Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint Trajectories of Cognitive Function and Depressive Symptoms and Risk of Incident Osteoporosis Among Older Adults With Diabetes: A Multinational Longitudinal Cohort Study From HRS, SHARE, and CHARLS

Hui Guo
Beijing University of Chinese Medicine

Zunwang Li
Beijing University of Chinese Medicine

Ruizheng Zhu
Beijing University of Chinese Medicine

Yunhui Zhang
Beijing University of Chinese Medicine

Junde Wu
Beijing University of Chinese Medicine Third Affiliated Hospital

Zhaojun Chen
zhaojunchen66@126.com

Beijing University of Chinese Medicine Third Affiliated Hospital

Research Article

Keywords: Cognitive function, Depressive symptoms, Osteoporosis, Diabetes mellitus, Group-based trajectory modeling, Older adults, Longitudinal cohort

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9760312/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Joint Trajectories of Cognitive Function and Depressive Symptoms and Risk of Incident Osteoporosis Among Older Adults With Diabetes: A Multinational Longitudinal Cohort Study From HRS, SHARE, and CHARLS

Hui Guo[a,b†], Zunwang Li[a,b†] , Ruizheng Zhu[a,b], Yunhui Zhang[a,b], Junde Wu[b], and Zhaojun Chen[b*]

a Beijing University of Chinese Medicine, Beijing100029, China

b Department of Orthopedics, Beijing University of Chinese Medicine Third Affiliated Hospital, 6 Beijing100029, China

*Corresponding Author

Correspondence to: Z. Chen, Department of Orthopedics, Beijing University of Chinese Medicine Third Affiliated Hospital, Beijing 100029, China.E-mail addresses: zhaojunchen66@126.com

†These authors have contributed equally to this work and share first authorship

Abstract

Background

Cognitive decline and depressive symptoms have both been recognized as important factors associated with abnormal bone metabolism and an increased risk of osteoporosis in older adults. However, the impact of their long-term joint trajectory patterns on incident osteoporosis among older adults with diabetes remains unclear. Therefore, this study aimed to investigate the association between synchronized trajectories of cognitive function and depressive symptoms and the risk of incident osteoporosis in older adults with diabetes.

Methods

This study included participants aged ≥65 years from the HRS (n=2811), SHARE (n=5911), and CHARLS (n=2088) cohorts who were free of osteoporosis at baseline and had at least three repeated measurements of cognitive function and depressive symptoms during the first three survey waves. GBTM was applied to identify synchronized trajectories of cognitive function and depression. Cox proportional hazards models were subsequently used to evaluate the associations between trajectory groups and the risk of incident osteoporosis. Fine–Gray competing risk models, discrete-time logistic regression models, and subgroup analyses stratified by age and sex were further conducted to assess the robustness of the findings. In addition, seven machine learning models were developed following feature selection using the LASSO and Boruta algorithms to identify important predictors of incident osteoporosis. SHAP analysis was applied to interpret feature contributions within the models. Clinical utility was further evaluated using calibration curves and DCA.

Results

GBTM consistently identified two trajectory patterns across the three cohorts: the High Cognition and Low CESD group and the Low Cognition and High CESD group. The median follow-up durations were 8 years in HRS, 4 years in SHARE, and 6 years in CHARLS. After full adjustment for potential confounders, participants in the Low Cognition and High CESD group exhibited a significantly higher risk of incident osteoporosis than those in the High Cognition and Low CESD group, with hazard ratios (HRs) of 1.26 (95% CI: 1.05–1.51) in HRS, 1.39 (95% CI: 1.24–1.56) in SHARE, and 1.34 (95% CI: 1.12–1.62) in CHARLS. The competing risk analyses and discrete-time models yielded findings consistent with the primary analyses. Machine learning analyses demonstrated that the CATBoost model achieved the best predictive performance across all three cohorts. Furthermore, SHAP analysis identified the cognition–depression interaction trajectory as one of the most influential predictive features.

Conclusions

An adverse synchronized trajectory characterized by low cognitive function and high depressive symptoms was significantly associated with an increased risk of incident osteoporosis among older adults with diabetes. This association remained consistent across cohorts from different countries and regions, suggesting that the combined dynamic changes in cognitive function and psychological health may serve as an important indicator for osteoporosis risk stratification in older adults with diabetes.

Keywords: Cognitive function; Depressive symptoms; Osteoporosis; Diabetes mellitus; Group-based trajectory modeling; Older adults; Longitudinal cohort

1 Introduction

Against the backdrop of rapid global population aging, the prevalence of diabetes among older adults has increased substantially, resulting in persistent metabolic dysregulation and widespread multisystem impairment (1). Among the major complications associated with diabetes, osteoporosis represents a particularly important yet frequently underrecognized clinical burden (2). Notably, older adults with type 2 diabetes mellitus (T2DM) often exhibit the so-called "diabetic bone paradox," in which bone mineral density (BMD) remains normal or even elevated despite a markedly increased risk of fragility fractures (3). This paradox is largely attributed to defects in bone microarchitecture and reduced bone material strength that cannot be adequately detected by conventional bone densitometry techniques (4). Consequently, traditional risk factors and routine diagnostic approaches may be insufficient for accurately predicting fracture risk in this population. These limitations highlight the urgent need to identify novel early warning indicators to reduce the growing but underrecognized burden of diabetic osteoporosis (5).

Beyond conventional metabolic parameters, accumulating evidence suggests the existence of important and potentially shared pathological mechanisms linking cognitive function, depression, and bone metabolism (6). Cognitive decline is not merely an isolated neurological disorder; it may directly increase fall risk through neurodegenerative impairment of motor function (7). In addition, cognitive impairment may accelerate bone loss through systemic inflammation, oxidative stress, and dysregulation of the hypothalamic–pituitary–adrenal (HPA) axis (8). Similarly, depression may contribute to osteoporosis progression through several related mechanisms, including chronic low-grade inflammation, persistent hypercortisolemia, reduced physical activity, inadequate nutritional intake, and the osteotoxic effects of psychotropic medications, particularly selective serotonin reuptake inhibitors (SSRIs) (9, 10). Importantly, this "brain–bone axis" may be further amplified in the context of diabetes, in which chronic hyperglycemia and advanced glycation end products (AGEs) synergistically exacerbate neuroinflammation and impair osteoblast function (11, 12).

Despite the biological plausibility of these associations, current clinical evidence remains limited by several important methodological shortcomings. First, most previous studies have been cross-sectional and have relied on single-time-point measurements, making them vulnerable to bias and unable to capture the long-term dynamic changes characteristic of degenerative disease processes (13). Second, earlier investigations have primarily examined cognitive decline or depression independently, without adequately considering their high clinical co-occurrence and potential synergistic effects on bone health (14). Third, studies investigating joint longitudinal trajectories remain scarce. Existing analytical frameworks rarely evaluate the longitudinal co-evolution of these neuropsychological factors and therefore may overlook the possibility that dynamic interaction patterns, rather than static associations alone, contribute substantially to skeletal vulnerability. Finally, robust cross-national and cross-cultural validation remains lacking, underscoring the importance of harmonized international datasets to establish generalizability and external validity.

To address these gaps, the present study utilized longitudinal data from the HRS, SHARE, and CHARLS cohorts to investigate the joint trajectories of cognitive function and depressive symptoms, assessed using CES-D scores, and their associations with incident osteoporosis among adults aged ≥65 years with diabetes. In addition, the consistency of these associations across different countries and regions was evaluated. Machine learning models were further applied to assess predictive performance, identify key phenotypic predictors, and improve multidimensional risk stratification. We hypothesized that older adults with diabetes who exhibited persistent depressive symptoms combined with sustained low cognitive function would have the highest risk of developing incident osteoporosis.

2 Materials and Methods

2.1 Study Population

The Health and Retirement Study (HRS), conducted by the University of Michigan with support from the National Institute on Aging, is a nationally representative longitudinal study designed to examine the dynamic interactions among health, economic, and social factors in non-institutionalized adults aged ≥51 years in the United States.

The Survey of Health, Ageing and Retirement in Europe (SHARE), coordinated by the SHARE-ERIC consortium and the Munich Center for the Economics of Aging at the Max Planck Institute for Social Law and Social Policy, is a multinational longitudinal study that collects harmonized data on health, socioeconomic status, and social and family networks among community-dwelling adults aged ≥50 years across multiple European countries.

The China Health and Retirement Longitudinal Study (CHARLS), led by the National School of Development at Peking University and the China Center for Economic Research, is a nationally representative longitudinal survey that collects harmonized data on health, socioeconomic conditions, and family structure among community-dwelling adults aged ≥45 years across 28 provinces in China.

The present study was based on longitudinal analyses using data obtained from the HRS, SHARE, and CHARLS cohorts. All three cohorts are nationally representative studies involving community-dwelling older adults. Standardized interviews and physical examinations were conducted every two years to collect detailed information on sociodemographic characteristics, health conditions, functional status, and health-related behaviors.

In the HRS cohort, 20,129 participants were initially screened. Among them, 17,318 participants were excluded according to predefined criteria: (1) missing diabetes information during the first three waves (n=3431); (2) prevalent osteoporosis during the first three waves (n=8617); (3) missing CES-D scores during the first three waves (n=653); (4) age <65 years (n=4354); and (5) missing follow-up information (n=263). Ultimately, 2811 participants were included in the final analysis (**Figure 1**).

In the SHARE cohort, 57,982 participants were initially enrolled. A total of 52,071 participants were excluded based on the following criteria: (1) missing diabetes information during the first three waves (n=7929); (2) prevalent osteoporosis during the first three waves (n=13,516); (3) missing cognitive assessment scores during the first three waves (n=13,011); (4) age <65 years (n=14,692); and (5) missing follow-up information (n=2923). Consequently, 5911 participants were included in the final analysis (Figure 1).

In the CHARLS cohort, 17,708 participants were initially included. Of these, 15,620 participants were excluded for the following reasons: (1) missing diabetes information during the first three waves (n=1362); (2) prevalent osteoporosis during the first three waves (n=5301); (3) missing CES-D scores during the first three waves (n=1034); (4) missing cognitive assessment scores during the first three waves (n=125); (5) age <65 years (n=7443); and (6) missing follow-up information (n=355). Finally, 2088 participants were included in the CHARLS cohort analysis (Figure 1).

2.2Assessment of Cognitive Function and Depression and Construction of Joint Trajectory Models

Depressive symptoms were assessed using standardized instruments across the three cohorts. Different versions of the CES-D were used in the HRS and CHARLS cohorts. Specifically, the HRS employed the 8-item short-form CES-D with dichotomous scoring (score range: 0–8), whereas CHARLS used the 10-item short-form CES-D with four-point response scoring (score range: 0–30). In the SHARE cohort, depressive symptoms were evaluated using the EURO-D scale, which consists of 12 dichotomously scored items (score range: 0–12). In all scales, higher scores indicated more severe depressive symptoms. To account for differences in assessment tools and scoring ranges across cohorts, depression scores were calculated according to the original study protocols and subsequently standardized using Z-scores to improve cross-cohort comparability.

Cognitive function was assessed using the cognitive measures available within each cohort. In the HRS cohort, cognitive assessments included immediate and delayed word recall, serial subtraction of 7s, and temporal orientation. SHARE evaluated cognitive performance using immediate and delayed word recall, verbal fluency, and numeracy tasks. In CHARLS, cognitive function was assessed through episodic memory and mental status evaluations, including temporal orientation, serial subtraction of 7s, and figure drawing tasks. Total cognitive scores in each cohort were calculated by summing scores across multiple cognitive domains, with higher scores indicating better cognitive performance. Given the differences in cognitive assessment tools and scoring systems among cohorts, all cognitive scores were further standardized using Z-scores to enhance comparability across studies.

During longitudinal data processing, participants with at least three repeated measurements of CES-D scores and cognitive scores during baseline and follow-up assessments were included. Follow-

up duration, defined as years since baseline, was used as the time variable to construct an individual-level repeated-measures dataset.

GBTM was subsequently applied to characterize the long-term trajectories of depressive symptoms and cognitive function and to identify their joint longitudinal developmental patterns. Models with different numbers of trajectory groups (2–4 groups) and polynomial orders (linear and quadratic terms) were tested during model construction. The optimal model was selected based on the Bayesian Information Criterion (BIC), the proportion of participants within each trajectory group, and overall model interpretability. The final trajectory classification scheme represented the optimal balance between statistical goodness-of-fit and clinical interpretability.

