体重指数和腰围对相关疾病危险因素异常的预测价值：适宜体重指数和腰围切点的研究［J］. 中华流行病学杂志，2002，23（1）：5-10. DOI：10.3760/j.issn：0254-6450.2002.01.003.
Coorperative Meta-analysis Group of China Obesity Task Force. Predictive values of body mass index and waist circumference to risk factors of related diseases in Chinese adult population［J］. Chin J Epidemiol, 2002, 23(1): 5-10. DOI: 10.3760/j.issn:0254-6450.2002.01.003.

［11］Li Y, Guo H, Xiao L, et al. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data［J］. Knowledge-Based Systems, 2016, 94: 88-104. DOI: 10.1016/j.knosys.2015.11.013.

［12］Royal College of Obstetricians and Gynaecologists. The investigation and management of the small for gestational age fetus. Green-top Guideline No.31［EB/OL］.（2013-03-22）［2023-02-16］. https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg31/.

［13］Papastefanou I, Wright D, Lolos M, et al. Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics, serum pregnancy-associated plasma protein-A and placental growth factor at 11-13 weeks' gestation［J］. Ultrasound Obstet Gynecol, 2021, 57(3): 392-400. DOI: 10.1002/uog.23118.

［14］Gürgen F, Zengin Z, Varol F. Intrauterine growth restriction (IUGR) risk decision based on support vector machines［J］. Expert Syst Appl, 2012, 39(3): 2872-2876. DOI: 10.1016/j.eswa.2011.08.147.

［15］Gardosi J, Madurasinghe V, Williams M, et al. Maternal and fetal risk factors for stillbirth: population based study［J］. BMJ, 2013, 346: f108. DOI: 10.1136/bmj.f108.

［16］Gurung S, Tong HH, Bryce E, et al. A systematic review on estimating population attributable fraction for risk factors for small-for-gestational-age births in 81 low-and middle-income countries［J］. J Glob Health, 2022, 12: 04024. DOI: 10.7189/jogh.12.04024.

［17］Parihar S, Singh S. Perinatal outcomes and intrahepatic cholestasis of pregnancy: a prospective study［J］. Int J Reprod Contracept Obstet Gynecol, 2019, 8(3): 1177-1182. DOI: 10.18203/2320-1770.ijrcog20190901.

［18］Natarajan V, Singh P, Vigneshwar NKV, et al. Maternal and placental risk factors for small gestational age and fetal malnutrition［J］. Curr Pediatr Rev, 2023, 19(2): 187-196. DOI: 10.2174/1573396318666220705154424.

［19］Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review［J］. J Biomed Inform, 2002, 35(5-6): 352-359. DOI: 10.1016/S1532-0464(03)00034-0.

［20］Vapnik VN, Kotz S. Estimation of dependences based on empirical data［M］. New York: Springer Science & Business Media, 2006: 232-457.

（收稿日期：2022-12-16）
（修回日期：2023-01-26）
本文编辑：王菲菲（中文）
李宝珠（英文）

·综述·

临床预测模型中处理时依性变量的策略及进展

于玥琳[1]　胥洋[2]　王俊峰[3]　詹思延[1]　王胜锋[1]

[1]北京大学公共卫生学院流行病与卫生统计学系/重大疾病流行病学教育部重点实验室，北京 100191；[2]北京大学临床研究所真实世界证据评价中心，北京 100191；[3]荷兰乌特勒支大学医学中心 Julius 健康科学与初级保健中心，乌得勒支 3508 TC

通信作者：王胜锋，Email：wangshengfeng@bjmu.edu.cn

【摘要】 预测模型中考虑时依性变量可改善模型的总体表现，提高其临床应用价值。界标模型、联合模型等基于传统回归策略在处理时依性变量个数和适用情境等方面存在局限，神经网络等机器学习算法有望对其灵活处理。本文针对传统模型、机器学习算法，总结各自纳入时依性变量的建模思路，梳理各方法的适用场景，概括现有方法仍存在的问题，以期为未来预测建模处理时依性变量提供方法学启示。

【关键词】 临床预测模型；　时依性变量；　动态预测；　机器学习

基金项目：国家自然科学基金（82173616）

Methodology and progress in adjusting time-dependent covariates in clinical prediction models

