Supplementary 3

Methods

Anthropometric measurements included systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR) and BMI by means of standardized measurements. Peripheral blood samples from each participant were collected through a standard phlebotomy protocol by personnel who received medical training. The biochemical parameters included TG, low-density lipoprotein cholesterol (LDL-C), HDL-C, total cholesterol (TC), FPG and glycosylated hemoglobin, type A1c (HbA1c). Other laboratory examinations included creatinine (Crea), uric acid (UA) and hemoglobin (HGB). From 2011 to 2020, five waves of statistics were recorded in CHARLS.

There were definitions of other covariates in this research. BMI was equal to weight (in kilograms) divided by height (in meters) squared. Marital status, smoking history, alcohol consumption history, cancer status and chronic kidney disease status were determined via self-reported health assessments.

To compare the baseline characteristics of the participants, normally distributed continuous variables were summarized as the means ± standard deviations, and nonnormally distributed continuous variables were descriptively presented as medians with interquartile ranges (IQRs). Analysis of variance (ANOVA) or the Kruskal–Wallis H test was used to compare the continuous variables when appropriate. Categorical variables are described as percentages and frequencies and were tested by the chi-square test or Fisher’s exact test.

Univariate and multivariate Cox proportional hazard regression models were used to estimate the associations between METS-IR and new-onset stroke in individuals within individuals with CKM syndrome stages 0-3. Controlled covariates were risk factors important to stroke and were selected in the utilization of three models. Specifically, Model 1 served as an unadjusted model. Model 2 included three adjustments for age, sex, and marital status. In Model 3, adjustments for smoking history, alcohol consumption history, hypertension, diabetes, heart problems, SBP, DBP, HR, LDL-C, CHO, Crea, UA and HGB were made. To evaluate the robustness of the association between METS-IR and future stroke risk, three additional sensitivity analyses were performed to confirm our findings. First, we excluded participants with a diagnosis of hypertension. Second, we excluded participants with a history of heart problems. Third, participants with a BMI<24 were excluded. The hazard ratio (HR) and 95% confidence interval (CI) were calculated via METS-IR quartiles, and we set the quartile 1 group as the reference in all models. Furthermore, subgroup analysis was performed via stratified Cox proportional hazards regression models to identify potential impact factors and evaluate the consistency of the associations between the METS-IR score and future stroke risk. Age (≥ 60 or <60), sex (male or female), current marital status (yes and no), smoking status (current/ever smoker and never smoker) and drinking status (current/ever alcohol drinking and never alcohol drinking) were included.

Kaplan–Meier (KM) curves can show different survival patterns across tow cohorts, and we used KM curves to reveal the cumulative incidence of stroke events in the four METS-IR quartiles. A log-rank test was then used to compare the differences within individuals with CKM syndrome stages 0-3, and the P value was recorded. In addition, considering the potential nonlinear relationship between METS-IR and future stroke risk, we employed restricted cubic splines (RCSs) and fitted RCS curves of different target groups to determine the linear trends or other shapes of the correlation; the number of knots was four. For the data of the 3-year transition of METS-IR, we carried out the strategy of K-means clustering, which is a technique that is able to divide N observations into K clusters. In this research, the CHARLS cohort of 4968 participants was divided into five classes, and the incidence of stroke in each class was compared. The odds ratio and 95% CI were calculated via logistic regression analyses to focus on the relevance of changes in METS-IR scores to future stroke risk.

In this study, to elucidate the predictive ability of METS-IR, we built new-onset stroke assessment machine learning models using 3 algorithms: Logistic Regression (LR), Extreme Gradient Boosting (XGBoost) and Random Forest (RF). The original dataset was divided into a training subset and a testing subset, with 70% allocated to the training set and 30% to the testing set. To assess three models' performance, we applied ROC curves and calculated accuracy, precision, recall and F1 score. Further details about three algorithms are available on supplementary materials.

Results

General characteristics of the study participants

Table 1 shows that the clinical and demographic baseline characteristics of the CHARLS cohort in individuals diagnosed with CKM syndrome stages 0–3 (n = 9516). **Table S1** shows the baseline characteristics of the NHANES cohort study participants in individuals diagnosed with CKM syndrome stages 0–3 (n = 5478) stratified by METS-IR quartiles. Participants from CHARLS were stratified according to the METS-IR quartiles: quartile 1 (Q1), METS-IR<29.62; quartile 2 (Q2), 29.62≤METS-IR≤34.3; quartile 3 (Q3), 34.31≤METS-IR≤40.29; and quartile 4 (Q4), METS-IR>40.29. In our study of 9516 eligible participants, the median age was 59.04±9.69 years, 5174 (54.4%) participants were males, and the average METS-IR was 35.6±8.52. Compared with those in the lower quartile of the METS-IR, the prevalence of hypertension, diabetes and heart disease were more likely to increase in the higher quartile. Moreover, parameters such as SBP, DBP, FPG, TG, TC, BMI and HbA1c gradually increased among participants in higher quartiles of METS-IR, whereas HDL-C levels were lower in those groups. In addition, the proportions of smokers and alcohol consumers tended to increase with increasing METS-IR, and the differences in Crea, UA, and HGB were not significant among the four METS-IR quartiles.