2.3 Assessment of Diabetes and Incident Osteoporosis

In all three cohorts, diabetes status was determined using self-reported medical questionnaire data. Participants who responded "yes" to the question, "Have you ever been told by a doctor that you have diabetes?" were classified as having diabetes. Incident osteoporosis was defined as the first occurrence of osteoporosis during follow-up. Follow-up time was calculated from the baseline assessment to the first diagnosis of osteoporosis or the last follow-up assessment, whichever occurred first.

To improve the robustness and comparability of the findings, outcome variables were independently constructed within each cohort using consistent definitions and analytical procedures. These outcome variables were subsequently used to evaluate the associations between different trajectories of cognitive function and depression and the risk of incident osteoporosis.

To ensure the appropriate temporal sequence between exposure and outcome, all participants included in the trajectory analyses were free of osteoporosis during the trajectory modeling period (the first three waves). This design ensured that changes in cognitive function and depressive symptoms preceded the onset of osteoporosis, thereby reducing the potential for reverse causality.

2.4 Covariates

Relevant sociodemographic and lifestyle-related variables were collected as covariates, including age, sex, race, educational attainment, wealth, body mass index (BMI), hypertension, drinking status, smoking status, cardiovascular disease (CVD), physical activity, and cancer. Smoking status was categorized as never/former smoker or current smoker, whereas drinking status was classified as non-drinker or current drinker. BMI was calculated as weight in kilograms divided by height in meters squared (kg/m^2).

2.5 Machine Learning Feature Preprocessing and Feature Selection

Machine learning models were further developed to identify factors associated with incident osteoporosis. Variance inflation factors (VIFs) and pairwise correlation coefficients were calculated for all candidate variables to assess multicollinearity. Variables with a VIF >3 or correlation coefficients >0.9 were considered collinear and were excluded to reduce redundancy among predictors.

Least absolute shrinkage and selection operator (LASSO) regression was subsequently applied for additional feature selection after removal of collinear variables. LASSO is an L1 regularization–based feature selection approach that incorporates a penalty term into the loss function, thereby shrinking the coefficients of less informative variables toward zero and enabling automatic variable selection. The optimal penalty parameter (λ) was determined using cross-validation. Multiple iterative fitting procedures were performed to identify variables with predictive value for the outcome. Variables with non-zero coefficients after LASSO selection were retained for subsequent machine learning model construction.

The Boruta algorithm was also used for feature selection following collinearity screening. Boruta is a random forest–based feature selection method that evaluates feature importance by generating "shadow features" corresponding to each original variable in the dataset. During the selection process, the importance of original variables was compared with that of shadow variables across 500 iterations until stable evaluation results were achieved. Variables identified as "confirmed" by the Boruta algorithm were subsequently retained for machine learning model development.

To address the reduced ability of machine learning models to identify minority-class samples, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance class distributions between minority and majority groups. SMOTE generates synthetic minority-class samples by calculating distances between minority-class observations, identifying the k-nearest neighbors for each

sample, and generating new synthetic observations along the direction connecting neighboring samples. This approach effectively alleviates class imbalance within the dataset. To further reduce the risk of overfitting, dimensionality reduction was applied to majority-class samples before SMOTE implementation. In addition, all continuous variables were standardized, and categorical variables were transformed using one-hot encoding to minimize disproportionate influences arising from differences in variable scales during model training.

2.6 Statistical Analyses

Continuous variables were expressed as means and standard deviations (SDs), whereas categorical variables were presented as frequencies and percentages. Group differences were assessed using the chi-square test or Student's t-test, as appropriate.

To identify distinct trajectory patterns of cognitive function and depressive symptoms, GBTM based on finite mixture models was performed. This approach uses maximum likelihood estimation to classify individuals with similar developmental trajectories into latent subgroups. Trajectory analyses were conducted separately within each cohort. Cognitive function and depressive symptom data from the first three interview waves were incorporated into the models, with years since baseline treated as the time variable. Models containing one to four trajectory groups were evaluated. The optimal number of trajectory groups was determined comprehensively according to the BIC, AIC, log-likelihood values, and mean posterior probabilities for each trajectory group. Final model selection was based on the following criteria: the lowest BIC value, a minimum group proportion >10%, and an average posterior probability ≥0.70. Each participant was ultimately assigned to the trajectory group corresponding to the highest posterior probability (**Supplementary Table 1**).

Multivariable Cox proportional hazards regression models were used to examine the associations between joint trajectories of cognitive function and depressive symptoms and the risk of incident osteoporosis. Event time was defined as the interval from the assessment period of the first three waves to the first occurrence of osteoporosis, censoring, or the end of follow-up. Hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated after adjustment for potential confounding factors. Model 1 was an unadjusted model. Model 2 was adjusted for age, sex, race, education, wealth, and BMI. Model 3 was further adjusted for hypertension, drinking status, and smoking status. Model 4 additionally included CVD, physical activity, and cancer.

Because death may influence the occurrence of osteoporosis, competing risk regression analyses were further conducted using the Fine–Gray subdistribution hazard model, with death treated as a competing event. Subdistribution hazard ratios (sHRs) and 95% CIs were reported to evaluate the robustness of the findings within a competing-risk framework. To further assess the stability of the associations under an alternative time-setting strategy, discrete-time logistic regression models were fitted using follow-up waves as the time unit. The dataset was restructured into a person-period format, in which each participant contributed one observation for each follow-up wave until the occurrence of osteoporosis or censoring.

The dataset was randomly divided into training and validation sets at a ratio of 7:3 to minimize the risk of model overfitting. Seven machine learning algorithms were developed within the MLR3 framework, including Random Forest, LightGBM, K-nearest neighbors (K-NN), Naive Bayes, support vector machine (SVM), XGBoost, and CatBoost.

Random Forest generates predictions by constructing multiple decision trees and aggregating their outputs. This method demonstrates strong generalizability, effectively handles multiclass problems, and is relatively resistant to overfitting. LightGBM is a gradient boosting decision tree–based algorithm that improves training efficiency through optimized learning strategies. It is characterized by low memory consumption, high computational efficiency, support for parallel learning, and strong predictive performance in both classification and regression tasks. K-NN performs classification or regression based on the nearest neighboring data points. Its advantages include algorithmic simplicity, ease of implementation, the absence of a training phase, and favorable performance in small datasets and nonlinear problems. Naive Bayes is a probabilistic classification model based on Bayes' theorem and the assumption of feature independence. It is computationally efficient and robust when handling incomplete or missing feature information. SVM achieves classification by identifying the optimal separating hyperplane and performs effectively in high-dimensional datasets with complex decision

boundaries. XGBoost is a gradient boosting–based ensemble learning algorithm that improves predictive performance by integrating multiple weak learners. It offers rapid optimization, robustness to missing values, and excellent performance in large-scale datasets. CatBoost is also a gradient boosting decision tree–based ensemble learning algorithm that improves model stability and reduces prediction bias through ordered boosting and categorical feature encoding techniques. CatBoost can directly process categorical variables, performs well in small-sample datasets, maintains stable training performance, and effectively controls overfitting, making it suitable for both classification and regression tasks.

Benchmarking analyses were conducted to systematically compare the performance of the different machine learning models. Model performance was evaluated using standardized datasets and consistent evaluation metrics. Given the classification nature of the machine learning tasks, seven evaluation metrics were selected, including accuracy, F-beta score, area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, area under the precision–recall curve (AUC-PR), and Brier score. Among these metrics, AUC-ROC was considered the primary indicator for identifying the optimal machine learning model, whereas the remaining metrics served as secondary evaluation measures. Ten-fold cross-validation was performed to improve the stability and robustness of model performance. Analysis of variance (ANOVA) and the Kruskal–Wallis H test were used to compare differences in evaluation metrics across machine learning models.

The SHAP (SHapley Additive exPlanations) algorithm was applied to interpret feature contributions in the best-performing machine learning model. SHAP is a model interpretation framework based on Shapley values derived from game theory and can quantify the contribution of each feature to model predictions. By considering all possible feature combinations, SHAP assigns an importance value to each variable, thereby facilitating interpretation of the model decision-making process. This approach is widely applicable across different machine learning algorithms and has been extensively used in feature importance analysis and model interpretation, providing interpretability support for complex predictive models. Calibration curves and decision curve analysis (DCA) were further used to evaluate the calibration performance and clinical utility of the machine learning models.

All data processing and statistical analyses were performed using IBM SPSS Statistics (version 24.0) and R software (version 4.3.0). A two-sided p value <0.05 was considered statistically significant.

3 Results

3.1 Baseline Characteristics of Participants According to Incident Osteoporosis Status

Tables 1–3 summarize the baseline characteristics of participants from the HRS, SHARE, and CHARLS cohorts according to incident osteoporosis status during follow-up.

A total of 2,811 participants without osteoporosis at baseline were included from the HRS cohort, with a mean age of 73.60 years (SD=6.87). During a median follow-up period of 8 years, 1,098 participants developed incident osteoporosis, with a mean age of 72.80 years (SD=6.34). Compared with participants who did not develop osteoporosis, those with incident osteoporosis were more likely to be younger (72.80 vs. 74.12 years, $p<0.001$), female (58.47% vs. 41.53%, $p<0.001$), and to have a higher BMI (26.25 vs. 25.41, $p<0.001$) (**Table 1**).

In the SHARE cohort, 5,911 participants without osteoporosis at baseline were included, with a mean age of 72.93 years (SD=6.08). During a median follow-up duration of 4 years, 2,009 participants developed incident osteoporosis, with a mean age of 73.40 years (SD=6.15). Compared with participants who remained free of osteoporosis, those who developed osteoporosis were more likely to be female (63.71% vs. 36.29%, $p<0.001$), older (73.40 vs. 72.69 years, $p<0.001$), and to have a higher BMI (27.06 vs. 26.41, $p<0.001$) (**Table 2**).

In the CHARLS cohort, 2,088 participants without osteoporosis at baseline were included, with a mean age of 71.92 years (SD=5.57). During a median follow-up period of 6 years, 554 participants developed incident osteoporosis, with a mean age of 71.47 years (SD=5.51). Compared with participants who did not develop osteoporosis, those with incident osteoporosis were more likely to be female (52.53% vs. 47.47%, $p<0.001$), younger (71.47 vs. 72.09 years, $p=0.012$), and to have lower household income levels (16,202.49 vs. 22,372.41, $p=0.759$) (**Table 3**).

3.2 Associations Between Synchronized Trajectories of Cognitive Function and Depression and the Risk of Incident Osteoporosis

Table 4 presents the associations between synchronized trajectories of cognitive function and depressive symptoms and the risk of incident osteoporosis across the three cohorts. GBTM identified two consistent trajectory patterns in all cohorts based on the BIC: High Cognition and Low CESD, and Low Cognition and High CESD (**Figure 2A–C)**. Detailed model evaluation indices are provided in Supplementary Table 1.

Using the High Cognition and Low CESD group as the reference category, crude analyses demonstrated positive associations between adverse cognition–depression trajectories and the risk of incident osteoporosis across all three cohorts. After adjustment for potential confounders in Model 4, including age, sex, race, education, wealth, BMI, hypertension, drinking status, smoking status, CVD, physical activity, and cancer, participants in the HRS cohort belonging to the Low Cognition and High CESD group exhibited a significantly increased risk of incident osteoporosis compared with those in the reference group (adjusted HR=1.26, 95% CI: 1.05–1.51, p=0.012) (Table 4). Similarly, in the SHARE cohort, participants in the Low Cognition and High CESD group had a 39% higher risk of incident osteoporosis compared with those in the High Cognition and Low CESD group (adjusted HR=1.39, 95% CI: 1.24–1.56, p<0.001) (Table 4). Consistent results were observed in the CHARLS cohort. Relative to the High Cognition and Low CESD group, participants in the Low Cognition and High CESD group demonstrated a 34% increased risk of incident osteoporosis (adjusted HR=1.34, 95% CI: 1.12–1.62, p=0.002) (**Table 4**).

Kaplan–Meier survival analyses further demonstrated significant differences in osteoporosis-free survival probabilities among trajectory groups across all three cohorts (HRS: log-rank p<0.001; SHARE: log-rank p<0.001; CHARLS: log-rank p<0.001) (**Figure 3**).

Figure 4 illustrates the feature selection procedures across the three cohorts. **Figures 4A–C** present the VIFs for candidate variables. Variables with VIF values >3 were considered to exhibit multicollinearity. However, all variables across the three cohorts demonstrated VIF values <3. Figures 4D–F show the correlation coefficients among variables. Correlation coefficients >0.9 were considered indicative of collinearity; nevertheless, all variables remained below this threshold.