Yu Yuelin[1], Xu Yang[2], Wang Junfeng[3], Zhan Siyan[1], Wang Shengfeng[1]

[1]Key Laboratory of Epidemiology of Major Diseases, Ministry of Education/Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; [2]Center for Real-world Evidence Evaluation, Peking University Clinical Research Institute, Beijing 100191, China; [3]Julius Center for Health Sciences and Primary Care, University of Utrecht, Utrecht 3508 TC, Netherlands

Corresponding author: Wang Shengfeng, Email: wangshengfeng@bjmu.edu.cn

【Abstract】 Adjusting time-dependent covariates into prediction models may help improve model performance and expand clinical applications. The methodology of handling time-dependent covariates is limited in traditional regression strategies (i. e., landmark model, joint model). For example, the number of predictors and practical situations which can be handled are restricted when using regression models. One new strategy is to use machine learning (i.e., neural networks). This review summarizes the methodology of handling time-dependent covariates in prediction models, such as applicable scenarios, strengths, and limitations, to offer methodological enlightenment for processing time-dependent covariates.

【Key words】 Clinical prediction model;　Time-dependent covariate;　Dynamic prediction;　Machine learning

Fund program:　National Natural Science Foundation of China (82173616)

电子医疗数据为临床预测模型开发提供丰富的患者个体层面纵向时序数据。但现有模型鲜有纳入时依性变量[1]，大多只利用基线信息而未考虑变量时依性变化对结局的影响，无法实现动态预测。相关方法学瓶颈在于如何处理时依性变量。变量的时依性通常有两种含义：时依性变量和时依性系数。前者的取值或状态随生存时间推移而

DOI：10.3760/cma.j.cn112338-20230128-00042

收稿日期　2023-01-28　　**本文编辑**　万玉立

引用格式：于玥琳，胥洋，王俊峰，等. 临床预测模型中处理时依性变量的策略及进展[J]. 中华流行病学杂志，2023，44(8)：1316-1320. DOI：10.3760/cma.j.cn112338-20230128-00042.

Yu YL, Xu Y, Wang JF, et al. Methodology and progress in adjusting time-dependent covariates in clinical prediction models [J]. Chin J Epidemiol, 2023, 44(8):1316-1320. DOI: 10.3760/cma.j.cn112338-20230128-00042.

发生改变；后者指在变量本身取值未变的情况下，变量对结局的作用随时间发生改变。电子医疗数据随时间不断更新，属于时依性变量范畴，故本文聚焦其处理策略。时依性变量又区分为内生和外生。内生时依性变量本身取值随时间变化，同时变量取值也受目标预测结局是否发生的影响[2]。例如，患者不同时点的生化指标值与目标疾病发生进展、病程迁延相互影响。如患者服药剂量或状态随时间的测量值会受患者是否发生药品不良反应事件的影响。相反，外生时依性变量对结局的作用效应不会受到前一时刻是否发生了结局事件的影响。如空气污染可作为哮喘发病的一个预测变量，但该变量不受患者是否哮喘发作（结局）的影响而改变其本身取值。本文按变量特征系统梳理现有时依性变量的处理策略，总结不同策略的适用场景及优缺点，以期为预测建模处理时依性变量提供方法学启示。

1. 基于传统回归模型拓展的时依性变量处理策略：

（1）基于单个传统回归模型拓展：基于基线信息开发的预测模型，实现预测值随预测变量在随访过程中的取值变化同步更新，最直接和朴素的方法是将变量更新后的取值代入原始“静态模型”，可达到更新结局风险的“伪动态预测”目的。但此方法在建模过程中仅使用了基线信息，忽略了随访过程中预测变量的变化，无法准确评估变量对结局的影响，未充分利用随访中获得的信息对模型的准确性进行改善。