Associations between METS-IR and stroke incidence

During the CHARLS follow-up period from 2011 to 2020, 849 (8.9%) participants in individuals diagnosed with CKM syndrome stages 0–3 developed their first incident case of stroke. Baseline METS-IR score was simultaneously presented as a continuous variable and a categorical variable (quartiles). As a categorical variable, stroke incidence rates for participants in the METS-IR quartiles were as follows: Q1: 142/2379 (6%); Q2: 187/2379 (7.9%); Q3: 220/2379 (9.2%); and Q4: 300/2379 (12.6%). The incidence of new-onset stroke gradually increases as the METS-IR increases. The results of the Cox proportional hazard model analysis revealed that with respect to Q1, the HR of Q2 was 1.421 (95% CI 1.131–1.786; P=0.003), the HR of Q3 was 1.723 (95% CI 1.452–2.012; P<0.001), and the HR of Q4 was 1.973 (95% CI 1.655–2.299; P<0.001). This indicates that when Q1 was used as the reference group, the risks of future stroke for Q2, Q3 and Q4 were 1.421-fold greater, 1.723-fold greater and 1.973-fold greater, respectively. Similarly, Model 2 adjusts for age, sex, and marital status, with Q1 as the reference group; Q2 has an HR of 1.344 (95% CI 1.072-1.685; P=0.01); Q3 has an HR of 1.657 (95% CI 1.396-1.935; P<0.001); and Q4 has an HR of 1.937 (95% CI 1.624-2.257; P<0.001). In Model 3, which was based on the full adjustment, there was also a strong association between METS-IR scores and future stroke risk (Q2 HR=1.215, 95% CI: 1.076-1.373, p=0.002; Q3 HR=1.594, 95% CI: 1.42-1.789, p<0.001; Q4 HR=1.894, 95% CI: 1.42-1.789, p<0.001). When treated as a continuous variable, METS-IR was linked with the incidence of stroke [Model 1: HR (95% CI) 1.030 (1.023-1.028), p<0.001; Model 2: HR (95% CI) 1.036 (1.028-1.043), p<0.001; Model 3: HR (95% CI) 1.026 (1.017-1.036), p<0.001]. In summary, in all three Cox regression models we fitted, the METS-IR score was significantly correlated with future stroke risk (**Table 2**).

Incidence patterns and the predictive value of METS-IR

The KM curves revealed that the increased future stroke risk coincides with the increasing METS-IR quartiles, and the highest risk of new-onset stroke was manifested among participants with the highest METS-IR quartile in CHARLS cohort with CKM syndrome stages 0–3(p<0.001) (**Fig. S1**) and NHANES cohort with CKM syndrome stages 0–3(p<0.001) (**Fig. S2**). Nonlinear relationships were observed between increased METS-IR and future stroke risk in model 1 through the application of RCS analysis (p for nonlinearity<0.05). In addition, METS-IR showed a consistent linear association with new-onset stroke in Model 2 (p for nonlinearity=0.472) and Model 3 (p for nonlinearity=0.113) of the total cohort in the plotted RCS curves (**Fig. 2A-C**). For NHANES database cohort, the results of RCS curves gained similar conclusions (**Fig. S3**).

Sensitivity analyses and subgroup analyses

Three sensitivity analyses were conducted to further elucidate the association between

METS-IR and future stroke risk in individuals diagnosed with CKM syndrome stages 0–3. In this study, several Cox regression models were used in a group without hypertension (n=5018), a group without a history of heart disease (n=8407) and participants whose BMI was less than 24 (n=5638). The analysis results revealed that the relationships between METS-IR and new-onset stroke remained significant. Our findings were adjusted for factors (age, sex, marital status, smoking history, alcohol consumption history, hypertension, diabetes, heart problems, SBP, DBP, HR, LDL-C, CHO, Crea, UA and HGB) in three models, and the robustness and consistency of the associations between stroke risk and high METS-IR quartiles were verified (**Table 3**). Considering that some potential risk factors might have latency interactions with the effect of METS-IR on future stroke risk in individuals diagnosed with CKM syndrome stages 0–3, a subgroup analysis stratified according to age, sex, current marital status, smoking status and drinking status was performed in our study. The results indicated that the interaction effects between age and METS-IR quartiles were statistically significant (p for interaction<0.01). Moreover, no pronounced interactions between METS-IR and other variables were discerned in any subgroup (gender: p for interaction=0.363; current married: p for interaction=0.066; smoking status: p for interaction=0.618; drinking status: p for interaction=0.854). Thus, on the basis of the results of the subgroup interaction analysis, the robustness of the associations between high METS-IR quartiles and future stroke risk reported in primary studies was examined (**Table S4**).