LASSO regression was subsequently applied to further screen variables that passed the multicollinearity assessments. The optimal penalty parameter (λ) was determined using 10-fold cross-validation. As log(λ) increased, regression coefficients progressively shrank toward zero. Cross-validation curves indicated that the optimal model fit was achieved at the λ values corresponding to the minimum mean squared error (HRS: 0.005; SHARE: 0.005; CHARLS: 0.002) (**Figures 4G–L**)

The Boruta algorithm was subsequently applied to variables retained after VIF, correlation, and LASSO screening. In the HRS cohort, all variables were identified as having “Confirmed” contributions to incident osteoporosis risk (Figure 4M). Similar findings were observed in the SHARE cohort, in which all retained variables were classified as having “Confirmed” importance (Figure 4N). In the CHARLS cohort, nine variables were identified as having “Confirmed” importance, whereas one variable was classified as “Rejected” (Figure 4O). A stringent variable selection strategy was adopted, and variables categorized as “Rejected” were excluded from subsequent analyses. Ultimately, only variables identified as having “Confirmed” importance were retained for machine learning model development.

3.4 Construction and Evaluation of Machine Learning Models

Figure 5 presents heatmaps summarizing the performance of seven machine learning models developed using the training and validation datasets from the HRS, SHARE, and CHARLS cohorts. The evaluated models included Random Forest, LightGBM, K-NN, Naive Bayes, SVM, XGBoost, and CatBoost. Model performance was assessed using the area under the AUC-ROC (**Figures 6A–F**), area under the AUC-PR (Figures 6G–L), Accuracy, F-beta score, Sensitivity, Specificity, and Brier score. Separate machine learning frameworks were independently established for the HRS (**Supplementary Tables 2–3**), SHARE (**Supplementary Tables 4–5**), and CHARLS cohorts (**Supplementary Tables 6–7**).

Among the models developed using the HRS cohort, CatBoost demonstrated the best overall performance in the training dataset. Specifically, CatBoost achieved the highest or near-highest values for AUC-ROC (0.721), Accuracy (0.650), F-beta score (0.672), Sensitivity (0.716), Specificity (0.588),

and AUC-PR (0.680), while yielding the lowest Brier score (0.210), indicating superior overall predictive performance. Random Forest and LightGBM showed the next-best performance, with AUC-ROC values of 0.720 and 0.701, AUC-PR values of 0.680 and 0.659, Accuracy values of 0.647 and 0.635, and F-beta scores of 0.666 and 0.649, respectively; however, their overall performance remained inferior to that of CatBoost. K-NN and SVM exhibited moderate predictive capability (**Supplementary Table 2; Figures 5A, 6A, and 6G**).

In the validation dataset of the HRS cohort, CatBoost maintained superior predictive performance, achieving an AUC-ROC of 0.699, Accuracy of 0.635, F-beta score of 0.657, Sensitivity of 0.705, Specificity of 0.566, AUC-PR of 0.656, and the lowest Brier score (0.218) (**Supplementary Table 3; Figures 5D, 6D, and 6J**). Collectively, these findings indicated that CatBoost was the optimal model for predicting incident osteoporosis risk in the HRS cohort.

Similarly, among the models developed using the SHARE cohort, CatBoost achieved the best overall performance in the training dataset, with an Accuracy of 0.699, F-beta score of 0.735, AUC-ROC of 0.761, Sensitivity of 0.844, Specificity of 0.557, AUC-PR of 0.699, and the lowest Brier score (0.188) (**Supplementary Table 4; Figures 5B, 6B, and 6H**). In the validation dataset, CatBoost continued to outperform the remaining models, achieving an Accuracy of 0.682, F-beta score of 0.717, AUC-ROC of 0.749, Sensitivity of 0.790, Specificity of 0.568, AUC-PR of 0.717, and a Brier score of 0.197 (**Supplementary Table 5; Figures 5E, 6E, and 6K**). These results further confirmed the robust predictive performance of CatBoost for incident osteoporosis in the SHARE cohort.

In the CHARLS cohort, Random Forest demonstrated the best overall performance in the training dataset, achieving an Accuracy of 0.713, F-beta score of 0.747, AUC-ROC of 0.757, Sensitivity of 0.832, Specificity of 0.588, AUC-PR of 0.696, and a Brier score of 0.189 (**Supplementary Table 6; Figures 5C, 6C, and 6I**). However, in the validation dataset, CatBoost emerged as the best-performing model, with an Accuracy of 0.710, F-beta score of 0.694, AUC-ROC of 0.795, Sensitivity of 0.704, Specificity of 0.714, AUC-PR of 0.747, and the lowest Brier score (0.187) (**Supplementary Table 7; Figures 5F, 6F, and 6L**). Given its superior validation performance and the consistent findings observed in both the HRS and SHARE cohorts, CatBoost was ultimately identified as the optimal model for predicting incident osteoporosis risk in the CHARLS cohort.

Overall, across both training and validation datasets from all three cohorts, CatBoost consistently exhibited the best predictive performance among the seven machine learning models for incident osteoporosis prediction. Significant differences in model performance were observed across all evaluation metrics ($P<0.001$) (**Supplementary Tables 2–7**).

3.5 Evaluation of Model Calibration and Clinical Utility

To further assess predictive accuracy and clinical applicability, calibration curves and DCA were performed for all machine learning models. Across the three cohorts, calibration analyses demonstrated that the CatBoost model produced prediction curves closest to the ideal calibration line, indicating excellent agreement between predicted and observed outcomes (**Figures 7A–C**).

DCA further revealed that CatBoost consistently achieved the highest net benefit across threshold probabilities ranging from 25% to 75%, outperforming all other machine learning models and substantially exceeding the clinical utility of the "Treat All" and "Treat None" strategies (**Figures 7D–F**).

3.6 SHAP Interpretation of Feature Importance in the Best-Performing Machine Learning Models Across Different Cohorts

The SHAP algorithm was applied to interpret feature contributions within the best-performing prediction models constructed from the HRS, SHARE, and CHARLS cohorts. **Figures 8A–C** illustrate the most influential variables contributing to incident osteoporosis prediction across the three cohorts. SHAP values were used to quantify the relative contribution of each feature to model predictions.

Notably, the cognition-and-CESD trajectory variable was consistently identified as one of the most important predictive features across all three cohorts, with particularly strong importance observed in the CHARLS cohort. Force plots and waterfall plots (**Figures 8D–I**) were further generated to provide case-level explanations for the CatBoost models. In these visualizations, light-colored features represented positive contributions toward the predicted outcome.

In the HRS-based model, the baseline predicted probability for preventing incident osteoporosis

was 0.408. After integrating the contributions of multiple features, the final predicted probability increased to 0.631 (**Figures 8D and 8G**). In the SHARE-based model, the baseline predicted probability increased from 0.374 to 0.680 after incorporating feature contributions (Figures 8E and 8H). Similarly, in the CHARLS-based model, the baseline predicted probability increased from 0.488 to 0.794 following integration of the feature effects (**Figures 8F and 8I**).

3.7 Sensitivity Analyses

A series of sensitivity analyses were conducted to further evaluate the robustness of the association between the joint trajectories of cognitive function and depressive symptoms and the risk of incident osteoporosis.

First, because death may influence the occurrence of osteoporosis, Fine–Gray competing risk regression analyses were performed with death treated as a competing event. The results of the competing risk analyses were highly consistent with those of the primary Cox regression models.

After trajectory classification using GBTM and adjustment for potential confounding variables, the High Cognition and Low CESD group remained the reference category. In the HRS cohort, participants in the Low Cognition and High CESD group exhibited an approximately 24% increased risk of incident osteoporosis (adjusted sHR=1.24, 95% CI: 1.05–1.47, p=0.011) (Supplementary Table 8). In the SHARE cohort, the Low Cognition and High CESD group also demonstrated a significantly elevated risk after accounting for death as a competing event (adjusted sHR=1.27, 95% CI: 1.15–1.40, p<0.001). Similarly, in the CHARLS cohort, the same trajectory group showed an approximately 31% higher risk of incident osteoporosis (adjusted sHR=1.31, 95% CI: 1.11–1.55, p=0.002) (Supplementary Table 8). The magnitude and direction of these associations were comparable to those observed in the Cox regression analyses, indicating that the relationship between synchronized cognition–depression trajectories and incident osteoporosis risk was not materially affected by the competing risk of death.

Second, discrete-time logistic regression models were fitted to further examine the association between cognition–depression trajectories and incident osteoporosis risk. Follow-up waves were treated as the time unit, and the dataset was restructured into a person-period format. In the HRS cohort, compared with the High Cognition and Low CESD group, the Low Cognition and High CESD group remained significantly associated with an increased risk of incident osteoporosis (adjusted OR=1.29, 95% CI: 1.06–1.56, p=0.011). Similar associations were observed in the SHARE cohort (adjusted OR=1.35, 95% CI: 1.17–1.56, p<0.001) and the CHARLS cohort (adjusted OR=1.48, 95% CI: 1.22–1.80, p<0.001). The discrete-time regression results were directionally consistent with those from the Cox and competing risk models, further supporting the robustness of the findings across different temporal modeling strategies (Supplementary Table 9).

Finally, subgroup analyses stratified by age and sex were conducted in the HRS, SHARE, and CHARLS cohorts (Figure 9). In the HRS cohort, compared with the High Cognition and Low CESD trajectory group, the Low Cognition and High CESD trajectory group was significantly associated with an increased risk of incident osteoporosis among participants aged ≤75 years (HR=1.26, 95% CI: 1.01–1.58, p=0.043) and among women (HR=1.28, 95% CI: 1.02–1.60, p=0.030). However, these associations did not reach statistical significance among participants aged >75 years (HR=1.28, 95% CI: 0.94–1.73, p=0.112) or among men (HR=1.23, 95% CI: 0.90–1.69, p=0.194). These findings suggest that the adverse cognition–depression trajectory may exert a stronger effect on osteoporosis risk among relatively younger older adults and women in the HRS cohort (Figure 9A).

In the SHARE cohort, the association remained statistically significant across all subgroups. The HRs for participants aged ≤75 years and >75 years were 1.42 (95% CI: 1.19–1.69, p<0.001) and 1.26 (95% CI: 1.08–1.47, p=0.003), respectively. Among women and men, the HRs were 1.32 (95% CI: 1.15–1.51, p<0.001) and 1.29 (95% CI: 1.03–1.60, p=0.025), respectively. These findings indicate that the trajectory characterized by low cognitive function and high depressive symptoms was significantly associated with increased incident osteoporosis risk regardless of age or sex stratification, although the magnitude of the association appeared greater among women and individuals aged ≤75 years (Figure 9B).

In the CHARLS cohort, significant associations were observed among participants aged ≤75 years (HR=1.36, 95% CI: 1.11–1.67, p=0.003) and among women (HR=1.56, 95% CI: 1.19–2.05, p=0.001). In contrast, no statistically significant associations were identified among participants aged >75 years

(HR=1.15, 95% CI: 0.74–1.76, p=0.537) or among men (HR=1.15, 95% CI: 0.88–1.50, p=0.311). These findings suggest that, among older Chinese adults, the adverse cognition–depression trajectory may exert a more pronounced effect on incident osteoporosis risk in individuals aged 65–75 years and in women (Figure 9C).

Overall, the sensitivity analyses across the three cohorts consistently supported the primary findings, indicating that the adverse joint trajectory characterized by low cognitive function and high depressive symptoms was associated with an increased risk of incident osteoporosis. These results further strengthened the robustness and cross-population generalizability of the study findings.

4 Discussion

Using data from the large international longitudinal cohorts HRS, SHARE, and CHARLS, this study systematically examined the associations between synchronized trajectories of cognitive function and depressive symptoms and the risk of incident osteoporosis among older adults with diabetes. Across populations from the United States, Europe, and China, individuals characterized by a trajectory of persistently low cognitive function and elevated depressive symptoms consistently exhibited a significantly increased risk of incident osteoporosis. These associations remained robust after multivariable adjustment, competing-risk analyses, and discrete-time modeling. Importantly, the present study moved beyond conventional single time-point assessments of cognition or depression by emphasizing the importance of their long-term dynamic interplay in relation to bone health. Furthermore, machine learning analyses confirmed that cognition–depression joint trajectories were among the most important predictors of osteoporosis risk. These findings suggest that cognitive decline and depressive symptoms may synergistically contribute to abnormal bone metabolism and osteoporosis development in older adults with diabetes through shared biological and behavioral pathways.