为了实现在建模过程中即纳入时间依赖变量，可以用分段（将随访时间进行分段）模型法对单个传统回归模型进行拓展（表1）。建模中只涉及对一个传统（Cox或logistic）回归模型的处理。典型模型包括时依性Cox比例风险回归模型（TDCM）[3]。如研究可将患者随时间变化的多时点化验指标测量值（血肌酐、血压等），按照测量时点分段展示时依性变量，对各测量时段利用对应收集到的指标测量值和结局情况，进而在随访时段切割状态下建立TDCM。TDCM虽然不要求整个随访时间内都满足比例风险恒定假设，但仍要求各分段内部满足该统计学假设[2,4-6]；并不对内生或外生时依性变量加以区别，相当于默认变量均为外生，忽略了内生时依性变量对模型估计的潜在影响；可同时处理的时依性变量个数受限[5,7]（大多纳入变量<10个）。上述局限可能致预测模型表现欠佳。

（2）综合多个传统回归模型拓展：单个传统回归模型受限于较严格的统计假设，处理时依性变量与结局之间关联的能力有限。为此，综合多个传统回归模型的建模思路旨在针对性加强模型构建和参数表达的灵活性。此时，将因分段处理被割裂的随访时间整体化是综合策略的方法学核心，其中关键步骤是通过连接函数、变量函数化等统计学“综合”手段将多个原始模型在数学表达上得到整合，最终形成一个嵌套函数表达的预测模型。常用的拓展模型包括联合模型、界标模型等（表1）。其中，联合模型适用于动态预测情境[10]。联合模型同时拟合纵向子模型和时间-事件模型，或同时拟合纵向子模型和处理分类结局的模型（如logistic回归模型）。具体操作上，常用纵向子模型如采用混合效应模型，思路为引入时间变量函数化表达的时依性变量，实现拟合时依性变量的动态变化，将函数化的时依性变量纳入时间-事件模型中，属于对多个不同种模型的综合。界标模型则是在分段模型法基础上，将数据集按事先拟定的界标时点及预测时段划分，在不同时段模拟变量随时间的动态变化，再利用光滑函数等连接函数实现对多时段的多个回归模型的综合，汇总各界标时点模型，属于对多个同

表1 基于传统回归模型拓展处理时依性变量的建模策略[2,4,8-9]

模型	适用情境	优势	局限
基于单个传统回归的模型拓展			
分段模型法	处理短期和长期效应明显不一致的时依性变量	二分类和生存结局资料均可处理；更适用于动态队列，对随访阶段新入组个体，可避免因使用基线信息造成的结局风险低估（即生存风险高估）	要求各分段时间内，有足够样本量（每个自变量至少有10个结局事件），且满足比例风险假设；假设所处理时依性变量均为外生，忽略了内生时依性变量对模型估计产生的潜在影响
综合多个传统回归的模型拓展			
多阶段模型	测量间隔不规则情况下估算缺失数据	二分类和生存结局资料均可处理；可处理时间间隔不规则的重复测量资料；计算效率高；两个阶段可分别采用不同的统计学模型	无法解释失访偏倚；需满足比例风险恒定假设
联合模型	纵向多次测量变量和生存结局（或分类结局），如用于反映疾病进展或治疗疗效多次重复测量的生物标志物水平	主要处理连续型时依性变量（如生物标志物或血压等随时间变化）；可根据临床背景需要选用纵向子模型，可应用的数据类型广泛；考虑了时依性变量内生或外生特征，从而避免模型错误假设	模型形式复杂；若时依性协变量数量较多，参数估计困难
界标模型	时依性协变量个数较多；时依性协变量随时间分布情况复杂	二分类和生存结局资料均可处理；每一界标时点内，仅利用截止时点之前的变量信息，不透支截止时点之后的变量信息；适用于动态队列；模型简单（相比联合模型），易进行参数估计；可提供结局风险随界标时点的动态变化趋势，模型性能（校准度和区分度）较优	易忽略协变量随时间的变化；要求有较长的随访时间；要求各界标时点的协变量和结局信息无缺失，现有缺失填补方法易造成偏倚；界标时点的选择和判断无统一标准，通常为研究者自拟