Three-year transition of the METS-IR

A total of 4968 participants diagnosed with CKM syndrome stages 0–3 with known 3-year transition data of the METS-IR were enrolled in the trial. In this analysis, when the cohort was divided into five groups, the effect of k-means clustering was superior to that of other options. The following classifications were designed: Class 1, the METS-IR index ranged from 28.2 in 2012 to 24.77 in 2015; Class 1, the METS-IR index ranged from 28.2 in 2012 to 24.77 in 2015, n=1492; Class 2, the METS-IR index ranged from 33.84 in 2012 to 34.92 in 2015, n=1897; Class 3, the METS-IR index ranged from 41.72 in 2012 to 43.57 in 2015, n=1152; Class 4, the METS-IR index ranged from 51.38 in 2012 to 54.4 in 2015, n=393; and Class 5, the METS-IR index ranged from 75.26 in 2012 to 77.3 in 2015, n=34. From Class 1 to Class 5, the basic levels of METS-IR increased and blood glucose was poorly controlled, which was accompanied by a higher rate of new-onset stroke. The proportions of Class 5, Class 4 and Class 3 among patients with stroke were higher than those in non-stroke participants (**Fig. S4**).

The logistic regression analyses revealed a significant association between the 3-year transition of METS-IR and future stroke risk. Compared with Class 1, which was set to the reference group, the risk of new-onset stroke increased gradually with the 3-year transition of the METS-IR from Class 2 to Class 3 (Class 2: OR: 1.236, 95% CI: 1.126-1.449, p=0.015; Class 3: OR: 1.866, 95% CI: 1.417-2.457, p<0.001). Furthermore, Class 5 was both the lowest performing group in terms of adequate METS-IR control and at the highest risk of new-onset stroke (OR: 11.48, 95% CI: 5.656-23.301, p<0.001), followed by Class 4 (OR: 2.634, 95% CI: 1.360-5.101, p=0.004). A sustained high METS-IR was observed in conjunction with high future stroke risk (**Table S5**). Essentially, the resulting associations were harmonized with the analysis results of the baseline METS-IR.

Machine learning algorithm confirm the link between METS-IR and stroke risk

Based on statistical analysis, we observed that increased METS-IR is associated with higher susceptibility to stroke in individuals diagnosed with CKM syndrome stages 0–3. To further validate this observation, we used three machine learning models (LR, XGBoost and RF) for additional analysis. The results showed that the AUC of all three models were greater than 0.6, indicating reasonable predictive accuracy in all models (**Fig. 3**). Among them, the LR model performed the best (CHARLS: AUC = 0.7; NHANES: AUC = 0.77), followed by XGBoost (CHARLS: AUC = 0.65; NHANES: AUC = 0.74) and RF (CHARLS: AUC = 0.64; NHANES: AUC = 0.71). As predictors, key variables were same as three models constructed in COX regression models.

Sensitivity analyses and subgroup analyses

Three sensitivity analyses were conducted to further elucidate the association between

METS-IR and future stroke risk in individuals diagnosed with CKM syndrome stages 0–3. In this study, several Cox regression models were used in a group without hypertension (n=5018), a group without a history of heart disease (n=8407) and participants whose BMI was less than 24 (n=5638). The analysis results revealed that the relationships between METS-IR and new-onset stroke remained significant. Our findings were adjusted for factors (age, sex, marital status, smoking history, alcohol consumption history, hypertension, diabetes, heart problems, SBP, DBP, HR, LDL-C, CHO, Crea, UA and HGB) in three models, and the robustness and consistency of the associations between stroke risk and high METS-IR quartiles were verified (**Table 3**). Considering that some potential risk factors might have latency interactions with the effect of METS-IR on future stroke risk in individuals diagnosed with CKM syndrome stages 0–3, a subgroup analysis stratified according to age, sex, current marital status, smoking status and drinking status was performed in our study. The results indicated that the interaction effects between age and METS-IR quartiles were statistically significant (p for interaction<0.01). Moreover, no pronounced interactions between METS-IR and other variables were discerned in any subgroup (gender: p for interaction=0.363; current married: p for interaction=0.066; smoking status: p for interaction=0.618; drinking status: p for interaction=0.854). Thus, on the basis of the results of the subgroup interaction analysis, the robustness of the associations between high METS-IR quartiles and future stroke risk in individuals with CKM syndrome stage 0-3 reported in primary studies was examined (**Table S4**).

Three-year transition of the METS-IR

A total of 4968 participants diagnosed with CKM syndrome stages 0–3 with known 3-year transition data of the METS-IR were enrolled in the trial. In this analysis, when the cohort was divided into five groups, the effect of k-means clustering was superior to that of other options. The following classifications were designed: Class 1, the METS-IR index ranged from 28.2 in 2012 to 24.77 in 2015; Class 1, the METS-IR index ranged from 28.2 in 2012 to 24.77 in 2015, n=1492; Class 2, the METS-IR index ranged from 33.84 in 2012 to 34.92 in 2015, n=1897; Class 3, the METS-IR index ranged from 41.72 in 2012 to 43.57 in 2015, n=1152; Class 4, the METS-IR index ranged from 51.38 in 2012 to 54.4 in 2015, n=393; and Class 5, the METS-IR index ranged from 75.26 in 2012 to 77.3 in 2015, n=34. From Class 1 to Class 5, the basic levels of METS-IR increased and blood glucose was poorly controlled, which was accompanied by a higher rate of new-onset stroke. The proportions of Class 5, Class 4 and Class 3 among patients with stroke were higher than those in non-stroke participants (**Fig. S4**).

The logistic regression analyses revealed a significant association between the 3-year transition of METS-IR and future stroke risk in individuals with CKM syndrome stage 0-3. Compared with Class 1, which was set to the reference group, the risk of new-onset stroke increased gradually with the 3-year transition of the METS-IR from Class 2 to Class 3 (Class 2: OR: 1.236, 95% CI: 1.126-1.449, p=0.015; Class 3: OR: 1.866, 95% CI: 1.417-2.457, p<0.001). Furthermore, Class 5 was both the lowest performing group in terms of adequate METS-IR control and at the highest risk of new-onset stroke (OR: 11.48, 95% CI: 5.656-23.301, p<0.001), followed by Class 4 (OR: 2.634, 95% CI: 1.360-5.101, p=0.004). A sustained high METS-IR was observed in conjunction with high future stroke risk (**Table S5**). Essentially, the resulting associations were harmonized with the analysis results of the baseline METS-IR.

Machine learning algorithm confirm the link between METS-IR and stroke risk

Based on statistical analysis, we observed that increased METS-IR is associated with higher susceptibility to stroke in individuals diagnosed with CKM syndrome stages 0–3. To further validate this observation, we used three machine learning models (LR, XGBoost and RF) for additional analysis. The results showed that the AUC of all three models were greater than 0.6, indicating reasonable predictive accuracy in all models (**Fig. 3**). Among them, the LR model performed the best (CHARLS: AUC = 0.7; NHANES: AUC = 0.77), followed by XGBoost (CHARLS: AUC = 0.65; NHANES: AUC = 0.74) and RF (CHARLS: AUC = 0.64; NHANES: AUC = 0.71). As predictors, key variables were same as three models constructed in COX regression models.

Supplementary of Discussions

METS-IR is a composite index containing FPG, TG, BMI and HDL-C, and increased METS-IR can partly represent high levels of FPG, TG, and BMI as well as low levels of HDL-C, although these associations are not usually consistent and proportional, and METS-IR cannot completely substitute for the other four parameters. We recently reported that TG interacts with vascular inflammation and artery atherosclerosis. Chylomicrons and very low-density lipoproteins related to high levels of TG may play distinct roles in atherosclerosis progression. The current literature supports the beneficial effects of HDL-C because increased HDL-C has been reported to be protective for vascular structure and function. Specific biological mechanisms are involved in reducing cellular death, dampening the inflammatory response, and shielding against pathological oxidation. Increased METS-IR is an evident risk factor for new-onset stroke, and our findings suggest that this effect may be mediated by HDL-C, which might be protective against atherosclerosis progression. In addition, abdominal obesity is causally linked to the likelihood of developing hemorrhagic stroke. BMI demonstrated a potent effect on stroke risk in a previous clinical study. High BMI, FPG and TG are cardinal risk factors for stroke and are frequently present in stroke development. Although the exact mechanisms are not entirely certain, they still offer a favorable basis for exploring the causal relationship between METS-IR and stroke incidence. Taken together, METS-IR represents a potentially promising biomarker for IR and stroke, and further research is warranted to decipher the mechanism through which it is associated with future stroke risk.