The detrimental effects of neuropsychiatric deterioration on skeletal health in patients with diabetes may be explained by a multifactorial “brain–bone–metabolism axis” (15). Chronic systemic inflammation likely represents one of the principal biological mechanisms. Elevated levels of proinflammatory cytokines, particularly IL-6, TNF-α, and CRP, may collectively accelerate neurodegeneration, exacerbate depressive phenotypes, and directly stimulate osteoclastogenesis through activation of the RANKL pathway, thereby promoting an accelerated state of “inflammaging” (16, 17). Dysregulation of the HPA axis may constitute another central mechanism. Persistent activation of the HPA axis in depression results in chronically elevated cortisol levels, which suppress osteoblast-mediated bone formation while promoting osteoclast-mediated bone resorption (18, 19). Concurrently, HPA-axis dysfunction may aggravate hippocampal atrophy and accelerate cognitive decline (20). In the context of diabetes, these pathological processes may be further amplified by metabolic disturbances. Systemic insulin resistance, accumulation of advanced glycation end products (AGEs), chronic oxidative stress, and widespread microvascular dysfunction may simultaneously impair the blood–brain barrier and skeletal microvasculature, resulting in parallel ischemic and metabolic injury to neural and bone tissues (21, 22). In addition, behavioral factors may further exacerbate these biological abnormalities. Severe depressive symptoms and cognitive impairment are frequently associated with reduced physical activity, increased fall risk, poor nutritional status, insufficient calcium and vitamin D intake, and lower adherence to antidiabetic and bone-protective therapies (23, 24).

A major methodological strength of the present study was the application of joint trajectory modeling, which shifted the analytical framework from static exposure assessment toward dynamic longitudinal co-evolution analysis. Previous studies have largely evaluated cognitive decline and depression as independent exposures, despite their high comorbidity and well-established bidirectional relationship in older populations (25, 26). Depression has been identified as a major contributor to accelerated cognitive decline, whereas progressive cognitive impairment may, in turn, worsen depressive symptoms (27, 28). Therefore, the long-term synchronized deterioration of cognition and mood may reflect not merely the coexistence of two conditions, but rather a more profound state of systemic neurobiological and metabolic aging (29). By applying GBTM, the present study successfully captured the longitudinal heterogeneity and temporal evolution of these neuropsychiatric factors (30). Compared with single time-point measurements, trajectory-based phenotyping provides a more comprehensive representation of the cumulative burden of neuropsychiatric deterioration and may therefore serve as a more informative biomarker of future osteoporosis risk (31).

The external validity and generalizability of the present findings were strengthened by the high degree of consistency observed across three distinct international cohort studies, namely the HRS, the SHARE, and the CHARLS (32-34). Despite substantial differences in healthcare systems, cultural attitudes toward mental health, dietary habits, and genetic backgrounds, the predictive significance of the cognition–depression joint trajectory remained remarkably consistent across populations. This cross-national reproducibility strongly suggests that the interaction among cognitive decline, depressive symptoms, and bone metabolism in diabetes may reflect a fundamental biological process (21, 35). Accordingly, this synchronized trajectory pattern may represent a reliable and globally applicable biomarker for osteoporosis risk assessment in aging populations.

From a clinical perspective, these findings have important implications. Current management strategies for older adults with diabetes primarily focus on glycemic control and the prevention of conventional macrovascular and microvascular complications, such as cardiovascular disease and nephropathy, whereas neuropsychiatric dysfunction and occult osteoporosis are frequently underrecognized (36). The present findings suggest that longitudinal joint assessment of cognitive function and depressive symptoms may provide a practical and cost-effective strategy for early identification of diabetic older adults at elevated risk of osteoporosis (37). Moreover, integration of machine learning approaches substantially enhanced the predictive utility of the study models. The CatBoost algorithm demonstrated superior stability and generalizability when handling complex nonlinear epidemiological data (38, 39). In parallel, SHAP improved model interpretability by quantifying the relative contributions of neuropsychiatric trajectories to osteoporosis risk prediction. Collectively, these approaches may facilitate individualized risk stratification and support precision prevention strategies in geriatric endocrinology.

Several strengths of this study should be highlighted, including the use of three large international longitudinal cohorts, prolonged follow-up duration, innovative application of joint trajectory analysis, comprehensive sensitivity analyses, and the integration of machine learning algorithms with SHAP-based interpretability for cross-national validation. Nevertheless, several limitations warrant consideration. First, incident osteoporosis was identified partly through self-reported physician diagnoses, which may have introduced recall bias and misclassification, although this approach has been widely validated in large-scale epidemiological studies. Second, heterogeneity existed among the psychological assessment instruments used across cohorts, including CES-D and EURO-D. However, Z-score standardization was applied before trajectory modeling to minimize inter-cohort measurement variability. Finally, despite extensive adjustment for potential confounders, residual confounding cannot be completely excluded, particularly with respect to unmeasured longitudinal changes in dietary patterns, inflammatory biomarkers, and cumulative medication exposure.

5 Conclusion

Using data from three large international longitudinal aging cohorts, this study demonstrated that a long-term synchronized trajectory characterized by persistently low cognitive function and elevated depressive symptoms was significantly associated with an increased risk of incident osteoporosis among older adults with diabetes. This association remained stable across different regions, statistical approaches, and multiple sensitivity analyses. These findings suggest that the joint dynamic evolution of cognitive impairment and depressive symptoms may serve as an important early warning indicator for osteoporosis risk in older diabetic populations. Future research and clinical practice should prioritize the early identification and comprehensive management of individuals with coexisting cognitive decline and depressive symptoms to reduce the risk of osteoporosis and related adverse outcomes.

Ethics approval and consent to participate

The Health and Retirement Study (HRS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the China Health and Retirement Longitudinal Study (CHARLS) were approved by the corresponding institutional ethics review boards in the United States, Europe, and China. All participants provided written informed consent before participation. The present study used publicly available de-identified datasets and was exempt from additional ethical approval.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes research and clinical practice. 2019;157:107843.
2. Hofbauer LC, Compston JE, Saag KG, Rauner M, Tsourdi E. Glucocorticoid-induced osteoporosis: novel concepts and clinical implications. The lancet Diabetes & endocrinology. 2025;13(11):964-79.
3. Bahat G, Erdogan T, Ozturk S, Selcukbiricik OS, Ozkok S, Yavuz DG, et al. Managing Bone Fragility in Older Adults with Diabetes: Pathophysiology, Assessment, and Therapeutic Considerations. Drugs & aging. 2026;43(4):341-60.
4. Yuan B, Zhang X, Hu H. Application of MRI-based vertebral bone quality scoring and its integration with BMD for identifying fragility fractures in postmenopausal women with type 2 diabetes mellitus. Endocrine. 2026;91(1).
5. Picke AK, Campbell G, Napoli N, Hofbauer LC, Rauner M. Update on the impact of type 2 diabetes mellitus on bone metabolism and material properties. Endocrine connections. 2019;8(3):R55-r70.
6. Yuan J, Meloni BP, Shi T, Bonser A, Papadimitriou JM, Mastaglia FL, et al. The Potential Influence of Bone-Derived Modulators on the Progression of Alzheimer's Disease. Journal of Alzheimer's disease : JAD. 2019;69(1):59-70.
7. Ren J, Liu M, Wang B, Bai C, Yang K, Ma X, et al. Secondary fracture risk after hip fracture in older patients: an updated sex-specific systematic review and meta analysis with GRADE and PAF assessment. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2026.
8. Hardy RS, Zhou H, Seibel MJ, Cooper MS. Glucocorticoids and Bone: Consequences of Endogenous and Exogenous Excess and Replacement Therapy. Endocrine reviews. 2018;39(5):519-48.
9. Kelly RR, McDonald LT, Jensen NR, Sidles SJ, LaRue AC. Impacts of Psychological Stress on Osteoporosis: Clinical Implications and Treatment Interactions. Frontiers in psychiatry. 2019;10:200.
10. Azuma K, Adachi Y, Hayashi H, Kubo KY. Chronic Psychological Stress as a Risk Factor of Osteoporosis. Journal of UOEH. 2015;37(4):245-53.
11. Biessels GJ, Despa F. Cognitive decline and dementia in diabetes mellitus: mechanisms and clinical implications. Nature reviews Endocrinology. 2018;14(10):591-604.
12. Farr JN, Khosla S. Determinants of bone strength and quality in diabetes mellitus in humans. Bone. 2016;82:28-34.
13. Chen X, Giles J, Yao Y, Yip W, Meng Q, Berkman L, et al. The path to healthy ageing in China: a Peking University-Lancet Commission. Lancet (London, England). 2022;400(10367):1967-2006.
14. Rice J, Falck RS, Seo YS, Dian L, Davis JC, Jehu DA, et al. Home-based exercise reduces fall risk in older adults with mild cognitive impairment who have sustained a hip fracture: A 6-month randomized controlled trial. Maturitas. 2026;206:108832.
15. Muka T, Oliver-Williams C, Kunutsor S, Laven JS, Fauser BC, Chowdhury R, et al. Association of Age at Onset of Menopause and Time Since Onset of Menopause With Cardiovascular Outcomes, Intermediate Vascular Traits, and All-Cause Mortality: A Systematic Review and Meta-analysis. JAMA cardiology. 2016;1(7):767-76.
16. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nature reviews Endocrinology. 2018;14(10):576-90.
17. Ferrucci L, Fabbri E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. Nature reviews Cardiology. 2018;15(9):505-22.
18. Shrestha P, Popuri SB, Hoffer C, Qureshi A, Barzilay J, Dapaah EA, et al. Management of glucocorticoid-induced osteoporosis in rheumatic diseases. Best practice & research Clinical rheumatology. 2026:102151.
19. Rosenblat JD, Gregory JM, Carvalho AF, McIntyre RS. Depression and Disturbed Bone Metabolism: A Narrative Review of the Epidemiological Findings and Postulated Mechanisms. Current molecular medicine. 2016;16(2):165-78.
20. Lee E, Lee MJ, Park B, Park I. Risk of fracture according to glucocorticoid use after renal biopsy: a nationwide population-based study. Scientific reports. 2020;10(1):13846.
21. Balducci S, Haxhi J, Mattia L, Vitale M, Pugliese L, Argento G, et al. Effects of exercise training on bone health in elderly people with type 2 diabetes without osteoporosis: A randomised clinical trial. Diabetes research and clinical practice. 2026:113303.
22. Srikanth V, Sinclair AJ, Hill-Briggs F, Moran C, Biessels GJ. Type 2 diabetes and cognitive dysfunction-

towards effective management of both comorbidities. The lancet Diabetes & endocrinology. 2020;8(6):535-45.
23. Morley JE, Vellas B, van Kan GA, Anker SD, Bauer JM, Bernabei R, et al. Frailty consensus: a call to action. Journal of the American Medical Directors Association. 2013;14(6):392-7.
24. Stubbs B, Koyanagi A, Schuch FB, Firth J, Rosenbaum S, Veronese N, et al. Physical activity and depression: a large cross-sectional, population-based study across 36 low- and middle-income countries. Acta psychiatrica Scandinavica. 2016;134(6):546-56.
25. Katon W, Pedersen HS, Ribe AR, Fenger-Grøn M, Davydow D, Waldorff FB, et al. Effect of depression and diabetes mellitus on the risk for dementia: a national population-based cohort study. JAMA psychiatry. 2015;72(6):612-9.
26. Mace RA, Law ME, Cohen JE, Ritchie CS, Okereke OI, Hoeppner BB, et al. A Mindfulness-Based Lifestyle Intervention for Dementia Risk Reduction: Protocol for the My Healthy Brain Feasibility Randomized Controlled Trial. JMIR research protocols. 2024;13:e64149.
27. Bennett S, Thomas AJ. Depression and dementia: cause, consequence or coincidence? Maturitas. 2014;79(2):184-90.
28. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet (London, England). 2020;396(10248):413-46.
29. Jacka FN, O'Neil A, Opie R, Itsiopoulos C, Cotton S, Mohebbi M, et al. A randomised controlled trial of dietary improvement for adults with major depression (the 'SMILES' trial). BMC medicine. 2017;15(1):23.
30. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annual review of clinical psychology. 2010;6:109-38.
31. Muthén B, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcoholism, clinical and experimental research. 2000;24(6):882-91.
32. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). International journal of epidemiology. 2014;43(2):576-85.
33. Börsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, et al. Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). International journal of epidemiology. 2013;42(4):992-1001.
34. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). International journal of epidemiology. 2014;43(1):61-8.
35. Vestergaard P. Discrepancies in bone mineral density and fracture risk in patients with type 1 and type 2 diabetes--a meta-analysis. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2007;18(4):427-44.
36. Lin XZ, Meng RL, Peng DD, Li C, Zheng XY, Xu HF, et al. Cross-sectional study on prevalence and risk factors for falls among the elderly in communities of Guangdong province, China. BMJ open. 2022;12(11):e062257.
37. Russo GT, Giandalia A, Romeo EL, Nunziata M, Muscianisi M, Ruffo MC, et al. Fracture Risk in Type 2 Diabetes: Current Perspectives and Gender Differences. International journal of endocrinology. 2016;2016:1615735.
38. Lyu Y, Wu HM, Yan HX, Guo R, Xiong YJ, Chen R, et al. Classification of coronary artery disease using radial artery pulse wave analysis via machine learning. BMC medical informatics and decision making. 2024;24(1):256.
39. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart (British Cardiac Society). 2018;104(14):1156-64.