类模型的综合。截至2022年底，PubMed检索联合模型和界标模型相关案例（含建立预测模型和利用模型进行统计分析）已分别达134和354例。

综合多个传统回归模型的时依性变量处理策略较单个传统回归可显著优化建模灵活性和模型表现，但模型本质仍为传统回归模型（使用连接函数、变量函数化等统计学方法“综合”多个传统回归），仍面临以下挑战：①各时段内部仍需满足比例风险恒定假设；②变量时依性关系复杂或时依性变量个数增加时，变量函数化或连接函数将难以表达。因此，连接函数等整合框架的选用可能克服如上局限，如考虑可反映效应改变速率的结构框架，或考虑将基于频率框架的传统模型整合贝叶斯概率框架，将具有先验信息的固定效应模型与生存子模型进行整合等，有待深入探索[11-13]。

2. 基于机器学习的时依性变量处理策略：相较传统模型，机器学习更适用于处理高维数据，尤其当各预测变量的关系尚未完全明确时，符合医学领域预测中未知关系相当普遍的场景[14]（表2）。此外，相较前两者基于传统回归的处理策略，机器学习可处理的在时间维度上测量时点更为密集的时依性变量。

（1）基于贝叶斯网络的机器学习处理方法：贝叶斯网络应用于时依性变量的建模策略主要包括动态贝叶斯网络（DBN）、连续时间贝叶斯网络等。突出特点是能够处理临床不确定情境，引入参数的不确定性避免过拟合，进一步提升模型的泛化能力。建模思路可归纳为网络结构训练和拓展时间维度进行概率分布推断。网络结构训练用于构建预测变量与目标结局的概率依赖关系，拓展时间维度用于拟合多个相邻时间片段前后时依性变量的概率依赖关系。如以时依性生化检查指标为预测变量（如癌症筛查项目中历次筛查所得的细胞学检查值、组织病理学检查结果、病毒检测结果），以发生癌前病变或发病为预测结局[24]，在DBN中即以各次筛查时点为不同的时间片段，各时间片段内部都建立生化指标与癌症发病的条件概率关联，相邻时间片段间构建时依性生化指标之间的概率关联（前一次生化指标测量值与后一次测量值之间的概率分布和关联强度）。其他DBN实例如Dagum等[25]应用DBN处理多时点心率监测记录预测患者睡眠呼吸暂停急性发作；Orphanou等[26]基于低密度脂蛋白等时依性生化指标和吸烟等基线信息，整合DBN与时间抽象强化学习构建中年男性首次冠心病发作的

表2 基于机器学习处理时依性变量的算法介绍及特点比较

模型	适用情境	优势	局限
基于贝叶斯概率框架的机器学习			
动态贝叶斯网络[15-18]	通常针对离散型时间序列资料；高危人群识别；尤其适用于筛检人群的动态预测；不完全数据集（含缺失）	不受限于Cox比例风险回归模型假设；根据变量间依赖关系易处理数据部分缺失问题；引入参数不确定性，避免模型过拟合；可得出定量的风险预测结果；可将横断面现有信息与历史信息结合，变量取值和变量随时间的发展趋势共同作为预测变量，提高模型预测准确性	数据缺失对模型参数估计准确性影响较大；模型中所考虑的时间序列信息为单一时间变量颗粒度信息
连续时间贝叶斯网络[15,19-20]	时间信息为连续型；多个变量间呈时间依赖性，尤其描述多种时间测量尺度、长时间跨度、细时间信息颗粒度、临床时依性信息较复杂场景	类似动态贝叶斯网络	时间序列长度（超参数τ）影响模型训练效果；时间片段间隔越大（颗粒度越大），模型偏倚越大；超参数优化的计算复杂，变量描述信息假定为统一范式
层级贝叶斯模型[21]	多个预测变量和多个结局；常应用于自身前后对照设计研究中短暂暴露与多结局的预测	避免参数过多造成过拟合；考虑了研究对象组成的层次结构；可探索变量间潜在关联	概率分解方法复杂
贝叶斯非参数	实时监测，相邻记录间存在时间依赖性	无须事先指定参数的数目，加强模型结构对数据的自适应能力	复杂，模型解释性和预测准确性稍低
基于神经网络的深度学习			
递归神经网络（RNN）	带时间戳的数据集；不平衡（非均衡）数据集	可记忆（建模时再利用前一阶段的输出）历史信息，并将历史输出作为下一步输入对后续输出做贡献；模型预测性能随纳入的时依性变量数据量增加而逐步提高，模型不仅从不断更新的时间变量实际值中获得学习，还从更新数据的时间趋势中学习	所需参数数量较多；每次更新都会抵消前次记忆，上一次记忆只能对下一次起作用；训练结果难以重复，因为权重赋值相同存在梯度过小/消失问题
卷积神经网络[22]	医疗文本记录信息	时间序列的特征提取是连续整体性的；预测阶段自动加权	模型偏差较大；无法处理图像旋转缩放等问题
自注意力机制	逐渐取代RNN，加入某些参数即可实现RNN功能，且运算速度快于RNN	可实现由最后一时点的输出回溯考虑最先时点的输入；各时点的输出无须考虑时序性（运算可不按照时间先后顺序），可通过一组输入同时实现平行化处理所有输出	–
贝叶斯神经网络[23]	样本集较小或不平衡（非均衡）数据集	克服传统神经网络无法过拟合缺陷；在参数上引入不确定性，增加结果置信度，引入人工干预；更好地处理不确定性，提升模型泛化能力	复杂，具体应用较少

预测模型等。除此外，DBN还可拓展为贝叶斯非参数，该算法进一步加强模型结构对数据的自适应能力。有研究表明加入时间序列因素的贝叶斯网络后验分类误差<0.04，模型准确性>96%[27]。

（2）基于神经网络的深度学习处理方法：当考虑的模型结构更为复杂（如网络层级、参数设置、待选择模型种类等过多），训练数据量（可获取电子医疗数据库中心数量、数据年限范围等）逐步扩大时，相比前述的DBN预测模型，深度学习算法应用“记忆”功能（网络会对前一时点信息进行记忆，并同后一时点的观测值共同输入计算中）处理时依性变量。即将前一时点的预测变量观测值作为输入得到该时点的预测结局信息，则在后一时点输入时，后一时点的预测变量观测值和前一时点的预测结局值将共同作为后一时点的输入信息，经网络内部超参数设定进行迭代学习后给出后一时点的预测结局。循环往复利用每一新时点的时依性变量信息实现最终目标时点的结局预测。目前较多研究利用递归神经网络对电子病历数据库和患者诊疗过程中的生命体征实时监测记录作时依性变量预测（如使用患者透析治疗或重症监护室中电子监测设备采集数据）[28-29]。

（3）贝叶斯神经网络：是基于贝叶斯概率框架利用神经网络/深度学习思路处理时依性变量。相当于在传统深度学习思路基础上，引入参数不确定性，将传统神经网络得到预测结局概率大小拓展为得到预测结局概率的“置信度”，克服了传统神经网络无法对不确定性建模而造成的过拟合缺陷。如Biazzo等[30]利用贝叶斯神经网络估计医院感染等中小型真实案例场景下的传染病传播模型。Fernandes等[31]将贝叶斯框架与卷积神经网络结合构建细菌性或病毒性肺炎的诊断模型。该方法结合前述贝叶斯网络和神经网络的两者优势，目前主要应用于组学或影像相关领域以处理高维数据[32-33]。

3. 不同时依性变量处理方法的选用依据：参考国际公认的预测模型研究中的偏倚风险评估与应用工具（PROBAST）建议[34-35]，结合数据层面（参考表1、2的模型适用情境）、模型层面（参考表1、2的模型主要优缺点）和临床实际应用3方面综合选用模型[4]。数据层面包括数据类型、变量个数、变量特点等。如当拟合场景逐渐复杂，所考虑变量个数较多，当结局变量与预测变量之间的耦合度逐渐降低时，或当多个症状发生所经历的时间跨度较短时，应用DBN相较连续时间贝叶斯网络是更优选择[14]。反之，若可获取的时点信息较少，仅有个别频次的随访时点分段且各时段内可测量的变量数据资源有限时，使用基于传统模型拓展的时依性变量处理策略在模型可解释性较机器学习会更清晰易懂，且两者表现相当。模型层面包括算法内在数学前提假设及模型表现、预测准确度等，避免模型错误设定。

需注意，精确预测虽然是预测模型的一大目标，但实际应用中通常选择“实用”的算法，即综合计算速度、算法实现的难易程度和模型可解释性。实用性偏向也是目前大多研究中的一大趋势[36]，研究者通常结合自身的临床经验选择更易进行解释的模型，如当复杂设计模型在预测性能上仅“稍”优于简单模型时，研究者常选用传统回归而非“黑箱”的机器学习。

4. 时依性变量处理的现存挑战：利用时依性变量的模型构建目前主要面临建模方法学、数据资源和模型质量评价方面的挑战。

首先，模型选用应考虑时依性变量特性。传统TDCM和现有新型机器学习算法在进行时依性变量处理时，前者无法区分、后者尚未明确区分时依性变量内生或外生变量特征的建模处理。大多数研究中所指的纵向资料聚焦于结局变量的重复测量，而对于变量重复测量关注较少[7]。电子医疗数据可提供大样本建模需求，但目前实际应用存在数据库结构和数据变量层面的方法学困境。多源异构数据库数据结构不一、数据类型多样，处理难度大，且可提供的变量侧重点各异，数据时点、指标定义、定量单位等均不统一。众多电子医疗数据资源仍有待开发，如我国药品不良反应数据库尚未被充分利用。此外，不同算法的效果优劣比较缺少实证支持，多数基于模拟数据的方法比较，由于数据生成过程偏向于特定算法，导致结论可靠性存疑。

模型质量评价方面，PROBAST已提供预测因素的定义、评估、测量时间的偏倚风险评估等指导（如评估“在模型使用的时点是否能够得到所有预测因素的信息”），但现有评估条目没有专门针对时依性变量的细节指标，这也是造成目前大多预测建模研究未充分考虑数据特征而直接根据可实现程度进行模型选择的不恰当操作原因之一。有待修订强调减少模型开发和实际应用断层[37]。

5. 总结：已有电子医疗数据和贴合真实临床情境的建模需求为时依性变量利用提供充足支持，利用纵向医疗信息进行动态预测也是对前者的充分应用。利用时依性变量则是动态预测建模的方法学重点工作，选择恰当的时依性变量处理策略则是具体实施的关键。目前策略主要为界标模型等代表的回归拓展和以贝叶斯网络和神经网络为代表的机器学习算法。实际处理时应结合可利用数据（如数据维度、变量类型等）、算法适用情境（如所需处理的数据颗粒度、需预测时长等）和临床需求（模型使用人群特殊性、可接受风险阈值等）3个方面权衡，并完善加强已有指南对于时依性变量的应用指导。

利益冲突　所有作者声明无利益冲突

参 考 文 献

[1] Goldstein BA, Navar AM, Pencina MJ, et al. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review[J]. J Am Med Inform Assoc, 2017, 24(1): 198-208. DOI: 10.1093/jamia/ocw042.

[2] 肖媛媛，许传志，赵耐青. 常用生存分析模型及其对时依性协变量效应的估计方法[J]. 中国卫生统计，2016, 33(3): 543-547, 552.
Xiao YY, Xu CZ, Zhao NQ. Common survival analysis models and their estimation methods for time-dependent covariate effects [J] Chin Health Statistics, 2016, 33 (3): 543-547, 552.

[3] 梁际洲, 郭晓晶, 许金芳, 等. 药物流行病学研究中的时依性变量处理方法简介及比较[J]. 药物流行病学杂志, 2022, 31(3): 190-197. DOI: 10.19960/j. cnki. issn1005-0698. 2022.03.009.
Liang JZ, Guo XJ, Xu JF, et al. Introduction and comparison of time-dependent variable processing methods in pharmacoepidemiology studies[J]. Chin J Pharmacoepidemiol, 2022, 31(3): 190-197. DOI: 10.19960/j.cnki.issn1005-0698.2022.03.009.
[4] Bull LM, Lunt M, Martin GP, et al. Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods[J]. Diagn Progn Res, 2020, 4:9. DOI:10.1186/s41512-020-00078-z.
[5] Sweeting MJ, Barrett JK, Thompson SG, et al. The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study[J]. Statist Med, 2017, 36(28): 4514-4528. DOI: 10.1002/sim.7144.
[6] Maziarz M, Heagerty P, Cai TX, et al. On longitudinal prediction with time-to-event outcome: comparison of modeling options[J]. Biometrics, 2017, 73(1): 83-93. DOI: 10.1111/biom.12562.
[7] Welten M, de Kroon MLA, Renders CM, et al. Repeatedly measured predictors: a comparison of methods for prediction modeling[J]. Diagn Progn Res, 2018, 2: 5. DOI: 10.1186/s41512-018-0024-7.
[8] 周江杰, 王胜锋. 界标模型介绍及在动态预测中的应用[J]. 中华流行病学杂志, 2022, 43(1): 112-117. DOI: 10.3760/cma.j.cn112338-20210122-00051.
Zhou JJ, Wang SF. Introduction of landmarking approach and its application in dynamic prediction[J]. Chin J Epidemiol, 2022, 43(1): 112-117. DOI: 10.3760/cma. j.cn112338-20210122-00051.
[9] Jenkins DA, Sperrin M, Martin GP, et al. Dynamic models to predict health outcomes: current status and methodological challenges[J]. Diagn Progn Res, 2018, 2: 23. DOI:10.1186/s41512-018-0045-2.
[10] Ibrahim JG, Chu HT, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data [J]. J Clin Oncol, 2010, 28(16): 2796-2801. DOI: 10.1200/JCO.2009.25.0654.
[11] Leiva-Yamaguchi V, Alvares D. A two-stage approach for Bayesian joint models of longitudinal and survival data: correcting bias with informative prior[J]. Entropy (Basel), 2020, 23(1):50. DOI:10.3390/e23010050.
[12] Riglet F, Mentre F, Veyrat-Follet C, et al. Bayesian individual dynamic predictions with uncertainty of longitudinal biomarkers and risks of survival events in a joint modelling framework: a comparison between Stan, Monolix, and NONMEM[J]. AAPS J, 2020, 22(2): 50. DOI: 10.1208/s12248-019-0388-9.
[13] Alsefri M, Sudell M, García-Fiñana M, et al. Bayesian joint modelling of longitudinal and time to event data: a methodological review[J]. BMC Med Res Methodol, 2020, 20(1):94. DOI:10.1186/s12874-020-00976-2.
[14] Orphanou K, Stassopoulou A, Keravnou E. Temporal abstraction and temporal Bayesian networks in clinical domains: a survey[J]. Artif Intell Med, 2014, 60(3): 133-149. DOI:10.1016/j.artmed.2013.12.007.
[15] Liu MX, Stella F, Hommersom A, et al. A comparison between discrete and continuous time Bayesian networks in learning from clinical time series data with irregularity [J]. Artif Intell Med, 2019, 95: 104-117. DOI: 10.1016/j.artmed.2018.10.002.
[16] Oniśko A. Application of dynamic Bayesian networks to risk assessmnet in medicine[J]. Zesz Nauk Politech Białost Inform, 2010, 5:35-49.
[17] Sucar LE. Dynamic and temporal Bayesian networks[M]// Sucar LE. Probabilistic graphical models: principles and applications. Cham: Springer, 2021: 181-202. DOI: 10.1007/978-3-030-61943-5_9.
[18] Leong TY. Multiple perspective dynamic decision making [J]. Artif Intell, 1998, 105(1/2): 209-261. DOI: 10.1016/S0004-3702(98)00082-4.
[19] Nodelman U, Shelton CR, Koller D. Continuous time Bayesian networks[C]//Proceedings of the eighteenth conference on uncertainty in artificial intelligence. Alberta, Canada: Morgan Kaufmann Publishers Inc, 2002: 378-387.
[20] Villa S, Stella F. Learning continuous time Bayesian networks in non-stationary domains (extended abstract) [C]//Proceedings of the 27th international joint conference on artificial intelligence. Stockholm: AAAI, 2019:5656-5660.
[21] Gyftodimos E, Flach PA. Hierarchical Bayesian networks:a probabilistic reasoning model for structured domains [C]//Proceedings of the ICML-2002 Workshop on Development of Representations. Sydney: ICML, 2002: 23-30.
[22] Jarrett D, Yoon J, van der Schaar M. Dynamic prediction in clinical survival analysis using temporal convolutional networks[J]. IEEE J Biomed Health Inform, 2020, 24(2): 424-436. DOI:10.1109/JBHI.2019.2929264.
[23] Winkler L, Ojeda C, Opper M. Stochastic control for Bayesian neural network training[J]. Entropy (Basel), 2022, 24(8):1097. DOI:10.3390/E24081097.
[24] Austin RM, Onisko A, Druzdzel MJ. The Pittsburgh cervical cancer screening model: a risk assessment tool[J]. Arch Pathol Lab Med, 2010, 134(5): 744-750. DOI: 10.5858/134.5.744.
[25] Dagum, Paul & Galper, Adam. Forecasting Sleep Apnea with Dynamic Network Models. [C]. Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93). 2013. DOI: 10.1016/B978-1-4832-1451-1.50012-3.
[26] Orphanou K, Stassopoulou A, Keravnou E. Risk assessment for primary coronary heart disease event using dynamic Bayesian networks[C]//15th conference on artificial intelligence in medicine. Pavia: Springer, 2015: 161-165. DOI:10.1007/978-3-319-19551-3_20.
[27] Sheidaei A, Foroushani AR, Gohari K, et al. A novel dynamic Bayesian network approach for data mining and survival data analysis[J]. BMC Med Inform Decis Mak, 2022, 22(1):251. DOI:10.1186/s12911-022-02000-7.
[28] Lee H, Yun D, Yoo J, et al. Deep learning model for real-time prediction of intradialytic hypotension[J]. Clin J Am Soc Nephrol, 2021, 16(3): 396-406. DOI: 10.2215/CJN.09280620.
[29] Thorsen-Meyer HC, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit:a retrospective study of high-frequency data in electronic patient records[J]. Lancet Digit Health, 2020, 2(4): e179-191. DOI: 10.1016/S2589-7500(20)30018-2.
[30] Biazzo I, Braunstein A, Dall'Asta L, et al. A Bayesian generative neural network framework for epidemic inference problems[J]. Sci Rep, 2022, 12(1): 19673. DOI: 10.1038/s41598-022-20898-x.
[31] Fernandes V, Junior GB, de Paiva AC, et al. Bayesian convolutional neural network estimation for pediatric pneumonia detection and diagnosis[J]. Comput Methods Programs Biomed, 2021, 208: 106259. DOI: 10.1016/j.cmpb.2021.106259.
[32] Van Bergen GHH, Duenk P, Albers CA, et al. Bayesian neural networks with variable selection for prediction of genotypic values[J]. Genet Sel Evol, 2020, 52(1): 26. DOI: 10.1186/s12711-020-00544-8.
[33] Ekong F, Yu YB, Patamia RA, et al. Bayesian depth-wise convolutional neural network design for brain tumor MRI classification[J]. Diagnostics (Basel), 2022, 12(7): 1657. DOI:10.3390/DIAGNOSTICS12071657.
[34] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies[J]. Ann Intern Med, 2019, 170(1): 51-58. DOI:10.7326/M18-1376.
[35] 陈茹, 王胜锋, 周家琛, 等. 预测模型研究的偏倚风险和适用性评估工具解读[J]. 中华流行病学杂志, 2020, 41(5): 776-781. DOI:10.3760/cma.j.cn112338-20190805-00580.
Chen R, Wang SF, Zhou JC, et al. Introduction of the prediction model risk of bias assessment tool: a tool to assess risk of bias and applicability of prediction model studies[J]. Chin J Epidemiol, 2020, 41(5): 776-781. DOI: 10.3760/cma.j.cn112338-20190805-00580.
[36] Andrinopoulou ER, Harhay MO, Ratcliffe SJ, et al. Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data[J]. Int J Epidemiol, 2021, 50(5): 1731-1743. DOI: 10.1093/ije/dyab047.
[37] Kyrimi E, Mclachlan S, Dube K, et al. A comprehensive scoping review of Bayesian networks in healthcare: past, present and future[J]. Artif Intell Med, 2021, 117:102108. DOI:10.1016/j.artmed.2021.102108.