Figure 1 Flowchart of participant selection.

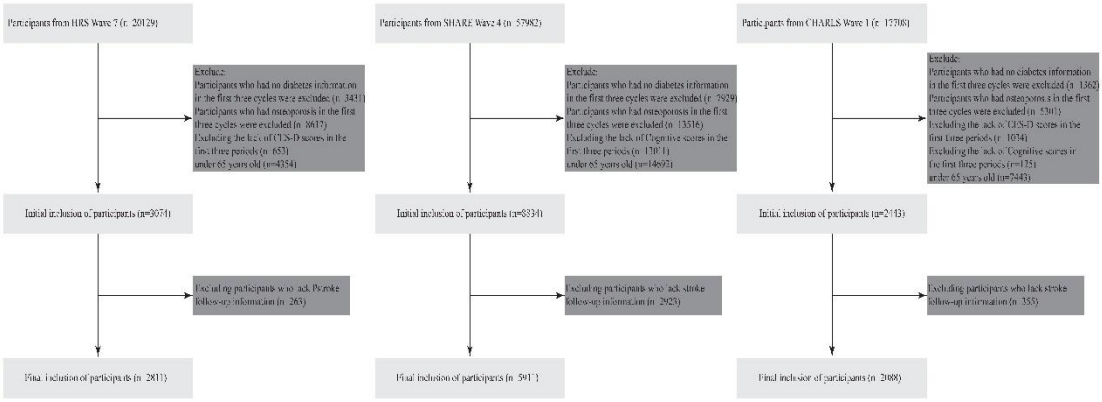


Figure 2 Group-based trajectory modeling (GBTM) identified the joint trajectories of cognitive function and grip strength across different cohorts. (A) HRS; (B) SHARE; (C) CHARLS.

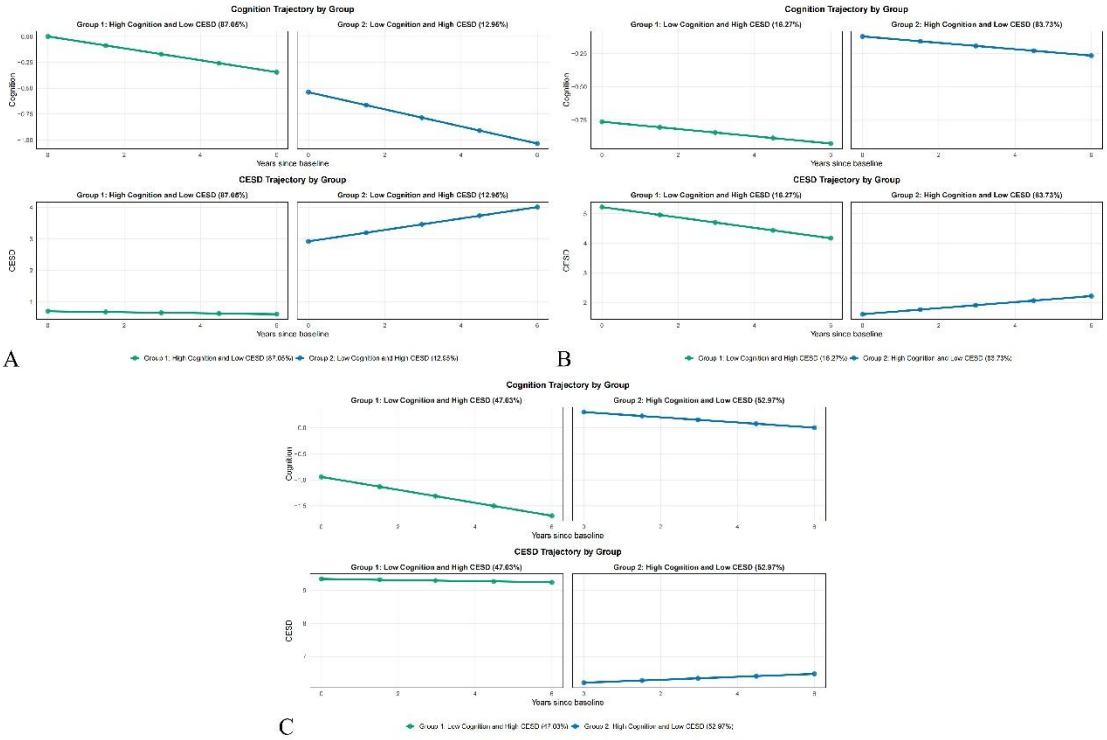


Figure 3 Kaplan–Meier (KM) curves illustrating osteoporosis-free survival probabilities across different trajectory groups in the three cohorts. (A) HRS; (B) SHARE; (C) CHARLS.

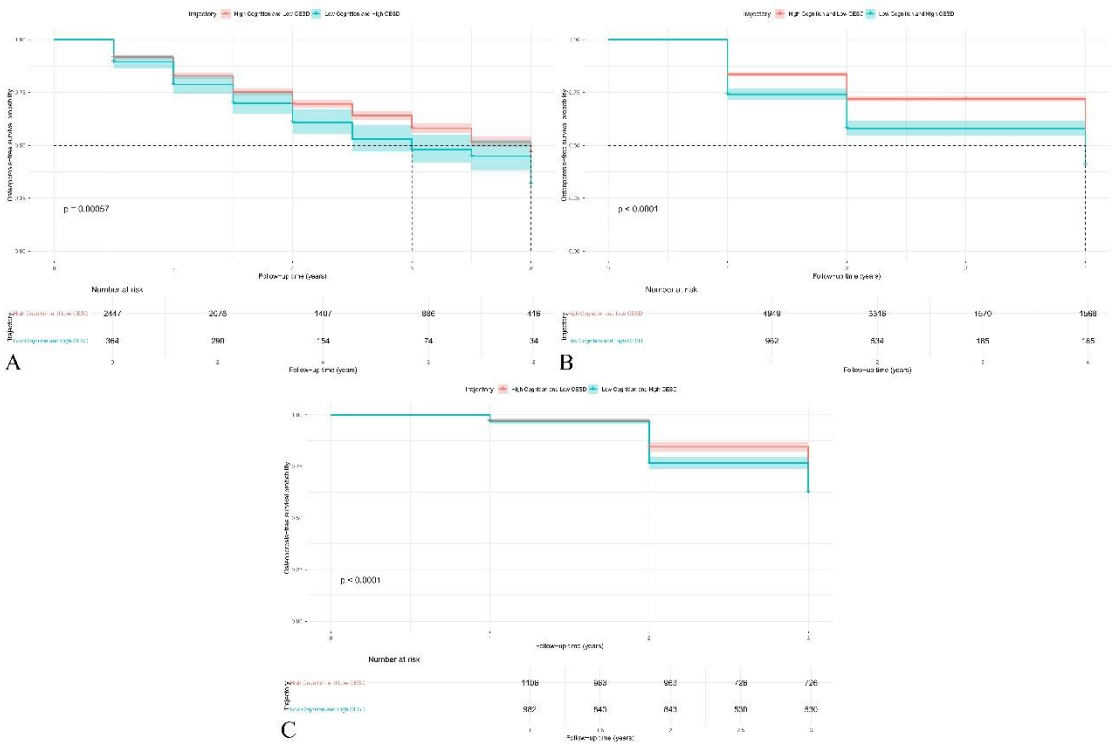


Figure 4 Feature selection using variance inflation factor (VIF) analysis, correlation analysis, and the Boruta algorithm. (A) VIF analysis in the HRS cohort; (B) VIF analysis in the SHARE cohort; (C) VIF analysis in the CHARLS cohort; (D) correlation heatmap in the HRS cohort; (E) correlation heatmap in the SHARE cohort; (F) correlation heatmap in the CHARLS cohort; (G) LASSO coefficient plot in the HRS cohort; (H) LASSO coefficient plot in the SHARE cohort; (I) LASSO coefficient plot in the CHARLS cohort; (J) LASSO cross-validation path plot in the HRS cohort; (K) LASSO cross-validation path plot in the SHARE cohort; (L) LASSO cross-validation path plot in the CHARLS cohort; (M) Boruta importance plot in the HRS cohort; (N) Boruta importance plot in the SHARE cohort; (O) Boruta importance plot in the CHARLS cohort.

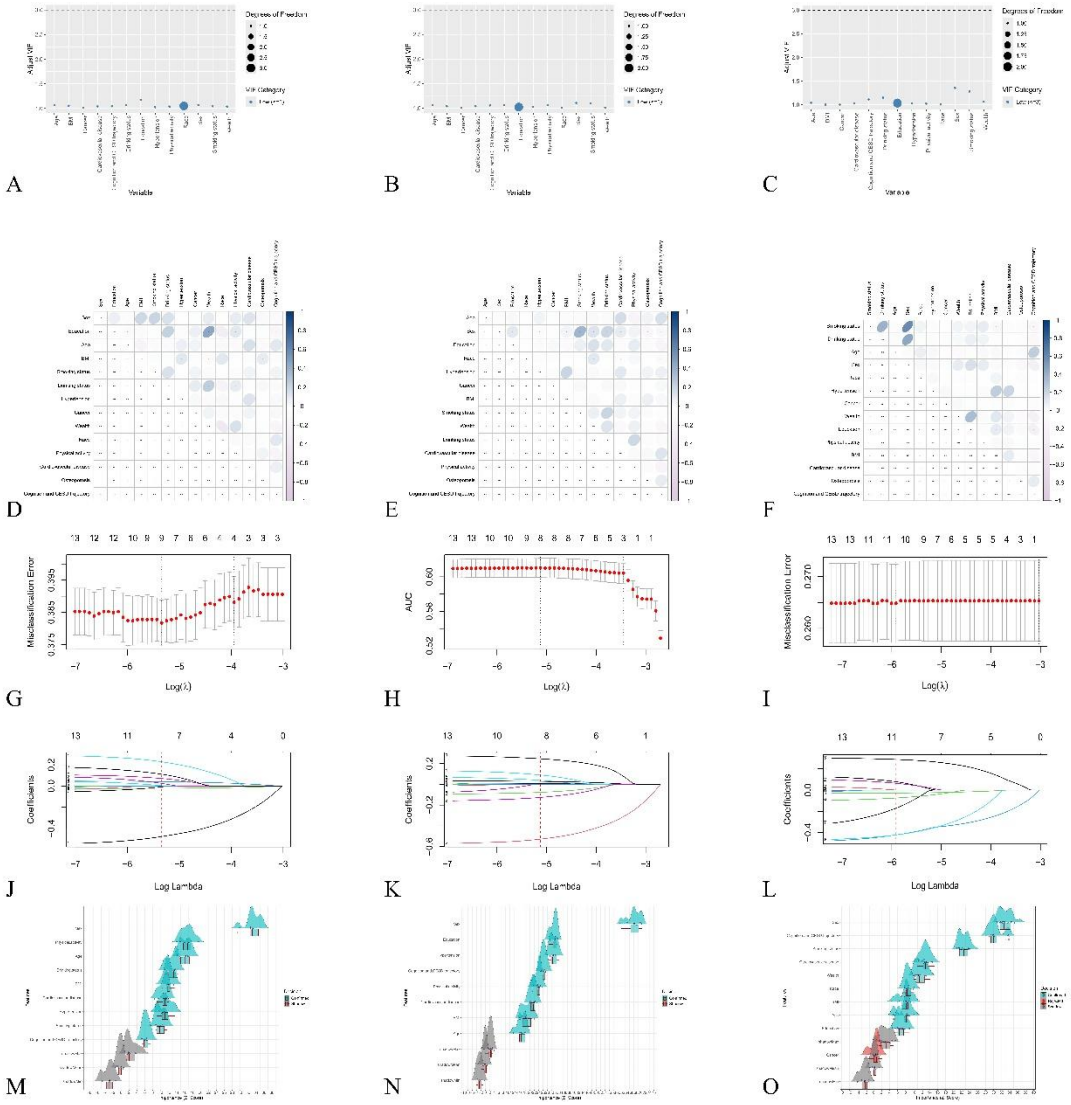


Figure 5 Heatmaps illustrating the performance of seven machine learning models across the three cohorts. (A) HRS training dataset; (B) SHARE training dataset; (C) CHARLS training dataset; (D) HRS validation dataset; (E) SHARE validation dataset; (F) CHARLS validation dataset.

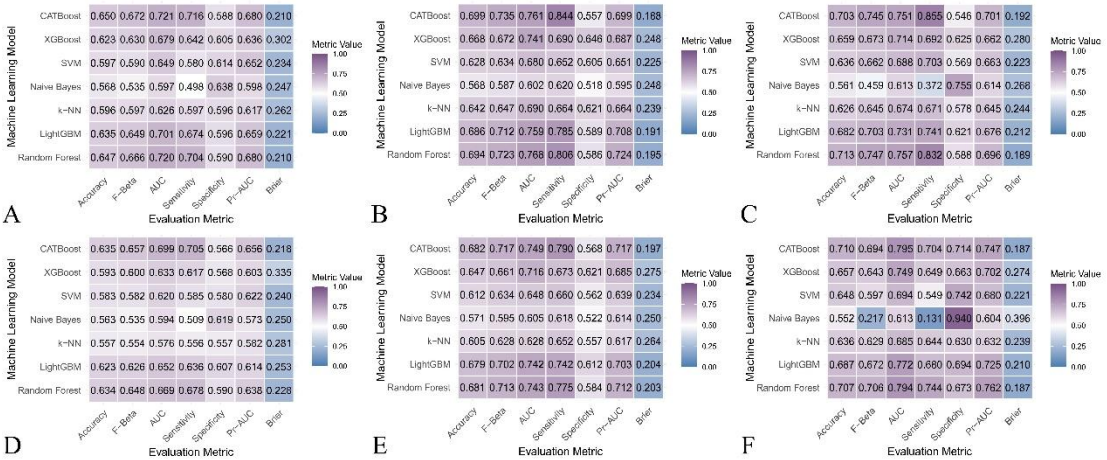


Figure 6 Receiver operating characteristic (ROC) curves and precision–recall (PR) curves of seven machine learning models across the three cohorts. (A) ROC curve for the HRS training dataset; (B)

ROC curve for the SHARE training dataset; (C) ROC curve for the CHARLS training dataset; (D) ROC curve for the HRS validation dataset; (E) ROC curve for the SHARE validation dataset; (F) ROC curve for the CHARLS validation dataset; (G) PR curve for the HRS training dataset; (H) PR curve for the SHARE training dataset; (I) PR curve for the CHARLS training dataset; (J) PR curve for the HRS validation dataset; (K) PR curve for the SHARE validation dataset; (L) PR curve for the CHARLS validation dataset.

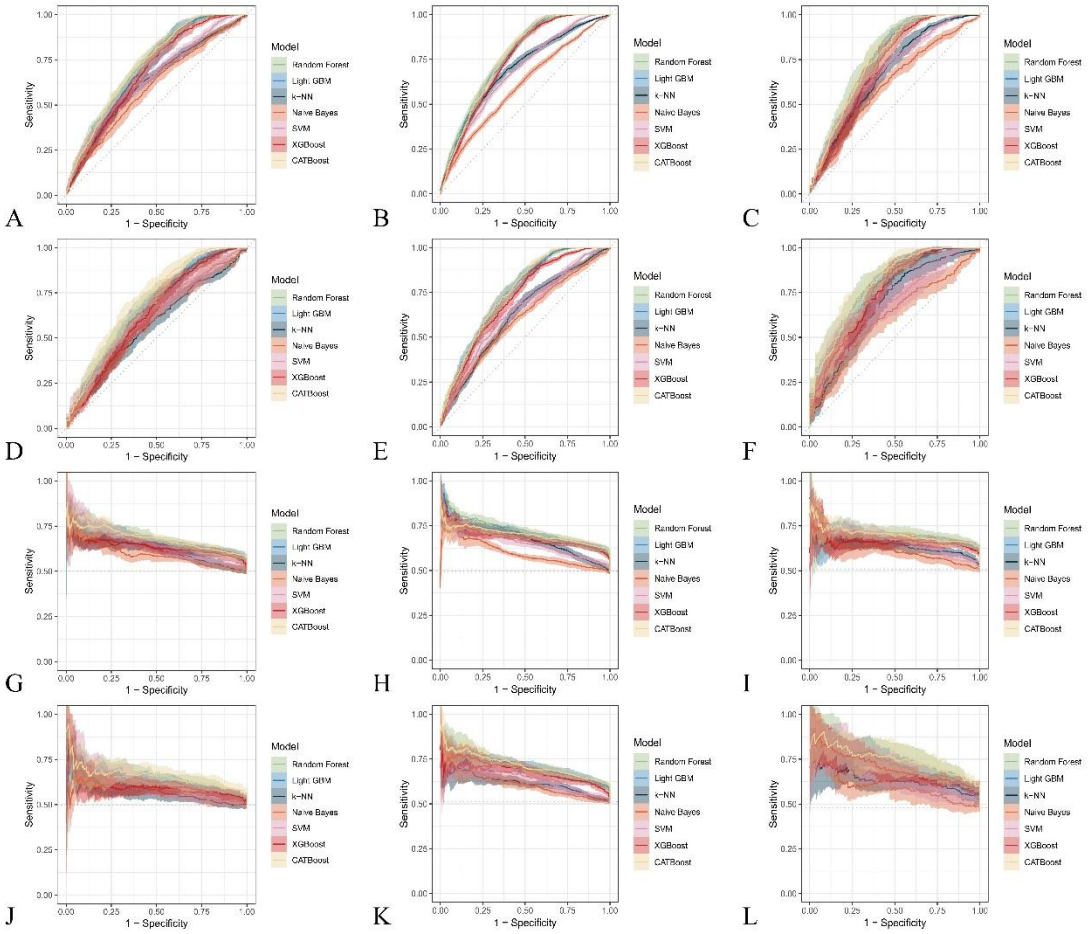


Figure 7 Calibration curves and decision curve analysis (DCA) evaluating the calibration performance and clinical utility of different machine learning models across the three cohorts. (A) Calibration curve for the HRS cohort; (B) calibration curve for the SHARE cohort; (C) calibration curve for the CHARLS cohort; (D) DCA curve for the HRS cohort; (E) DCA curve for the SHARE cohort; (F) DCA curve for the CHARLS cohort.

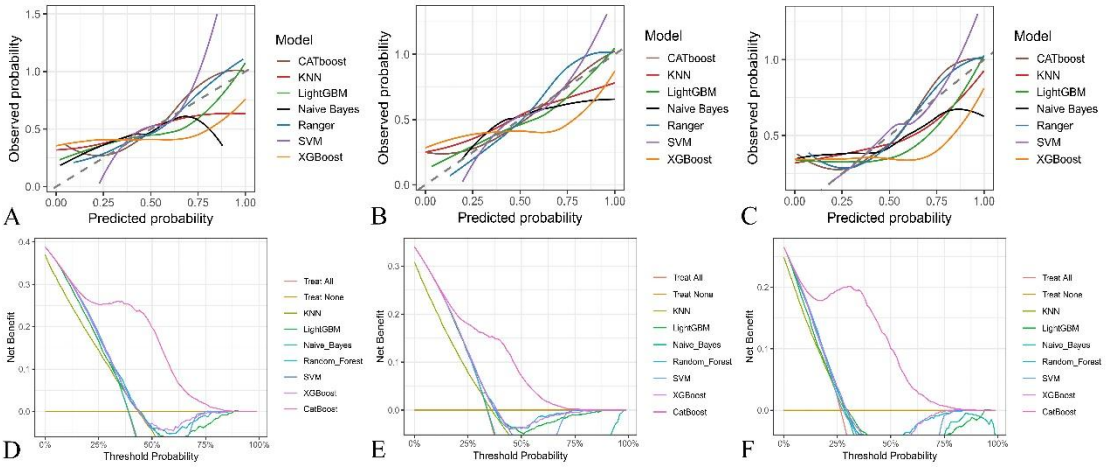


Figure 8 SHAP analysis interpreting the performance of the CatBoost model across different cohorts. (A) SHAP summary plot for the HRS cohort; (B) SHAP summary plot for the SHARE cohort; (C) SHAP summary plot for the CHARLS cohort; (D) force plot for the HRS cohort; (E) force plot for the SHARE cohort; (F) force plot for the CHARLS cohort; (G) waterfall plot for the HRS cohort; (H) waterfall plot for the SHARE cohort; (I) waterfall plot for the CHARLS cohort.

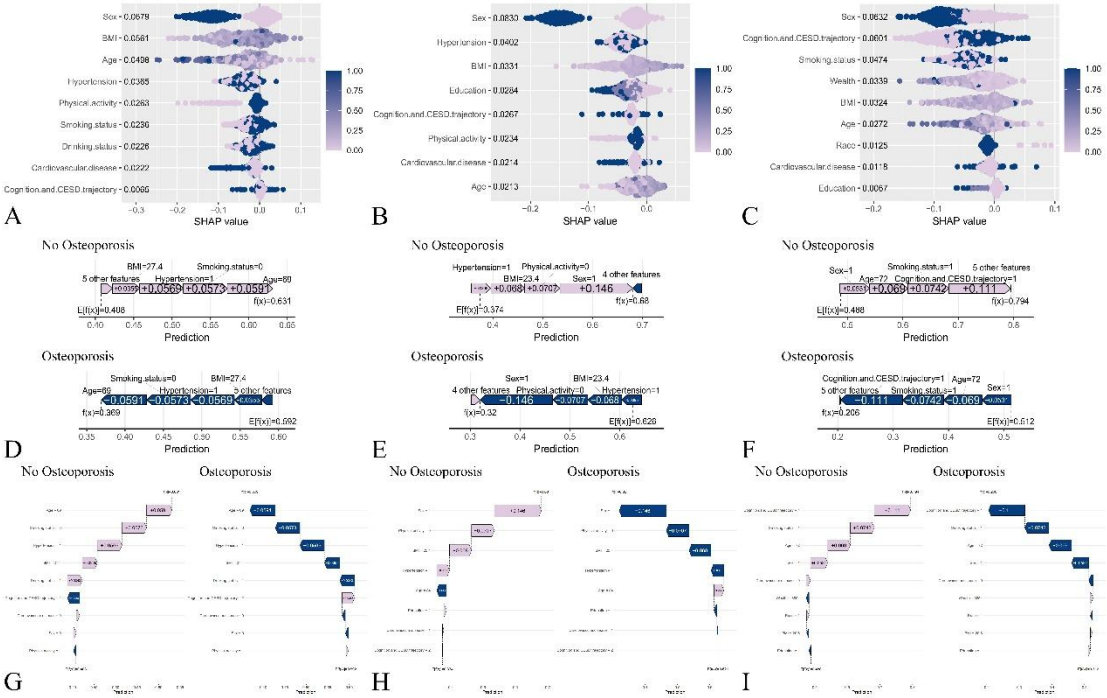


Figure 9 Associations between the joint trajectories of cognitive function and CES-D scores and the risk of incident osteoporosis among adults aged ≥65 years with diabetes across different age and sex subgroups in the three cohorts. (A) HRS; (B) SHARE; (C) CHARLS.

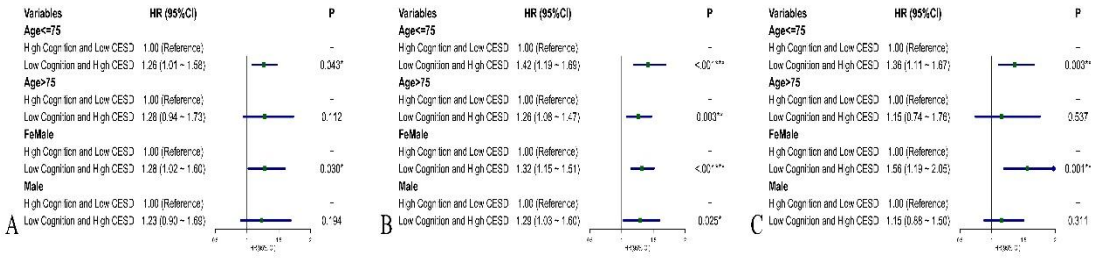


Table 1 Baseline characteristics of participants in the HRS cohort.

Characteristic	**N**	**Overall** N = 2,811	**Normal** N = 1,713	**New-onset Osteoporosis** N = 1,098	**p-value**
Age(year)[a], Mean ± SD	2,811	73.60 ± 6.87	74.12 ± 7.14	72.80 ± 6.34	<0.001
Sex [b], n(%)	2,811				<0.001
Female		1,467 (52.19%)	825 (48.16%)	642 (58.47%)	
Male		1,344 (47.81%)	888 (51.84%)	456 (41.53%)	
Race[b], n(%)	2,811				0.803
Non-Hispanic white		2,233 (79.44%)	1,367 (79.80%)	866 (78.87%)	
Non-Hispanic Blacks		290 (10.32%)	169 (9.87%)	121 (11.02%)	
Hispanic		227 (8.08%)	139 (8.11%)	88 (8.01%)	
Other Race		61 (2.17%)	38 (2.22%)	23 (2.09%)	
Education[b], n(%)	2,811				0.783
High school or less		638 (22.70%)	396 (23.12%)	242 (22.04%)	
Some college		1,520 (54.07%)	919 (53.65%)	601 (54.74%)	
College graduate or above		653 (23.23%)	398 (23.23%)	255 (23.22%)	
Wealth [a], Mean ±	2,811	575,589.98 ± 1,448,024.65	575,033.18 ± 1,393,862.52	576,458.66 ± 1,529,346.36	0.895
BMI [a], Mean ±	2,811	25.74 ± 4.38	25.41 ± 4.31	26.25 ± 4.46	<0.001
Hypertension	2,811				0.813
No		1,559 (55.46%)	947 (55.28%)	612 (55.74%)	
Yes		1,252 (44.54%)	766 (44.72%)	486 (44.26%)	
Smoking status[b], n(%)	2,811				0.781
No		1,253 (44.57%)	760 (44.37%)	493 (44.90%)	
Yes		1,558 (55.43%)	953 (55.63%)	605 (55.10%)	

Characteristic	N	Overall N = 2,811	Normal N = 1,713	New-onset Osteoporosis N = 1,098	p-value
Drinking status[b], n(%)	2,811				0.289
No		1,392 (49.52%)	862 (50.32%)	530 (48.27%)	
Yes		1,419 (50.48%)	851 (49.68%)	568 (51.73%)	
Cancer [b], n(%)	2,811				0.501
No		2,401 (85.41%)	1,457 (85.06%)	944 (85.97%)	
Yes		410 (14.59%)	256 (14.94%)	154 (14.03%)	
Cardiovascular disease[b], n(%)	2,811				0.501
No		2,134 (75.92%)	1,293 (75.48%)	841 (76.59%)	
Yes		677 (24.08%)	420 (24.52%)	257 (23.41%)	
Physical activity [b], n(%)	2,811				0.013
No		573 (20.38%)	375 (21.89%)	198 (18.03%)	
Yes		2,238 (79.62%)	1,338 (78.11%)	900 (81.97%)	
Cognition and CESD trajectory	2,811				0.310
High Cognition and Low CESD		2,447 (87.05%)	1,500 (87.57%)	947 (86.25%)	
Low Cognition and High CESD		364 (12.95%)	213 (12.43%)	151 (13.75%)	

a: Student t-test, b: Chi-square test, SD: standard deviation

Table 2 Baseline characteristics of participants in the SHARE cohort.

Characteristic	N	Overall N = 5,911	Normal N = 3,902	New-onset Osteoporosis N = 2,009	p-value
Sex[b], n(%)	5,911				<0.001
FeMale		3,185 (53.88%)	1,905 (48.82%)	1,280 (63.71%)	
Male		2,726 (46.12%)	1,997 (51.18%)	729 (36.29%)	
Age(year)[a], Mean ± SD	5,911	72.93 ± 6.08	72.69 ± 6.02	73.40 ± 6.15	<0.001
Native [b], n(%)	5,911				0.659
No		287 (4.86%)	186 (4.77%)	101 (5.03%)	
Yes		5,624 (95.14%)	3,716 (95.23%)	1,908 (94.97%)	

Characteristic	N	Overall N = 5,911	Normal N = 3,902	New-onset Osteoporosis N = 2,009	p-value
Education[b], n(%)	5,911				<0.001
Before high school		2,338 (39.55%)	1,444 (37.01%)	894 (44.50%)	
High school and vocational training		2,391 (40.45%)	1,629 (41.75%)	762 (37.93%)	
Higher Education		1,182 (20.00%)	829 (21.25%)	353 (17.57%)	
Wealth [a], Mean ±SD	5,911	402,795.88 ± 800,303.59	411,719.80 ± 795,093.72	385,463.33 ± 810,245.08	<0.001
BMI [a], Mean ±SD	5,911	26.63 ± 4.33	26.41 ± 4.25	27.06 ± 4.46	<0.001
Hypertension [b], n(%)	5,911				<0.001
No		3,263 (55.20%)	2,219 (56.87%)	1,044 (51.97%)	
Yes		2,648 (44.80%)	1,683 (43.13%)	965 (48.03%)	
Smoking status[b], n(%)	5,911				<0.001
No		3,667 (62.04%)	2,338 (59.92%)	1,329 (66.15%)	
Yes		2,244 (37.96%)	1,564 (40.08%)	680 (33.85%)	
Drinking status [b], n(%)	5,911				<0.001
No		969 (16.39%)	594 (15.22%)	375 (18.67%)	
Yes		4,942 (83.61%)	3,308 (84.78%)	1,634 (81.33%)	
Cancer [b], n(%)	5,911				0.392
No		5,451 (92.22%)	3,590 (92.00%)	1,861 (92.63%)	
Yes		460 (7.78%)	312 (8.00%)	148 (7.37%)	
Physical activity [b], n(%)	5,911				<0.001
No		953 (16.12%)	564 (14.45%)	389 (19.36%)	
Yes		4,958 (83.88%)	3,338 (85.55%)	1,620 (80.64%)	
Cardiovascular disease [b], n(%)	5,911				0.003
No		4,644 (78.57%)	3,110 (79.70%)	1,534 (76.36%)	
Yes		1,267 (21.43%)	792 (20.30%)	475 (23.64%)	
Cognition and CESD trajectory [b], n(%)	5,911				<0.001

Characteristic	N	Overall N = 5,911	Normal N = 3,902	New-onset Osteoporosis N = 2,009	p-value
High Cognition and Low CESD		4,949 (83.73%)	3,360 (86.11%)	1,589 (79.09%)	
Low Cognition and High CESD		962 (16.27%)	542 (13.89%)	420 (20.91%)	

a: Student t-test, b: Chi-square test, SD: standard deviation

Table 3 Baseline characteristics of participants in the CHARLS cohort.

Characteristic	N	Overall N = 2,088	Normal N = 1,534	New-onset Osteoporosis N = 554	p-value
Age(year)[a], Mean ± SD	2,088	71.92 ± 5.57	72.09 ± 5.58	71.47 ± 5.51	0.012
Sex [b], n(%)	2,088				<0.001
Female		908 (43.49%)	617 (40.22%)	291 (52.53%)	
Male		1,180 (56.51%)	917 (59.78%)	263 (47.47%)	
Race [b], n(%)	2,088				0.014
Other		106 (5.08%)	67 (4.37%)	39 (7.04%)	
Han Chinese		1,982 (94.92%)	1,467 (95.63%)	515 (92.96%)	
Education[b], n(%)	2,088				0.019
Before high school		1,908 (91.38%)	1,386 (90.35%)	522 (94.22%)	
High school and vocational training		123 (5.89%)	100 (6.52%)	23 (4.15%)	
Higher Education		57 (2.73%)	48 (3.13%)	9 (1.62%)	
Wealth [a], Mean ±	2,088	20,735.37 ± 30,359.80	22,372.41 ± 32,038.85	16,202.49 ± 24,595.24	<0.001
BMI [a], Mean ±	2,088	23.04 ± 36.38	23.31 ± 42.38	22.30 ± 3.88	0.633
Hypertension	2,088				0.870
No		1,453 (69.59%)	1,069 (69.69%)	384 (69.31%)	
Yes		635 (30.41%)	465 (30.31%)	170 (30.69%)	
Smoking status[b], n(%)	2,088				0.062
No		1,105 (52.92%)	793 (51.69%)	312 (56.32%)	
Yes		983 (47.08%)	741 (48.31%)	242 (43.68%)	

Characteristic	N	Overall N = 2,088	Normal N = 1,534	New-onset Osteoporosis N = 554	p-value
Drinking status [b], n(%)	2,088				0.124
No		1,239 (59.34%)	895 (58.34%)	344 (62.09%)	
Yes		849 (40.66%)	639 (41.66%)	210 (37.91%)	
Cancer	2,088				0.583
No		2,071 (99.19%)	1,520 (99.09%)	551 (99.46%)	
Yes		17 (0.81%)	14 (0.91%)	3 (0.54%)	
Physical activity [b], n(%)	2,088				0.764
No		1,731 (82.90%)	1,274 (83.05%)	457 (82.49%)	
Yes		357 (17.10%)	260 (16.95%)	97 (17.51%)	
Cardiovascular disease[b], n(%)	2,088				0.713
No		1,759 (84.24%)	1,295 (84.42%)	464 (83.75%)	
Yes		329 (15.76%)	239 (15.58%)	90 (16.25%)	
Cognition and CESD trajectory [b], n(%)	2,088				<0.001
High Cognition and Low CESD		1,106 (52.97%)	857 (55.87%)	249 (44.95%)	
Low Cognition and High CESD		982 (47.03%)	677 (44.13%)	305 (55.05%)	

a: Student t-test, b: Chi-square test, SD: standard deviation

Table 4 Cox regression analyses of the associations between the joint trajectories of cognitive function and CES-D scores and the risk of incident osteoporosis among adults aged ≥65 years with diabetes.

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
HRS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.35 (1.14 ~ 1.61)	<.001***	1.26 (1.06 ~ 1.50)	0.010**	1.24 (1.04 ~ 1.48)	0.016*	1.26 (1.05 ~ 1.51)	0.012*
SHARE								
Cognition and CESD trajectory								

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.64 (1.47 ~ 1.83)	<.001***	1.46 (1.31 ~ 1.64)	<.001***	1.45 (1.29 ~ 1.62)	<.001***	1.39 (1.24 ~ 1.56)	<.001***
CHARLS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.54 (1.30 ~ 1.82)	<.001***	1.35 (1.12 ~ 1.62)	0.002**	1.34 (1.12 ~ 1.62)	0.002**	1.34 (1.12 ~ 1.62)	0.002**

Model 1: Crude

Model 2: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI

Model 3: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status

Model 4: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status, Cardiovascular disease, Physical activity,Ccancer

Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Supplementary Table 1 Model fit indices for GBTM trajectory selection across different cohorts.

Model	AIC	BIC	Log-Likelihood
HRS			
1-class	80,753.56	80,824.85	-40,364.78
2-class	66,297.36	66,398.36	-33,131.68
3-class	65,843.08	65,973.79	-32,899.54
4-class	65,299.59	65,460.00	-32,622.79
Share			
1-class	195,475.5	195,555.7	-97,725.73
2-class	160,612.8	160,726.5	-80,289.40
3-class	160,131.6	160,278.7	-80,043.81
4-class	160,129.7	160,310.2	-80,037.86
Charls			
1-class	90,470.39	90,538.12	-45,223.19
2-class	74,252.94	74,348.88	-37,109.47

Model	AIC	BIC	Log-Likelihood
3-class	74,023.14	74,147.31	-36,989.57
4-class	73,914.56	74,066.95	-36,930.28

Supplementary Table 2 Benchmarking results for the training dataset in the HRS cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.647 (0.626-0.668)	0.666 (0.648-0.684)	0.720 (0.696-0.743)	0.704 (0.682-0.727)	0.590 (0.555-0.625)	0.680 (0.646-0.714)	0.210 (0.203-0.217)
Light GBM	0.635 (0.612-0.658)	0.649 (0.628-0.670)	0.701 (0.670-0.733)	0.674 (0.650-0.699)	0.596 (0.548-0.645)	0.659 (0.627-0.692)	0.221 (0.208-0.234)
K-KNN	0.596 (0.579-0.612)	0.597 (0.586-0.607)	0.626 (0.609-0.643)	0.597 (0.580-0.614)	0.596 (0.555-0.637)	0.617 (0.590-0.643)	0.262 (0.255-0.270)
Naive Bayes	0.568 (0.542-0.594)	0.535 (0.503-0.568)	0.597 (0.572-0.623)	0.498 (0.460-0.537)	0.638 (0.616-0.661)	0.598 (0.576-0.619)	0.247 (0.240-0.253)
SVM	0.597 (0.573-0.620)	0.590 (0.570-0.610)	0.649 (0.628-0.671)	0.580 (0.554-0.607)	0.614 (0.577-0.652)	0.652 (0.627-0.677)	0.234 (0.229-0.238)
XGBoost	0.623 (0.603-0.643)	0.630 (0.607-0.653)	0.679 (0.652-0.706)	0.642 (0.614-0.670)	0.605 (0.574-0.635)	0.636 (0.604-0.669)	0.302 (0.283-0.320)
CATBoost	0.650 (0.635-0.666)	0.672 (0.657-0.686)	0.721 (0.702-0.740)	0.716 (0.687-0.744)	0.588 (0.544-0.632)	0.680 (0.651-0.708)	0.210 (0.206-0.214)
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 3 Benchmarking results for the validation dataset in the HRS cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.634 (0.589-0.679)	0.648 (0.604-0.693)	0.669 (0.620-0.717)	0.678 (0.642-0.715)	0.590 (0.525-0.654)	0.638 (0.579-0.697)	0.228 (0.214-0.242)

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Light GBM	0.623 (0.598-0.647)	0.626 (0.597-0.654)	0.652 (0.618-0.686)	0.636 (0.607-0.665)	0.607 (0.570-0.645)	0.614 (0.554-0.674)	0.253 (0.237-0.269)
K-KNN	0.557 (0.523-0.591)	0.554 (0.514-0.594)	0.576 (0.540-0.613)	0.556 (0.511-0.601)	0.557 (0.508-0.607)	0.582 (0.541-0.624)	0.281 (0.264-0.297)
Naive Bayes	0.563 (0.534-0.592)	0.535 (0.500-0.571)	0.594 (0.557-0.631)	0.509 (0.461-0.557)	0.619 (0.565-0.673)	0.573 (0.526-0.620)	0.250 (0.238-0.261)
SVM	0.583 (0.553-0.612)	0.582 (0.549-0.615)	0.620 (0.577-0.664)	0.585 (0.553-0.617)	0.580 (0.527-0.633)	0.622 (0.577-0.668)	0.240 (0.234-0.247)
XGBoost	0.593 (0.555-0.632)	0.600 (0.558-0.643)	0.633 (0.603-0.662)	0.617 (0.566-0.669)	0.568 (0.525-0.612)	0.603 (0.558-0.647)	0.335 (0.311-0.359)
CATBoost	0.635 (0.588-0.683)	0.657 (0.605-0.708)	0.699 (0.653-0.745)	0.705 (0.657-0.752)	0.566 (0.507-0.625)	0.656 (0.602-0.710)	0.218 (0.209-0.227)
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 4 Benchmarking results for the training dataset in the SHARE cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.694 (0.680-0.709)	0.723 (0.707-0.738)	0.768 (0.753-0.783)	0.806 (0.789-0.822)	0.586 (0.562-0.611)	0.724 (0.700-0.747)	0.195 (0.190-0.200)
Light GBM	0.686 (0.673-0.699)	0.712 (0.696-0.727)	0.759 (0.746-0.771)	0.785 (0.769-0.801)	0.589 (0.569-0.609)	0.708 (0.685-0.731)	0.191 (0.184-0.197)
K-KNN	0.642 (0.626-0.658)	0.647 (0.629-0.665)	0.690 (0.672-0.709)	0.664 (0.641-0.688)	0.621 (0.595-0.648)	0.664 (0.640-0.688)	0.239 (0.230-0.247)
Naive Bayes	0.568 (0.555-0.581)	0.587 (0.571-0.602)	0.602 (0.590-0.615)	0.620 (0.605-0.636)	0.518 (0.498-0.538)	0.595 (0.573-0.617)	0.248 (0.245-0.252)

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
SVM	0.628 (0.620-0.637)	0.634 (0.621-0.648)	0.680 (0.667-0.693)	0.652 (0.633-0.672)	0.605 (0.592-0.619)	0.651 (0.625-0.677)	0.225 (0.222-0.229)
XGBoost	0.668 (0.660-0.675)	0.672 (0.659-0.686)	0.741 (0.732-0.750)	0.690 (0.672-0.708)	0.646 (0.630-0.662)	0.687 (0.663-0.711)	0.248 (0.242-0.253)
CATBoost	0.699 (0.679-0.719)	0.735 (0.715-0.755)	0.761 (0.745-0.777)	0.844 (0.830-0.859)	0.557 (0.529-0.585)	0.699 (0.675-0.724)	0.188 (0.181-0.194)
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 5 Benchmarking results for the validation dataset in the SHARE cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.681 (0.667-0.696)	0.713 (0.695-0.730)	0.743 (0.725-0.761)	0.775 (0.755-0.795)	0.584 (0.566-0.601)	0.712 (0.675-0.749)	0.203 (0.198-0.209)
Light GBM	0.679 (0.651-0.707)	0.702 (0.672-0.732)	0.742 (0.720-0.765)	0.742 (0.714-0.771)	0.612 (0.580-0.645)	0.703 (0.668-0.737)	0.204 (0.193-0.215)
K-KNN	0.605 (0.584-0.627)	0.628 (0.606-0.649)	0.628 (0.608-0.647)	0.652 (0.629-0.676)	0.557 (0.532-0.581)	0.617 (0.587-0.647)	0.264 (0.255-0.274)
Naive Bayes	0.571 (0.551-0.591)	0.595 (0.573-0.617)	0.605 (0.578-0.631)	0.618 (0.598-0.639)	0.522 (0.492-0.552)	0.614 (0.575-0.652)	0.250 (0.240-0.259)
SVM	0.612 (0.587-0.637)	0.634 (0.605-0.664)	0.648 (0.618-0.678)	0.660 (0.628-0.691)	0.562 (0.533-0.591)	0.639 (0.596-0.683)	0.234 (0.228-0.241)
XGBoost	0.647 (0.627-0.668)	0.661 (0.639-0.682)	0.716 (0.691-0.740)	0.673 (0.647-0.699)	0.621 (0.579-0.664)	0.685 (0.645-0.725)	0.275 (0.260-0.290)
CATBoost	0.682 (0.665-0.699)	0.717 (0.696-0.737)	0.749 (0.734-0.764)	0.790 (0.767-0.814)	0.568 (0.549-0.587)	0.717 (0.687-0.747)	0.197 (0.190-0.204)

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 6 Benchmarking results for the training dataset in the CHARLS cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.713 (0.686-0.741)	0.747 (0.725-0.769)	0.757 (0.716-0.799)	0.832 (0.815-0.850)	0.588 (0.536-0.641)	0.696 (0.646-0.746)	0.189 (0.174-0.203)
Light GBM	0.682 (0.658-0.707)	0.703 (0.686-0.721)	0.731 (0.692-0.771)	0.741 (0.723-0.759)	0.621 (0.577-0.664)	0.676 (0.635-0.716)	0.212 (0.194-0.229)
K-KNN	0.626 (0.606-0.646)	0.645 (0.621-0.670)	0.674 (0.642-0.707)	0.671 (0.639-0.703)	0.578 (0.542-0.613)	0.645 (0.610-0.681)	0.244 (0.228-0.261)
Naive Bayes	0.561 (0.529-0.592)	0.459 (0.404-0.514)	0.613 (0.585-0.641)	0.372 (0.313-0.430)	0.755 (0.720-0.790)	0.614 (0.571-0.658)	0.268 (0.254-0.281)
SVM	0.636 (0.609-0.664)	0.662 (0.636-0.689)	0.688 (0.665-0.711)	0.703 (0.672-0.733)	0.569 (0.526-0.611)	0.663 (0.626-0.700)	0.223 (0.216-0.229)
XGBoost	0.659 (0.627-0.691)	0.673 (0.644-0.702)	0.714 (0.673-0.755)	0.692 (0.653-0.730)	0.625 (0.581-0.668)	0.662 (0.619-0.704)	0.280 (0.255-0.305)
CATBoost	0.703 (0.668-0.738)	0.745 (0.718-0.773)	0.751 (0.719-0.784)	0.855 (0.816-0.893)	0.546 (0.493-0.599)	0.701 (0.668-0.734)	0.192 (0.180-0.204)
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 7 Benchmarking results for the validation dataset in the CHARLS cohort.

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
Random Forest	0.707 (0.666-	0.706 (0.656-	0.794 (0.743-	0.744 (0.682-	0.673 (0.605-	0.762 (0.689-	0.187 (0.168-

Model	Accuracy	F Beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve	Brier
	0.748)	0.756)	0.845)	0.806)	0.742)	0.835)	0.205)
Light GBM	0.687 (0.658-0.716)	0.672 (0.631-0.714)	0.772 (0.734-0.810)	0.680 (0.610-0.751)	0.694 (0.636-0.751)	0.725 (0.665-0.786)	0.210 (0.191-0.229)
K-KNN	0.636 (0.594-0.678)	0.629 (0.588-0.669)	0.685 (0.638-0.732)	0.644 (0.604-0.683)	0.630 (0.571-0.689)	0.632 (0.576-0.688)	0.239 (0.215-0.263)
Naive Bayes	0.552 (0.524-0.579)	0.217 (0.174-0.260)	0.613 (0.552-0.674)	0.131 (0.101-0.161)	0.940 (0.918-0.962)	0.604 (0.535-0.674)	0.396 (0.358-0.435)
SVM	0.648 (0.592-0.705)	0.597 (0.522-0.672)	0.694 (0.619-0.770)	0.549 (0.482-0.616)	0.742 (0.665-0.818)	0.680 (0.575-0.785)	0.221 (0.202-0.241)
XGBoost	0.657 (0.633-0.682)	0.643 (0.621-0.666)	0.749 (0.708-0.790)	0.649 (0.605-0.692)	0.663 (0.595-0.731)	0.702 (0.640-0.765)	0.274 (0.250-0.297)
CATBoost	0.710 (0.659-0.761)	0.694 (0.621-0.767)	0.795 (0.740-0.850)	0.704 (0.620-0.787)	0.714 (0.656-0.772)	0.747 (0.662-0.831)	0.187 (0.170-0.204)
P	<.001[a]	<.001[a]	<.001[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]

[a]ANOVA test; [b]Kruskal-Wallis

Supplementary Table 8 Fine–Gray competing risk regression analyses examining the associations between the joint trajectories of cognitive function and CES-D scores and the risk of incident osteoporosis among adults aged ≥65 years with diabetes after accounting for death as a competing event.

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value
HRS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.33 (1.14 ~ 1.56)	<.001***	1.25 (1.06 ~ 1.47)	0.008**	1.23 (1.04 ~ 1.45)	0.014*	1.24 (1.05 ~ 1.47)	0.011*
ELSA								

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.56 (1.42 ~ 1.71)	<.001***	1.32 (1.20 ~ 1.46)	<.001***	1.32 (1.19 ~ 1.45)	<.001***	1.27 (1.15 ~ 1.40)	<.001***
CHARLS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.49 (1.27 ~ 1.74)	<.001***	1.31 (1.11 ~ 1.56)	0.002**	1.31 (1.11 ~ 1.55)	0.002**	1.31 (1.11 ~ 1.55)	0.002**

Model 1: Crude
Model 2: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI
Model 3: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status
Model 4: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status, Cardiovascular disease, Physical activity,Ccancer
Significance: * p < 0.05, ** p < 0.01, *** p < 0.001

Supplementary Table 9 Discrete-time logistic regression analyses examining the associations between the joint trajectories of cognitive function and CES-D scores and the risk of incident osteoporosis under different time-setting approaches among adults aged ≥65 years with diabetes.

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value
HRS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.33 (1.14 ~ 1.56)	<.001***	1.25 (1.06 ~ 1.47)	0.008**	1.23 (1.04 ~ 1.45)	0.014*	1.24 (1.05 ~ 1.47)	0.011*
ELSA								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition	1.56 (1.42 ~	<.001***	1.32 (1.20 ~	<.001*	1.32 (1.19 ~	<.001**	1.27 (1.15 ~	<.001**

Exposure Variable	Model 1		Model 2		Model 3		Model 4	
	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value	sHR (95% CI)	P-value
and High CESD	1.71)		1.46)	**	1.45)	*	1.40)	*
CHARLS								
Cognition and CESD trajectory								
High Cognition and Low CESD	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-	1.00 (Reference)	-
Low Cognition and High CESD	1.49 (1.27 ~ 1.74)	<.001***	1.31 (1.11 ~ 1.56)	0.002**	1.31 (1.11 ~ 1.55)	0.002**	1.31 (1.11 ~ 1.55)	0.002**

Model 1: Crude
Model 2: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI
Model 3: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status
Model 4: Adjust: Age, Sex, Race, Education, marital status, Wealth, BMI, Hypertension, Drinking status, Smoking status, Cardiovascular disease, Physical activity,Ccancer
Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx