


# BMJ Open Validation of type 2 diabetes subgroups by simple clinical parameters: a retrospective cohort study of NHANES data from 1999 to 2014

Jing Xie <sup>1</sup>, Hua Shao,<sup>2</sup> Tao Shan,<sup>3</sup> Shenqi Jing,<sup>3</sup> Yaxiang Shi,<sup>4</sup> Junjie Wang,<sup>3</sup> Jie Hu,<sup>3</sup> Yong Li,<sup>5</sup> Ruochen Huang,<sup>3</sup> Naifeng Liu,<sup>1,6</sup> Yun Liu<sup>3</sup>

**To cite:** Xie J, Shao H, Shan T, *et al.* Validation of type 2 diabetes subgroups by simple clinical parameters: a retrospective cohort study of NHANES data from 1999 to 2014. *BMJ Open* 2022;**12**:e055647. doi:10.1136/bmjopen-2021-055647

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-055647>).

Received 02 August 2021  
Accepted 27 February 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

**Correspondence to**  
Dr Naifeng Liu;  
[liunf@seu.edu.cn](mailto:liunf@seu.edu.cn) and  
Dr Yun Liu;  
[liuyun@njmu.edu.cn](mailto:liuyun@njmu.edu.cn)

## ABSTRACT

**Objectives** To verify whether a simplified method based on age, body mass index (BMI) and glycated haemoglobin (HbA1c) is feasible in classifying patients with type 2 diabetes (T2D), and evaluate the predictive ability of subgroups in several health and mortality outcomes.

**Design** Retrospective cohort study.

**Setting** The National Health and Nutrition Examination Survey 1999–2014 cycle.

**Participants** A total of 1960 participants with diabetes and the age at diagnosis greater than 30.

**Primary and secondary outcome measures** Participants with T2D were assigned to previously defined (by Ahlqvist) subgroups based on five variables: age, BMI, HbA1c, homoeostasis model assessment (HOMA) 2 estimates of  $\beta$ -cell function (HOMA2-B), and insulin resistance (HOMA2-IR), and on three variables: age, BMI and HbA1c. The classification performances of the three variables were evaluated based on 10-fold cross validation, with accuracy, precision and recall as evaluation criteria. Outcomes were assessed using logistic regression and Cox regression analysis.

**Results** Without HOMA measurements, it is difficult to identify severe insulin-resistant diabetes, but other subgroups can be ideally identified. There is no significant difference between the five variables and the three variables in the ability to predict the prevalence of poor cardiovascular health (CVH), chronic kidney disease, non-alcoholic fatty liver disease and advanced liver fibrosis, and the risk of all-cause, cardiovascular disease and cancer-related mortality ( $p > 0.05$ ), except the prevalence of poor CVH in mild age-related diabetes ( $p < 0.05$ ).

**Conclusions** A simple classification based on age, BMI and HbA1c could be used to identify T2D with several health and mortality risks, which is accessible in most individuals with T2D. Due to its simplicity and practicality, more patients with T2D can benefit from subgroup specific treatment paradigms.

## INTRODUCTION

Type 2 diabetes (T2D) is a public health concern worldwide.<sup>1</sup> It is generally acknowledged that T2D is a heterogeneous disease that progresses through multiple physiopathological mechanisms<sup>2</sup> and requires

## Strengths and limitations of this study

- Our study compares the results of three variables with five variables, while Kahkoska's study did not. We also explored the classification performance of mild age-related diabetes and mild obesity-related diabetes, whereas Ahlqvist simply provided explanations for identifying those initially diagnosed with severe insulin-resistant diabetes and severe insulin-deficient diabetes.
- Our study conducts a comparison of comprehensive risk prediction between the two sets of variables regarding the prevalence of cardiovascular health, chronic kidney disease, non-alcoholic fatty liver disease and advanced liver fibrosis at diagnosis, as well as the risks of all-cause, cardiovascular disease and cancer-related mortality many years later.
- Our study explores the influence of sex and different stages of diabetes on the results.
- Though those younger than 30 were excluded, type 1 diabetes may also confound our results.
- There is no information on the risk of complications and the observation time of mortality is not long enough. Further studies are needed to understand the value of classifying type 2 diabetes based on simple variables.

multi-dimensional and sustained treatment.<sup>3</sup> Ahlqvist and colleagues tried to categorise diabetes subgroups, which offers an exciting approach to identifying diabetes heterogeneity.<sup>4</sup> According to the method of Ahlqvist *et al.*, T2D can be classified into four subgroups: severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD) and mild age-related diabetes (MARD). Studies have shown different risks of complications and different intervention strategies across the four subgroups. There is a more rapid progression of kidney disease and higher prevalence of non-alcoholic fatty liver disease (NAFLD) in the SIRD,<sup>4 5</sup> while retinopathy

and neuropathy are more prevalent in the SIDD.<sup>4,5</sup> Those in the SIRD subgroup had a greater glucose reduction efficiency with the use of thiazolidinediones, and those in MARD had a greater glucose reduction efficiency with the use of sulfonylureas.<sup>6</sup> These findings help to improve prevention and treatment and allow targeted medication against diabetes.

This sub-classification approach has been tested in many countries, including India,<sup>7</sup> Japan<sup>8</sup> and China,<sup>9</sup> which suggest that the four T2D subgroups were stable and reproducible in non-European populations. However, fasting C-peptide or fasting insulin are required for homoeostasis model assessment (HOMA) variables which are not routinely performed in clinical practice and are not well standardised, particularly in low-income and middle-income countries. The approach is therefore unworkable for two-thirds of the world's population with diabetes.<sup>10</sup>

Kahkoska *et al* used three simple clinical variables: age at diabetes diagnosis, glycated haemoglobin (HbA1c) and body mass index (BMI), attempting to validate T2D subgroups and evaluate their association with diabetes complications.<sup>11</sup> They showed that three-variable clustering demonstrated similar results with those obtained with the five-variable clustering approach in All New Diabetics in Scania (ANDIS). Subsequently, Ahlqvist used the same method in ANDIS, and found the simplified clustering strategy based on three variables had a good performance in the identification of the SIDD subgroup and a poor performance in the identification of the SIRD subgroup,<sup>12</sup> which emphasised the importance of properly validating alternative clustering methods.

However, there are some limitations in these studies. Kahkoska's study with only three variables lacked a comparison with five variables. Ahlqvist simply provided explanations for identifying those initially diagnosed with SIRD and SIDD, but did not mention the classification performance of MARD and MOD. Furthermore, according to previous studies, SIRD is the least prevalent T2D among all patients,<sup>12</sup> so it remains to be studied whether the overall prediction, clinical risk prediction and therapeutic intervention make sense.

Against this background, the objectives of this report are to (1) explore what are the changes in diabetes subgroups when HOMA2 indicators are missing, and discover the correctness of the division of each subgroup; (2) explore whether sex and different stages of diabetes (newly diagnosed and already suffering from diabetes) have an impact on the results; (3) test the utility of a simpler clustering method in predicting the prevalence of diabetic complications (cardiovascular health (CVH), chronic kidney disease (CKD), nonalcoholic fatty liver disease (NAFLD), advanced liver fibrosis) and the risk of mortality (cancer, cardiovascular disease (CVD) and all-cause mortality).

## MATERIALS AND METHODS

### Patient and public involvement

No patient was involved in the design and conduct of this study.

### Study population

Data were obtained from the National Health and Nutrition Examination Survey (NHANES) (1999–2014; n=82091). The NHANES is a nationally representative, population-based, multistage, cross-sectional study carried out and approved by the US NCHS.<sup>13</sup>

Newly diagnosed diabetes was categorised as 'No' based on the question 'Have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?' and fasting plasma glucose (FPG)  $\geq 7.0$  mmol L<sup>-1</sup> or HbA1c level  $\geq 6.5\%$  (according to American Diabetes Association diabetes diagnostic criteria).<sup>9</sup> Already suffering from diabetes was defined as 'Yes' to the question 'Have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?'.

As information about types of diabetes was not available within the NHANES data, we were uncertain whether the participants had type 1 diabetes or T2D. To address this issue, we excluded participants diagnosed with diabetes before the age of 30.<sup>14</sup> The data cleaning algorithm was shown in online supplemental figure S1, and the final sample size was 1960.

### Outcomes definition

CVH was evaluated according to the American Heart Association's Life's Simple 7 (LS7),<sup>15</sup> which includes: smoking, weight, physical activity, diet, blood cholesterol, blood glucose and blood pressure. AHA definitions of CVH for each metric are shown in table 1. Each LS7 component received a score of 0, 1 or 2 to reflect poor, intermediate or ideal health, respectively. An overall LS7 score ranging from 0 to 14 was calculated as the sum of the LS7 component scores. This score was categorised as low (0–4) and ideal (5–14) CVH.<sup>16</sup>

CKD was defined as kidney damage, with an estimated glomerular filtration rate (eGFR) of less than 60 mL min<sup>-1</sup> per 1.73 m<sup>2</sup>. eGFR was calculated using the CKD-EPI study equation based on serum creatinine.<sup>17</sup>

The criteria to categorise NAFLD included a US Steatosis Index (USFLI) score of  $\geq 30$ , no excessive alcohol consumption (average  $\leq 1$  alcoholic drink per day for women and  $\leq 2$  alcoholic drinks per day for men), negative hepatitis C antibody, and negative Hepatitis B surface antigen.<sup>18</sup> The formula for USFLI score is shown as equation (1).<sup>19</sup>

Advanced liver fibrosis was determined using two non-invasive markers of liver fibrosis: the fibrosis-4 (FIB-4) score and the NAFLD Fibrosis Score (NFS). Their cut-off values were FIB-4 2.67 or NFS  $> 0.676$ .<sup>20</sup> NFS<sup>21</sup> and FIB-4<sup>22</sup> indexes were calculated as shown in equations (2), (3).

**Table 1** AHA definitions of cardiovascular health for each metric

Health metric	Level	Definition
Smoking status*	Ideal	Smoked less than 100 cigarettes
	Intermediate	Smoked at least 100 cigarettes lifetime, but currently not smoking at all
	Poor	Smoked at least 100 cigarettes lifetime and currently smoking
Body mass index†	Ideal	<25 kg/m <sup>2</sup>
	Intermediate	25–29 kg/m <sup>2</sup>
	Poor	≥30 kg/m <sup>2</sup>
Physical activity‡	Ideal	≥150 min/week moderate or ≥75 min/week vigorous or ≥150 min/week moderate+vigorous
	Intermediate	1–149 min/week moderate or 1–74 min/week vigorous or 1–149 min/week moderate+vigorous
	Poor	None
Healthy diet score§	Ideal	≥81
	Intermediate	51–80
	Poor	≤50
Blood pressure¶	Ideal	SBP<120/DBP<80 mm Hg
	Intermediate	SBP 120–139 or DBP 80–89 mm Hg or treated** to SBP<120/DBP<80 mm Hg
	Poor	SBP ≥140 or DBP≥90 mm Hg or treated** not to SBP <120/DBP<80 mm Hg
Total cholesterol††	Ideal	<200 mg/dL
	Intermediate	200–239 mg/dL or treated** to <200 mg/dL
	Poor	≥240 mg/dL or treated** not to <200 mg/dL
Fasting plasma glucose††	Ideal	<100 mg/dL
	Intermediate	100–125 mg/dL or treated** to <100 mg/dL
	Poor	≥126 mg/dL or treated** not to <100 mg/dL

\*Cigarette smoking was determined based on self-reports ‘Smoked at least 100 cigarettes in life’ and ‘Do you now smoke cigarettes’.

†BMI (kg/m<sup>2</sup>) was calculated based on objectively measured weight and height.

‡Physical activity was assessed by self-reported frequency and duration of moderate-intensity and vigorous-intensity recreational activities.

§Dietary intake was obtained from 24 hours dietary recall interviews. From 1999 to 2002 based on the just one dietary interview, whereas from 2003 to 2014 based on the average of first and second dietary interviews, Healthy Eating Index (HEI) scores were calculated by R language using the ‘hei’ package.

¶Systolic (SBP) and diastolic (DBP) blood pressures were measured four times consecutively by certified examiners, the average of all available blood pressure measures for each person was calculated.

\*\*Self-reported history of medications for hypertension, diabetes and high cholesterol were also obtained to determine whether participants received treatments for the conditions above, based on the questions ‘are you now taking prescribed medicines for high blood pressure/cholesterol/blood sugar’.

††Total cholesterol (mg/dL) and fasting plasma glucose (mg/dL) were assessed using fasting blood samples.

$$\begin{aligned}
 \text{US FLI} &= (\exp(-0.8073 \times \text{non-Hispanic black} \\
 &+ 0.3458 \times \text{Mexican American} + 0.0093 \times \\
 &\text{age} + 0.6151 \times \log_e(\text{GGT}) + 0.0249 \times \\
 &\text{waist circumference} + 1.1792 \times \log_e(\text{insulin}) \\
 &+ 0.8242 \times \log_e(\text{glucose}) - 14.7812 / \\
 &(1 + \exp(-0.8073 \times \text{non-Hispanic black} \\
 &+ 0.3458 \times \text{Mexican American} + 0.0093 \times \text{age} + \\
 &0.6151 \times \log_e(\text{GGT}) + 0.0249 \times \text{waist circumference} + \\
 &1.1792 \times \log_e(\text{insulin}) + 0.8242 \\
 &\times \log_e(\text{glucose}) - 14.7812 \times 100 \\
 \text{NFS} &= -1.675 + 0.037 \times \text{age} + 0.094 \times \text{BMI} + 1.13 \\
 &\times (\text{impaired fasting glucose or diabetes}) + 0.99 \\
 &\times (\text{AST/ALT}) - 0.013 \times \text{platelet} - 0.66 \times \text{albumin} \\
 \text{FIB} - 4 &= (\text{age} \times \text{AST}) / (\text{platelet} \times \sqrt[3]{\text{ALT}})
 \end{aligned}
 \tag{1}$$

Where ‘non-Hispanic Black’ and ‘Mexican American’ have a value of 1 if the person is of that ethnicity and 0 if the person is not; where impaired fasting glucose/diabetes has a value of 1 if the subjects have impaired fasting glucose or diabetes and a value of 0 if they do not; age (years), BMI (kg/m<sup>2</sup>), waist circumference (cm), glucose (mg/dL), insulin (pmol/L), GGT (U/L), AST (U/L), ALT (U/L), platelet (×10<sup>9</sup>/L), albumin (g/dL).

Mortality data of the NHANES (1999–2014) participants were provided by the National Centre for Health Statistics (NCHS) using probabilistic record matching with death certificate data found in the National Death Index (NCHS Linked Mortality File) by 31 December 2015. We set mortality-related parameters as all-cause, CVD and cancer-related mortality, based on ICD10 code defined in the NHANES. Follow-up time was the period between the NHANES examination date and the last known date about each person living or death.<sup>23</sup> A total of 1960 participants were followed for a median of 98 (95%

CI 96 to 101) months, calculated by the reverse Kaplan-Meier method.

### Statistical analysis

To demonstrate the reliability of the k-means clustering method based on Euclidean distance, we calculated the silhouette coefficient of different clustering algorithms, and applied the t-SNE visualisation in subgroups.

The classification performance of this study was assessed using 10-fold cross-validation method.<sup>24</sup> First, we calculated the centre of MARD, MOD, SIDD and SIRD subgroups in the training set through k-means clustering based on age, BMI, HbA1c, HOMA2 estimates of  $\beta$ -cell function (HOMA2-B) and insulin resistance (HOMA2-IR). Second, the distance from the five variables to the centre of the four subgroups was measured in the testing set to determine which subgroup did participants belong to (by the nearest distance). Third, the same method was used in the same testing set to determine which subgroup did participants belong to based on the distance between age, BMI and HbA1c to the centre of the four subgroups. Finally, we treated the five-variable results as the real value, the three-variable results as predicted value in the testing sets, and used the ‘confusionMatrix’ function in the R package caret to evaluate the performance of the three-variable classifier. Acknowledging the differences between our study and the study of Ahlqvist *et al*,<sup>4</sup> we chose letter-based cluster labels, which correspond to the four subgroups proposed by Ahlqvist: Cluster A (MARD), Cluster B (MOD), Cluster C (SIDD) and Cluster D (SIRD). In order to visualise the participants’ redistribution among different subgroups, the Sankey diagram was used and plotted by the ‘sankeyNetwork’ function in the R package networkD3.

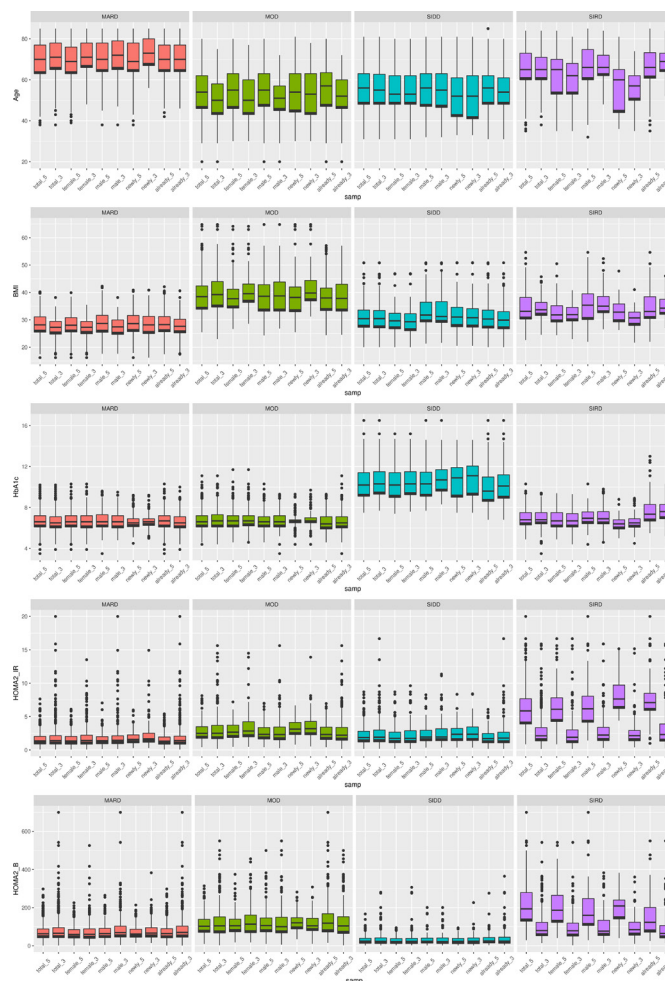
Additional analyses were conducted to compare the five-variable classifier and the three-variable classifier in the ability to predict CVH, CKD, NAFLD, advanced liver fibrosis, all-cause, CVD and cancer-related mortality. Odds ratio (OR) was used to evaluate the prevalence of poor CVH, CKD, NAFLD and advanced liver fibrosis across different classification approaches and subgroups. Logistic regression was conducted to calculate OR. The long-term mortality risks between different classification methods and different subgroups estimated from the Cox regression models were expressed as hazard ratio (HR).

As laboratory techniques varied in the long term as studied in our analysis, we calibrated FPG in 2005–2014 and insulin in 2003–2014 to earlier NHANES surveys, but did not calibrate HbA1c as it is not necessary according to NHANES recommendations.<sup>25 26</sup> We also used fasting subsample weights recommended by the NHANES, and the survey package in R (V.4.1-1) to account for the complex sampling design.

## RESULTS

### Subgroups characteristics

The study population included 1960 adults from the NHANES (1999–2014) database. Through the t-SNE



**Figure 1** Distributions and of characteristics in the four type 2 diabetes subgroups based on 5-variable and 3-variable. BMI, body mass index; HbA1c, glycated haemoglobin; HOMA2-B, homoeostasis model assessment 2 of  $\beta$ -cell function; HOMA2-IR, homoeostasis model assessment 2 of insulin resistance; MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes. total\_5, All sample clustering based on 5-variable; total\_3, All sample clustering based on 3-variable; female\_5, Female sample clustering based on 5-variable; female\_3, Female sample clustering based on 3-variable; male\_5, Male sample clustering based on 5-variable; male\_3, Male sample clustering based on 3-variable; newly\_5, Newly diagnosed with T2D sample clustering based on 5-variable; newly\_3, Newly diagnosed with T2D sample clustering based on 3-variable; already\_5, Already diagnosed with T2D sample clustering based on 5-variable; already\_3, Already diagnosed with T2D sample clustering based on 3-variable.

visualisation of four subgroups and the silhouette coefficient of different clustering algorithm, we found that the k-means clustering method based on Euclidean distance was reliable (online supplemental figure S2, supplemental table S1).

As shown in figure 1, the demographic characteristics of the four subgroups (MARD, MOD, SIDD, SIRD) based on five-variable were consistent with many previous reports,

**Table 2** Performance in classifying the four subgroups

	Criteria	Cluster A	Cluster B	Cluster C	Cluster D
Total	Sensitivity	0.76	0.70	0.89	0.42
	Specificity	0.95	0.95	0.99	0.79
	Pos pred value	0.93	0.84	0.95	0.16
	Neg pred value	0.79	0.90	0.98	0.93
	Balanced accuracy	0.85	0.83	0.94	0.60
Male	Sensitivity	0.73	0.67	0.89	0.40
	Specificity	0.94	0.97	0.99	0.76
	Pos pred value	0.93	0.87	0.94	0.12
	Neg pred value	0.75	0.91	0.98	0.93
	Balanced accuracy	0.84	0.82	0.94	0.58
Female	Sensitivity	0.76	0.73	0.82	0.39
	Specificity	0.94	0.94	0.99	0.80
	Pos pred value	0.92	0.83	0.94	0.15
	Neg pred value	0.81	0.89	0.97	0.93
	Balanced accuracy	0.85	0.83	0.91	0.59
Newly diagnosed	Sensitivity	0.74	0.68	0.89	0.36
	Specificity	0.94	0.95	1.00	0.79
	Pos pred value	0.92	0.85	0.97	0.08
	Neg pred value	0.78	0.87	0.98	0.96
	Balanced accuracy	0.84	0.81	0.94	0.57
Already diagnosed	Sensitivity	0.82	0.77	0.79	0.41
	Specificity	0.92	0.94	0.99	0.86
	Pos pred value	0.91	0.80	0.93	0.18
	Neg pred value	0.84	0.92	0.95	0.95
	Balanced accuracy	0.87	0.85	0.89	0.64

Sensitivity (recall)=TP/(TP+FN).

Specificity=TN/(FP+TN).

Pos pred value (precision)=TP/(TP+FP).

Neg pred value=TN/(FN+TN).

Balanced accuracy= (TP/(TP+FN)+TN/(FP+TN))/2.

FN, false negative (incorrectly classified positive cases); FP, false positive (incorrectly classified negative cases); TN, true negative (correctly classified negative cases); TP, true positive (correctly classified positive cases).

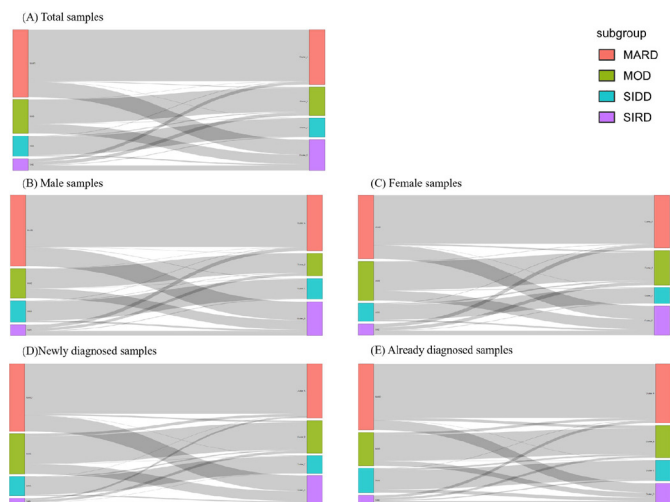
which shows that the classification was stable. The SIRD subgroup was characterised by higher HOMA2-IR and HOMA2-B. The SIDD subgroup had the highest HbA1c and the lowest HOMA2-B. The MOD subgroup presented a lower age and higher BMI, average blood glucose,  $\beta$ -cell function and insulin resistance. The MARD subgroup demonstrated a higher age and lower BMI and HbA1c, average  $\beta$ -cell function and insulin resistance.

We found that the age of cluster A was generally higher than that of MARD, and that the age of newly diagnosed patients was higher than that of patients with a history of diabetes. Comparing cluster B and MOD, we found that the age of cluster B is generally lower than that of MOD. The BMI level in cluster B was higher than that in MOD. When comparing cluster C with SIDD, the results of all variables were essentially the same, and the BMI for males was generally higher than that for females. Compared

with SIRD, cluster D had a lower age in the newly diagnosed sample. The BMI of the newly diagnosed was lower than that of the already diagnosed. HbA1c was essentially the same between cluster D and SIRD. The HbA1c of the newly diagnosed was significantly lower than that of the already diagnosed. In terms of HOMA2-B and HOMA2-IR, cluster D was significantly lower than SIRD, which indicated that it is difficult to identify the high insulin resistance characteristic without the HOMA2 index. However, when the HOMA2 indicators were missing, the characteristics of cluster A, B, C and MARD, MOD and SIDD were basically the same.

### Classification performance

Online supplemental table S2 shows the performance evaluation of the three-variable features compared with the five-variable in classifying subgroups. The overall



**Figure 2** Patients redistribution among different subgroups. MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.

classification accuracy was 0.74 (0.72, 0.76), kappa was 0.62. Females are more accurate in identifying the subgroups than males (accuracy: 0.73 vs 0.72). The accuracy of already diagnosed diabetes is higher than that of newly diagnosed diabetes (accuracy: 0.77 vs 0.73).

The precision (pos pred value) of cluster A, cluster B, cluster C and cluster D were 0.93, 0.84, 0.95 and 0.16, their recall (sensitivity) 0.76, 0.70, 0.89 and 0.42, and balanced accuracy 0.85, 0.83, 0.94 and 0.60, respectively. Except for cluster D, we found other subgroups had a relatively ideal classification effect, and cluster C had the highest accuracy (0.94). The results of males, females, the newly diagnosed and the already diagnosed were similar to those of total samples. In the already diagnosed, cluster A had a better classification effect than total samples. Overall, the sex distribution and stages of diabetes did not differ markedly across the subgroups (table 2).

The pattern of participants redistribution was shown in a Sankey diagram (figure 2). Overall, a total of 517 (26.38%) participants switched into different subgroups, mainly from MARD and MOD to cluster D. The redistribution of male, female, newly diagnosed and already diagnosed samples were similar to that of total samples.

### Risk-subgroups association

In the total samples, except for MARD subgroup, there were no significant differences in predicting the prevalence of poor CVH, CKD, NAFLD and advanced liver fibrosis in every two paired subgroups obtained using three-variable and five-variable classification approaches (all the 95% CIs of OR included value 1). The poor CVH prevalence of cluster A is significantly lower than that of MARD, the adj. OR (95% CI) were 0.7 (0.56 to 0.89). In predicting the risk of mortality, the mortality showed no differences in every two paired groups classified by two different approaches (all the 95% CIs of OR included

value 1). Different sexes and courses of diabetes had substantially similar results, as shown in table 3.

Table 4 lists the ORs of diabetic complications among different subgroups based on three-variable. Compared with cluster A, cluster B had the highest poor CVH prevalence (adj. OR: 5.95, 95% CI 4.08 to 8.66), cluster C and cluster D's poor CVH prevalence were also significantly higher. Cluster C had the lowest prevalence of CKD (adj. OR: 0.21, 95% CI 0.11 to 0.41), cluster B also had a significantly lower prevalence. Cluster D had a significantly higher prevalence of NAFLD (adj. OR: 1.68, 95% CI 1.10 to 2.57), and there were no significant difference between cluster A and other subgroups. Among the four subgroups, cluster C had the lowest prevalence of advanced liver fibrosis (adj. OR: 0.31, 95% CI 0.17 to 0.57), while cluster D has the highest prevalence (adj. OR: 1.89, 95% CI 1.31 to 2.73).

After a median of 98 (95% CI 96 to 101) months follow-up, the HR of all-cause, CVD and cancer-related mortality among the subgroups were shown in table 4. Compared with cluster A, the adjusted HR (95% CI) of all-cause mortality of cluster B, cluster C and cluster D were 0.42 (0.3 to 0.6), 0.38 (0.21 to 0.69) and 0.66 (0.48 to 0.93), respectively. Cluster A had the highest all-cause mortality. Cluster B had a significantly lower CVD mortality risk than cluster A (adj. HR: 0.35, 95% CI 0.15 to 0.81), and there was no significant difference between the other subgroups. In terms of cancer-related mortality, there was no significant difference among the four subgroups (all the 95% CIs of HR included value 1).

In male samples, we found the prevalence of non-alcoholic fatty liver and advanced liver fibrosis of cluster D showed no significant difference compared with cluster A (adj. OR: 1.18, 95% CI 0.70 to 1.98, and 1.22, 95% CI 0.76 to 1.98). Also, there was no significant difference between cluster B and cluster A regarding the risk of CVD mortality (adj. HR: 0.64, 95% CI 0.34 to 1.2). In female samples, the prevalence of advanced liver fibrosis displayed no significant difference in cluster C and cluster A (adj. OR: 0.71, 95% CI 0.30 to 1.67). Predictions of mortality risk in each subgroup of females were consistent with those of males. In the newly diagnosed samples, cluster C and cluster D had no significant difference in terms of poor CVH compared with cluster A (the 95% CIs of OR included value 1). Compared with cluster A, cluster D had no significant difference in the prevalence of NAFLD (adj. OR: 1.30, 95% CI 0.68 to 2.48), significantly lower prevalence of advanced liver fibrosis (adj. OR: 0.40, 95% CI 0.17 to 0.94). In already diagnosed samples, cluster D had no significantly higher prevalence of NAFLD (adj. OR: 1.61, 95% CI 0.88 to 2.95). Also, cluster D had no significant lower all-cause mortality risk than cluster A, the adjusted HR (95% CI) was 0.78 (0.52 to 1.16).

### DISCUSSION

In this study, we tested reliability when HOMA2 indicators are absent to stratify T2D as well as the utility of

**Table 3** Outcome risks for each subgroup between the five-variable method and three-variable approaches

	Poor CVH			CKD			NAFLD			Advanced liver fibrosis			All-cause mortality			CVD-related mortality			Cancer-related mortality		
	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	
Total†																					
MARD vs cluster A*	0.76 (0.61 to 0.94)	0.71 (0.56 to 0.89)	1.16 (0.94 to 1.43)	1.14 (0.92 to 1.41)	0.99 (0.77 to 1.26)	0.99 (0.77 to 1.27)	0.92 (0.75 to 1.13)	0.99 (0.77 to 1.27)	0.99 (0.77 to 1.27)	1.09 (0.93 to 1.28)	1.13 (0.81 to 1.59)	1.12 (0.8 to 1.58)	1.09 (0.93 to 1.28)	1.12 (0.8 to 1.58)	0.95 (0.62 to 1.47)	0.96 (0.62 to 1.47)	0.95 (0.62 to 1.47)	1.12 (0.8 to 1.58)	0.95 (0.62 to 1.47)	0.96 (0.62 to 1.47)	0.95 (0.62 to 1.47)
MOD vs cluster B	1.01 (0.78 to 1.31)	1.09 (0.8 to 1.49)	0.86 (0.58 to 1.28)	0.89 (0.57 to 1.4)	0.86 (0.63 to 1.17)	0.85 (0.61 to 1.17)	0.86 (0.65 to 1.14)	0.85 (0.61 to 1.17)	0.86 (0.65 to 1.14)	0.91 (0.63 to 1.32)	0.69 (0.33 to 1.46)	0.69 (0.33 to 1.46)	0.91 (0.63 to 1.32)	0.69 (0.33 to 1.46)	0.7 (0.28 to 1.78)	0.76 (0.3 to 1.95)	0.7 (0.28 to 1.78)	0.69 (0.33 to 1.46)	0.7 (0.28 to 1.78)	0.76 (0.3 to 1.95)	0.7 (0.28 to 1.78)
SIDD vs cluster C	0.98 (0.71 to 1.37)	1.00 (0.68 to 1.47)	0.98 (0.53 to 1.81)	0.96 (0.5 to 1.82)	0.94 (0.61 to 1.46)	0.94 (0.61 to 1.46)	1.05 (0.63 to 1.76)	0.94 (0.61 to 1.46)	0.94 (0.61 to 1.46)	0.98 (0.66 to 1.45)	1.21 (0.63 to 2.33)	1.13 (0.58 to 2.17)	0.98 (0.66 to 1.45)	1.13 (0.58 to 2.17)	0.92 (0.4 to 2.12)	0.92 (0.4 to 2.12)	0.92 (0.4 to 2.12)	0.98 (0.66 to 1.45)	0.92 (0.4 to 2.12)	0.92 (0.4 to 2.12)	0.92 (0.4 to 2.12)
SIRD vs cluster D	1.07 (0.75 to 1.52)	1.03 (0.69 to 1.53)	0.66 (0.44 to 0.99)	0.69 (0.45 to 1.08)	0.96 (0.64 to 1.44)	0.97 (0.64 to 1.48)	1.03 (0.72 to 1.48)	0.97 (0.64 to 1.48)	1.08 (0.74 to 1.57)	0.76 (0.53 to 1.05)	0.67 (0.35 to 1.29)	0.72 (0.37 to 1.4)	0.76 (0.53 to 1.05)	0.67 (0.35 to 1.29)	1.1 (0.52 to 2.35)	1.1 (0.52 to 2.35)	1.1 (0.52 to 2.35)	0.76 (0.53 to 1.05)	1.1 (0.52 to 2.35)	1.1 (0.52 to 2.35)	1.1 (0.52 to 2.35)
Male†																					
MARD vs cluster A	0.81 (0.6 to 1.1)	0.81 (0.58 to 1.12)	1.21 (0.91 to 1.61)	1.18 (0.88 to 1.59)	1.05 (0.78 to 1.42)	1.07 (0.79 to 1.45)	1.02 (0.78 to 1.35)	1.07 (0.79 to 1.45)	1.00 (0.74 to 1.33)	1.19 (0.96 to 1.48)	1.21 (0.73 to 1.99)	1.2 (0.73 to 1.98)	1.19 (0.96 to 1.48)	1.21 (0.73 to 1.99)	1.08 (0.64 to 1.83)	1.08 (0.64 to 1.83)	1.08 (0.64 to 1.83)	1.19 (0.96 to 1.48)	1.08 (0.64 to 1.83)	1.08 (0.64 to 1.83)	1.08 (0.64 to 1.83)
MOD vs cluster B	1.08 (0.73 to 1.61)	1.23 (0.77 to 1.99)	0.86 (0.46 to 1.6)	0.73 (0.36 to 1.46)	0.93 (0.6 to 1.42)	0.92 (0.58 to 1.44)	1.01 (0.66 to 1.53)	0.92 (0.58 to 1.44)	1.03 (0.66 to 1.6)	0.88 (0.58 to 1.42)	1.07 (0.51 to 2.26)	1 (0.47 to 2.09)	0.88 (0.58 to 1.42)	1.07 (0.51 to 2.26)	0.7 (0.23 to 2.09)	0.7 (0.23 to 2.09)	0.7 (0.23 to 2.09)	0.88 (0.58 to 1.42)	0.7 (0.23 to 2.09)	0.7 (0.23 to 2.09)	0.7 (0.23 to 2.09)
SIDD vs cluster C	0.91 (0.58 to 1.43)	0.77 (0.45 to 1.34)	1.38 (0.61 to 3.14)	1.3 (0.55 to 3.04)	0.91 (0.52 to 1.6)	0.87 (0.49 to 1.54)	1.06 (0.49 to 2.31)	0.87 (0.49 to 1.54)	1.02 (0.45 to 2.33)	1.15 (0.7 to 1.88)	1.04 (0.48 to 2.29)	0.97 (0.44 to 2.13)	1.15 (0.7 to 1.88)	1.04 (0.48 to 2.29)	1.74 (0.49 to 6.19)	1.74 (0.49 to 6.19)	1.74 (0.49 to 6.19)	1.15 (0.7 to 1.88)	1.74 (0.49 to 6.19)	1.74 (0.49 to 6.19)	1.74 (0.49 to 6.19)
SIRD vs cluster D	0.89 (0.54 to 1.47)	0.87 (0.5 to 1.54)	0.61 (0.32 to 1.16)	0.71 (0.36 to 1.41)	0.69 (0.41 to 1.16)	0.7 (0.41 to 1.19)	1.02 (0.59 to 1.77)	0.7 (0.41 to 1.19)	1.14 (0.62 to 2.07)	0.51 (0.32 to 0.81)	0.57 (0.24 to 1.37)	0.62 (0.25 to 1.52)	0.51 (0.32 to 0.81)	0.57 (0.24 to 1.37)	0.78 (0.27 to 2.22)	0.78 (0.27 to 2.22)	0.78 (0.27 to 2.22)	0.51 (0.32 to 0.81)	0.78 (0.27 to 2.22)	0.78 (0.27 to 2.22)	0.78 (0.27 to 2.22)
Female†																					
MARD vs cluster A	0.71 (0.53 to 0.96)	0.69 (0.5 to 0.95)	1.18 (0.87 to 1.6)	1.15 (0.84 to 1.57)	1.06 (0.68 to 1.65)	1.08 (0.69 to 1.69)	0.88 (0.65 to 1.21)	1.08 (0.69 to 1.69)	0.86 (0.62 to 1.18)	1.11 (0.87 to 1.41)	1.19 (0.75 to 1.88)	1.2 (0.76 to 1.89)	1.11 (0.87 to 1.41)	1.19 (0.75 to 1.88)	0.88 (0.42 to 1.84)	0.88 (0.42 to 1.84)	0.88 (0.42 to 1.84)	1.11 (0.87 to 1.41)	0.88 (0.42 to 1.84)	0.88 (0.42 to 1.84)	0.88 (0.42 to 1.84)
MOD vs cluster B	0.93 (0.66 to 1.32)	0.96 (0.63 to 1.46)	0.68 (0.4 to 1.15)	0.74 (0.42 to 1.3)	0.83 (0.53 to 1.3)	0.81 (0.52 to 1.29)	0.82 (0.56 to 1.19)	0.81 (0.52 to 1.29)	0.86 (0.58 to 1.26)	1.06 (0.6 to 1.86)	0.28 (0.03 to 2.48)	0.36 (0.04 to 3.39)	1.06 (0.6 to 1.86)	0.28 (0.03 to 2.48)	0.93 (0.29 to 3.06)	0.93 (0.29 to 3.06)	0.93 (0.29 to 3.06)	1.06 (0.6 to 1.86)	0.93 (0.29 to 3.06)	0.93 (0.29 to 3.06)	0.93 (0.29 to 3.06)
SIDD vs cluster C	1.00 (0.60 to 1.65)	1.00 (0.56 to 1.76)	0.86 (0.36 to 2.05)	0.81 (0.31 to 2.11)	1.02 (0.5 to 2.08)	1.02 (0.5 to 2.09)	0.98 (0.5 to 1.92)	1.02 (0.5 to 1.92)	1.04 (0.51 to 2.09)	1.13 (0.59 to 2.17)	1.54 (0.34 to 6.86)	1.32 (0.29 to 5.93)	1.13 (0.59 to 2.17)	1.54 (0.34 to 6.86)	0.94 (0.29 to 3.07)	0.94 (0.29 to 3.07)	0.94 (0.29 to 3.07)	1.13 (0.59 to 2.17)	0.94 (0.29 to 3.07)	0.94 (0.29 to 3.07)	0.94 (0.29 to 3.07)
SIRD vs cluster D	1.55 (0.93 to 2.58)	1.59 (0.89 to 2.82)	0.69 (0.4 to 1.21)	0.82 (0.44 to 1.54)	1.03 (0.51 to 2.07)	1.05 (0.52 to 2.14)	1.07 (0.64 to 1.81)	1.05 (0.52 to 2.14)	1.14 (0.66 to 1.96)	0.6 (0.37 to 0.99)	0.57 (0.21 to 1.57)	0.64 (0.21 to 1.88)	0.6 (0.37 to 0.99)	0.57 (0.21 to 1.57)	1.23 (0.34 to 4.47)	1.23 (0.34 to 4.47)	1.23 (0.34 to 4.47)	0.64 (0.21 to 1.88)	1.23 (0.34 to 4.47)	1.23 (0.34 to 4.47)	1.23 (0.34 to 4.47)
Newly diagnosed§																					
MARD vs cluster A	1.01 (0.67 to 1.5)	1.08 (0.7 to 1.65)	1.35 (0.89 to 2.04)	1.36 (0.89 to 2.09)	1.03 (0.68 to 1.56)	1.00 (0.65 to 1.54)	1.22 (0.81 to 1.83)	1.22 (0.81 to 1.83)	1.21 (0.8 to 1.84)	1.34 (1 to 1.79)	1.27 (0.74 to 2.19)	1.31 (0.76 to 2.27)	1.34 (1 to 1.79)	1.27 (0.74 to 2.19)	1.3 (0.64 to 2.64)	1.3 (0.64 to 2.64)	1.3 (0.64 to 2.64)	1.34 (1 to 1.79)	1.3 (0.64 to 2.64)	1.3 (0.64 to 2.64)	1.3 (0.64 to 2.64)
MOD vs cluster B	1.03 (0.64 to 1.65)	1.25 (0.73 to 2.14)	0.65 (0.25 to 1.68)	0.62 (0.23 to 1.63)	0.91 (0.53 to 1.57)	0.95 (0.53 to 1.71)	1.15 (0.7 to 1.89)	1.15 (0.7 to 1.89)	1.15 (0.7 to 1.89)	1.1 (0.59 to 2.04)	1.59 (0.49 to 5.22)	1.85 (0.54 to 6.37)	1.1 (0.59 to 2.04)	1.59 (0.49 to 5.22)	0.96 (0.26 to 3.58)	0.96 (0.26 to 3.58)	0.96 (0.26 to 3.58)	1.1 (0.59 to 2.04)	0.96 (0.26 to 3.58)	0.96 (0.26 to 3.58)	0.96 (0.26 to 3.58)
SIDD vs cluster C	1.08 (0.55 to 2.14)	0.98 (0.42 to 2.29)	1.1 (0.3 to 3.97)	1.07 (0.17 to 6.59)	0.95 (0.43 to 2.09)	0.94 (0.42 to 2.1)	0.98 (0.37 to 2.58)	0.98 (0.37 to 2.58)	0.93 (0.34 to 2.57)	1.16 (0.49 to 2.75)	1.11 (0.28 to 4.44)	1.05 (0.26 to 4.21)	1.16 (0.49 to 2.75)	1.11 (0.28 to 4.44)	1.09 (0.15 to 7.72)	1.09 (0.15 to 7.72)	1.09 (0.15 to 7.72)	1.16 (0.49 to 2.75)	1.09 (0.15 to 7.72)	1.09 (0.15 to 7.72)	1.09 (0.15 to 7.72)
SIRD vs cluster D	0.45 (0.18 to 1.08)	0.26 (0.09 to 0.74)	1.6 (0.19 to 13.62)	2.46 (0.25 to 23.99)	1.4 (0.48 to 4.06)	1.6 (0.5 to 5.15)	1.13 (0.3 to 4.26)	1.13 (0.3 to 4.26)	1.22 (0.28 to 5.22)	0.4 (0.16 to 1.01)	0.24 (0.08 to 0.66)	-¶	0.4 (0.16 to 1.01)	0.24 (0.08 to 0.66)	0.33 (0.06 to 1.99)	0.33 (0.06 to 1.99)	0.33 (0.06 to 1.99)	0.4 (0.16 to 1.01)	0.33 (0.06 to 1.99)	0.33 (0.06 to 1.99)	0.33 (0.06 to 1.99)

Continued

Table 3 Continued

	Poor CVH		CKD		NAFLD		Advanced liver fibrosis		All-cause mortality		CVD-related mortality		Cancer-related mortality	
	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)
Already diagnosed†	0.8 (0.63 to 1.02)	0.78 (0.6 to 1)	1.01 (0.8 to 1.29)	1.02 (0.8 to 1.31)	1.15 (0.86 to 1.54)	1.14 (0.84 to 1.55)	0.88 (0.7 to 1.1)	0.88 (0.7 to 1.1)	1.01 (0.84 to 1.23)	1.02 (0.84 to 1.23)	1.05 (0.69 to 1.6)	1.06 (0.69 to 1.62)	0.94 (0.57 to 1.56)	0.94 (0.57 to 1.56)
MOD vs cluster B	0.93 (0.69 to 1.25)	0.96 (0.67 to 1.36)	0.73 (0.49 to 1.1)	0.81 (0.52 to 1.27)	0.78 (0.55 to 1.1)	0.77 (0.54 to 1.11)	0.85 (0.62 to 1.18)	0.95 (0.67 to 1.33)	0.96 (0.64 to 1.42)	1.08 (0.72 to 1.62)	0.61 (0.25 to 1.47)	0.6 (0.25 to 1.47)	0.71 (0.29 to 2.25)	0.89 (0.35 to 2.25)
SIDD vs cluster C	0.93 (0.65 to 1.34)	0.95 (0.62 to 1.46)	0.96 (0.45 to 2.04)	0.96 (0.44 to 2.09)	0.82 (0.49 to 1.36)	0.83 (0.5 to 1.38)	1 (0.55 to 1.82)	1.05 (0.57 to 1.95)	0.89 (0.58 to 1.36)	0.85 (0.55 to 1.3)	0.96 (0.45 to 2.05)	0.85 (0.4 to 1.82)	1.05 (0.38 to 2.9)	1.07 (0.39 to 2.96)
SIRD vs cluster D	1.51 (0.95 to 2.4)	1.76 (1.04 to 2.97)	0.91 (0.56 to 1.49)	0.91 (0.53 to 1.54)	1.03 (0.59 to 1.8)	0.98 (0.55 to 1.74)	1.13 (0.71 to 1.8)	1.09 (0.66 to 1.79)	0.94 (0.64 to 1.39)	0.92 (0.62 to 1.37)	1.05 (0.52 to 2.13)	1.08 (0.53 to 2.19)	1.68 (0.55 to 5.19)	1.52 (0.49 to 4.72)

\*All comparisons take MARD, MOD, SIDD and SIRD as reference subgroups.

†Adjust for sex, race, ever smoking, education level, total cholesterol, systolic blood pressure and duration of diabetes.

‡Adjust for race, ever smoking, education level, total cholesterol, systolic blood pressure and duration of diabetes.

§Adjust for sex, race, ever smoking, education level, total cholesterol and systolic blood pressure.

¶Since there is too little cardiovascular mortality in SIRD and cluster D, there is no calculation result.

CKD, chronic kidney disease; CVD, cardiovascular disease; CVH, cardiovascular health; MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; NAFLD, non-alcoholic fatty liver disease; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.

the simplified clustering method to predict the risk of diabetes-related outcomes, including CVH, CKD, NAFLD, advanced liver fibrosis, cancer-related mortality, CVD-related mortality and all-cause mortality.

Our study shows that without HOMA measurements, MARD, MOD and SIDD subgroups can be ideally identified. The characteristic of SIDD is that its HbA1c was significantly higher than other subgroups, MARD was older and MOD had a higher BMI. Therefore, age, BMI and HbA1c can ideally identify the three subgroups. It is difficult to identify SIRD due to the absence of clinical characteristics of insulin resistance. We also attempted to add reliable and inexpensive surrogate biomarkers of insulin resistance, such as triglyceride-glucose index (TyG),<sup>27 28</sup> TyG-body mass index (TyG-BMI),<sup>29</sup> and TyG-waist circumference (TyG-WC),<sup>29</sup> triglyceride to HDL-C ratio (TG/HDL-C).<sup>30</sup> However, no indicators were found to correspond to the distribution of HOMA2-IR, as shown in online supplemental figure S3. In the NHANES 1999–2014, we observed low correlation coefficients among TyG, TyG-BMI, TyG-WC, TG/HDL-C and HOMA2-IR. The correlation coefficient in the group without diabetes was found to be higher than in the group with diabetes, the correlation coefficient between TyG-WC and HOMA2-IR was the greatest in the group with diabetes, as shown in online supplemental table S3, but TyG-WC cannot replace the HOMA2-IR. The relationship between these indexes and HOMA2-IR may be affected by many factors such as race, obesity and health status, which require further investigation. Furthermore, our study showed that three-variable classification and five-variable method had similar effects in risk prediction.

In the total samples, cluster A was found to have the lowest prevalence of poor CVH, while cluster B the highest. The CVH of males was worse than that of females in cluster B. An increasing number of studies indicate that CVH is related to the risk of CVD,<sup>31 32</sup> and non-CVD outcomes.<sup>33 34</sup> Therefore, we need to pay more attention to males in cluster B to prevent future CVD and non-CVD, and urge them to change their daily life to improve their CVH, such as improve diet quality, change dietary behaviours<sup>35</sup> and enhance aerobic exercise.<sup>36</sup> For CKDs, we found that cluster A had the highest prevalence of diagnosis, which may be related to their higher age.<sup>37</sup> Hence, for newly diagnosed cluster A, the impact of CKD on drug selection should be taken into account while formulating treatment plans. The prevalence of both NAFLD and advanced liver fibrosis was high in cluster D, especially among women. This may be related to insulin resistance.<sup>38</sup> Patients in this subgroup should be prioritised for liver assessment. Cluster A presented the highest risk of all-cause and CVD mortality, largely due to higher age. Therefore, more attention should be paid to this subgroup for its high risk of mortality, particularly the screening of CVD, although it had a low prevalence of poor CVD. For patients who are diagnosed as cluster B, the risk of cardiovascular events may be relatively high, and it is also necessary to pay attention to screening for cardiovascular events for this subgroup.



**Table 4** Different risks among the four subgroups

	Poor CVH			CKD			NAFLD			Advanced liver fibrosis			All-cause mortality			CVD-related mortality			Cancer-related mortality		
	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Adj. OR (95% CI)	Crude OR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)	Crude HR (95% CI)	Adj. HR (95% CI)				
<b>Total*</b>	Reference	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Cluster A	3.13 (2.31 to 4.24)	5.95 (4.08 to 8.66)	0.24 (0.16 to 0.38)	0.32 (0.20 to 0.52)	1.54 (1.04 to 2.28)	1.36 (0.90 to 2.05)	0.96 (0.71 to 1.30)	1.29 (0.92 to 1.82)	0.32 (0.22 to 0.47)	0.42 (0.3 to 0.6)	0.35 (0.15 to 0.81)	0.28 (0.12 to 0.63)	0.35 (0.15 to 0.81)	0.1 (0.01 to 1.53)	0.11 (0.01 to 1.58)	0.1 (0.01 to 1.53)	0.11 (0.01 to 1.58)	0.1 (0.01 to 1.53)	0.11 (0.01 to 1.58)	0.1 (0.01 to 1.53)	
Cluster B	3.02 (2.16 to 4.22)	4.10 (2.76 to 6.08)	0.19 (0.09 to 0.37)	0.21 (0.11 to 0.41)	1.24 (0.77 to 2.00)	1.30 (0.80 to 2.11)	0.24 (0.13 to 0.45)	0.31 (0.17 to 0.57)	0.34 (0.18 to 0.65)	0.38 (0.21 to 0.69)	0.45 (0.16 to 1.28)	0.43 (0.15 to 1.24)	0.45 (0.16 to 1.28)	0.65 (0.21 to 2.04)	0.69 (0.25 to 1.89)	0.65 (0.21 to 2.04)	0.69 (0.25 to 1.89)	0.65 (0.21 to 2.04)	0.69 (0.25 to 1.89)	0.65 (0.21 to 2.04)	
Cluster C	3.21 (2.35 to 4.38)	4.38 (3.03 to 6.32)	0.63 (0.43 to 0.93)	0.67 (0.45 to 1.00)	1.69 (1.11 to 2.58)	1.68 (1.10 to 2.57)	1.70 (1.20 to 2.43)	1.89 (1.31 to 2.73)	0.64 (0.46 to 0.88)	0.66 (0.48 to 0.93)	0.57 (0.25 to 1.26)	0.55 (0.25 to 1.21)	0.57 (0.25 to 1.26)	0.84 (0.45 to 1.57)	0.83 (0.44 to 1.56)	0.84 (0.45 to 1.57)	0.83 (0.44 to 1.56)	0.84 (0.45 to 1.57)	0.83 (0.44 to 1.56)	0.84 (0.45 to 1.57)	
<b>Male†</b>	Reference	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Cluster A	4.92 (3.03 to 7.98)	11.27 (6.10 to 20.82)	0.21 (0.09 to 0.51)	0.26 (0.10 to 0.66)	1.11 (0.67 to 1.82)	0.96 (0.57 to 1.63)	0.91 (0.57 to 1.47)	1.34 (0.81 to 2.20)	0.41 (0.25 to 0.68)	0.5 (0.31 to 0.83)	0.64 (0.34 to 1.2)	0.55 (0.3 to 1)	0.64 (0.34 to 1.2)	0.15 (0.02 to 1.31)	0.18 (0.02 to 1.61)	0.15 (0.02 to 1.31)	0.18 (0.02 to 1.61)	0.15 (0.02 to 1.31)	0.18 (0.02 to 1.61)	0.15 (0.02 to 1.31)	
Cluster B	3.74 (2.26 to 6.20)	3.86 (1.92 to 7.75)	0.25 (0.11 to 0.55)	0.28 (0.12 to 0.64)	0.83 (0.44 to 1.55)	0.94 (0.50 to 1.77)	0.21 (0.09 to 0.49)	0.29 (0.13 to 0.67)	0.53 (0.29 to 0.94)	0.51 (0.28 to 0.95)	0.42 (0.08 to 2.13)	0.51 (0.1 to 2.52)	0.42 (0.08 to 2.13)	0.52 (0.13 to 2.06)	0.61 (0.15 to 2.44)	0.52 (0.13 to 2.06)	0.61 (0.15 to 2.44)	0.52 (0.13 to 2.06)	0.61 (0.15 to 2.44)	0.52 (0.13 to 2.06)	
Cluster C	3.23 (2.04 to 5.13)	4.05 (2.36 to 6.94)	0.35 (0.20 to 0.60)	0.41 (0.23 to 0.70)	1.21 (0.72 to 2.01)	1.18 (0.70 to 1.98)	0.96 (0.61 to 1.51)	1.22 (0.76 to 1.98)	0.52 (0.36 to 0.76)	0.56 (0.38 to 0.82)	0.57 (0.26 to 1.25)	0.54 (0.26 to 1.13)	0.57 (0.26 to 1.25)	0.64 (0.28 to 1.47)	0.63 (0.28 to 1.4)	0.64 (0.28 to 1.47)	0.63 (0.28 to 1.4)	0.64 (0.28 to 1.47)	0.63 (0.28 to 1.4)	0.64 (0.28 to 1.47)	
<b>Female‡</b>	Reference	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Cluster A	1.72 (1.18 to 2.51)	2.61 (1.62 to 4.20)	0.23 (0.14 to 0.38)	0.34 (0.19 to 0.60)	2.06 (1.15 to 3.71)	1.70 (0.89 to 3.23)	1.03 (0.59 to 1.80)	1.33 (0.74 to 2.39)	0.21 (0.1 to 0.42)	0.3 (0.15 to 0.62)	0.07 (0 to 1.91)	0.06 (0 to 1.69)	0.07 (0 to 1.91)	0.19 (0.01 to 3.68)	0.2 (0.01 to 5)	0.19 (0.01 to 3.68)	0.2 (0.01 to 5)	0.19 (0.01 to 3.68)	0.2 (0.01 to 5)	0.19 (0.01 to 3.68)	
Cluster B	2.47 (1.30 to 4.68)	3.42 (1.85 to 6.30)	0.19 (0.08 to 0.44)	0.21 (0.09 to 0.50)	2.22 (0.92 to 5.37)	2.21 (0.84 to 5.85)	0.59 (0.25 to 1.42)	0.71 (0.30 to 1.67)	0.31 (0.13 to 0.76)	0.35 (0.14 to 0.83)	0.33 (0.07 to 1.46)	0.37 (0.09 to 1.55)	0.33 (0.07 to 1.46)	1.4 (0.36 to 5.41)	1.28 (0.21 to 7.93)	1.4 (0.36 to 5.41)	1.28 (0.21 to 7.93)	1.4 (0.36 to 5.41)	1.28 (0.21 to 7.93)	1.4 (0.36 to 5.41)	
Cluster C	3.11 (2.01 to 4.79)	4.92 (3.02 to 8.00)	0.99 (0.60 to 1.65)	0.94 (0.55 to 1.62)	2.39 (1.25 to 4.58)	2.49 (1.32 to 4.72)	2.45 (1.45 to 4.11)	2.41 (1.41 to 4.14)	0.54 (0.32 to 0.9)	0.52 (0.32 to 0.85)	0.5 (0.19 to 1.31)	0.57 (0.24 to 1.37)	0.5 (0.19 to 1.31)	0.89 (0.38 to 2.09)	0.96 (0.42 to 2.2)	0.89 (0.38 to 2.09)	0.96 (0.42 to 2.2)	0.89 (0.38 to 2.09)	0.96 (0.42 to 2.2)	0.89 (0.38 to 2.09)	
<b>Newly diagnosed§</b>	Reference	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Cluster A	2.28 (1.28 to 4.07)	4.19 (1.98 to 8.91)	0.16 (0.06 to 0.42)	0.17 (0.06 to 0.51)	1.05 (0.56 to 1.96)	1.00 (0.52 to 1.91)	1.04 (0.57 to 1.89)	1.20 (0.63 to 2.30)	0.35 (0.21 to 0.58)	0.45 (0.26 to 0.76)	0.38 (0.18 to 0.82)	0.29 (0.13 to 0.64)	0.45 (0.26 to 0.76)	0.2 (0.04 to 1.02)	0.24 (0.06 to 1.05)	0.2 (0.04 to 1.02)	0.24 (0.06 to 1.05)	0.2 (0.04 to 1.02)	0.24 (0.06 to 1.05)	0.2 (0.04 to 1.02)	
Cluster B	1.72 (0.81 to 3.65)	1.77 (0.77 to 4.06)	0.14 (0.06 to 0.38)	0.16 (0.07 to 0.38)	0.64 (0.28 to 1.48)	0.67 (0.28 to 1.56)	0.26 (0.10 to 0.71)	0.41 (0.14 to 1.18)	0.21 (0.06 to 0.77)	0.23 (0.06 to 0.84)	0.13 (0 to 17.69)	0.1 (0 to 19.58)	0.23 (0.06 to 0.84)	0.42 (0.15 to 1.18)	0.46 (0.15 to 1.41)	0.42 (0.15 to 1.18)	0.46 (0.15 to 1.41)	0.42 (0.15 to 1.18)	0.46 (0.15 to 1.41)	0.42 (0.15 to 1.18)	
Cluster C	1.25 (0.66 to 2.38)	0.92 (0.45 to 1.86)	0.13 (0.04 to 0.43)	0.14 (0.04 to 0.51)	1.34 (0.69 to 2.60)	1.30 (0.68 to 2.48)	0.36 (0.16 to 0.81)	0.40 (0.17 to 0.94)	0.21 (0.09 to 0.51)	0.2 (0.08 to 0.47)	-\$	-\$	0.2 (0.08 to 0.47)	0.17 (0.02 to 1.44)	0.14 (0.01 to 1.46)	0.17 (0.02 to 1.44)	0.14 (0.01 to 1.46)	0.17 (0.02 to 1.44)	0.14 (0.01 to 1.46)	0.17 (0.02 to 1.44)	
<b>Already diagnosed*</b>	Reference	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Cluster A	2.42 (1.64 to 3.55)	3.43 (2.24 to 5.26)	0.31 (0.20 to 0.49)	0.46 (0.29 to 0.74)	1.50 (0.95 to 2.38)	1.44 (0.89 to 2.31)	0.76 (0.53 to 1.10)	1.06 (0.71 to 1.57)	0.31 (0.21 to 0.46)	0.39 (0.26 to 0.6)	0.24 (0.07 to 0.89)	0.22 (0.06 to 0.78)	0.39 (0.26 to 0.6)	0.14 (0.01 to 2.63)	0.14 (0.01 to 2.63)	0.24 (0.07 to 0.89)	0.14 (0.01 to 2.63)	0.24 (0.07 to 0.89)	0.14 (0.01 to 2.63)	0.24 (0.07 to 0.89)	
Cluster B	2.67 (1.61 to 4.42)	2.51 (1.43 to 4.41)	0.15 (0.06 to 0.38)	0.17 (0.06 to 0.45)	1.20 (0.66 to 2.20)	1.35 (0.69 to 2.64)	0.28 (0.13 to 0.58)	0.35 (0.17 to 0.73)	0.44 (0.24 to 0.83)	0.45 (0.25 to 0.82)	0.43 (0.18 to 1.03)	0.5 (0.21 to 1.21)	0.45 (0.25 to 0.82)	0.56 (0.17 to 1.78)	0.55 (0.15 to 2.05)	0.56 (0.17 to 1.78)	0.55 (0.15 to 2.05)	0.56 (0.17 to 1.78)	0.55 (0.15 to 2.05)	0.56 (0.17 to 1.78)	
Cluster C	4.24 (2.78 to 6.46)	4.87 (3.06 to 7.76)	1.21 (0.79 to 1.84)	1.02 (0.65 to 1.58)	1.36 (0.77 to 2.43)	1.61 (0.88 to 2.95)	1.84 (1.23 to 2.75)	1.66 (1.08 to 2.54)	1.33 (0.64 to 2.76)	0.78 (0.52 to 1.16)	1.26 (0.57 to 2.79)	1.33 (0.64 to 2.76)	0.78 (0.52 to 1.16)	1.06 (0.61 to 1.85)	1.05 (0.61 to 1.81)	1.26 (0.57 to 2.79)	1.06 (0.61 to 1.85)	1.26 (0.57 to 2.79)	1.06 (0.61 to 1.85)	1.26 (0.57 to 2.79)	

\*Adjust for sex, race, ever smoking, education level, total cholesterol, systolic blood pressure and duration of diabetes.

†Adjust for race, ever smoking, education level, total cholesterol, systolic blood pressure and duration of diabetes.

‡Adjust for sex, race, ever smoking, education level, total cholesterol and systolic blood pressure.

§Since there is too little cardiovascular mortality in cluster D, there is no calculation result.

CKD, chronic kidney disease; CVD, cardiovascular disease; CVH, cardiovascular health; NAFLD, non-alcoholic fatty liver disease.

Our study has several strengths. First, we used the results of five variables as a comparison. Second, we conducted a comparison on comprehensive risk prediction, including the prevalence of CVH, CKDs, NAFLD, advanced liver fibrosis at diagnosis, as well as the risk of mortality many years later. Last but not least, we explored the effects of different sexes and stages of diabetes on results. However, several limitations should be noted. First, though those younger than 30 were excluded, type 1 diabetes may also confound our results. Second, since the FIB-4 or NFS are based on metabolic variables, these scores may not perform well in diabetic patients compared with the general NAFLD population.<sup>39</sup> Third, there is no information on the risk of complication and the observation time of the mortality is not long enough. Further studies are needed to better understand the value of classification of T2D according to simple variables.

## CONCLUSIONS

In conclusion, a simple classification based on age, BMI and HbA1c could be used to identify T2D regarding several health and mortality risks. Since this method is clinically feasible in nearly all individuals with T2D, it has a great potential to be widely applied in clinical practice and allows more T2D patients to benefit from subgroup specific treatment paradigms.

### Author affiliations

<sup>1</sup>College of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu, China

<sup>2</sup>Department of Pharmacy, Southeast University Zhongda Hospital, Nanjing, Jiangsu, China

<sup>3</sup>Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu, China

<sup>4</sup>Department of Information, Southeast University Zhongda Hospital, Nanjing, Jiangsu, China

<sup>5</sup>Department of Cardiology, Jiangsu Province People's Hospital and Nanjing Medical University First Affiliated Hospital, Nanjing, Jiangsu, China

<sup>6</sup>Department of Cardiology, Southeast University Zhongda Hospital, Nanjing, Jiangsu, China

**Acknowledgements** We thank all participants who volunteered as part of the NHANES. We also thank Weiwei Duan for his help with research methods, and Wanke Cao and Min Zhang for their help with the revision of the manuscript. We also thank Guangzhou Huati United Translation Service for its contribution to the language polishing of the manuscript.

**Contributors** NFL had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JX, YL and NFL made substantial contributions to the conception and design of the work. TS, SJ, JH and RH collected and assembled the data. JX, JW and YL contributed to the analysis and interpretation of data for the work. JX, HS and YS drafted the manuscript. All authors gave final approval of the version to be published and agree to be responsible for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** This work was supported by grants from the big data industry development pilot demonstration project of Ministry of Industry and Information Technology of China [grant numbers 2019243, 202084].

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** All participants gave written informed consent prior to participation, and the methods were approved by the National Centre for Health Statistics.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iD

Jing Xie <http://orcid.org/0000-0002-8142-3237>

## REFERENCES

- 1 NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* 2016;387:1513–30.
- 2 Tuomi T, Santoro N, Caprio S, *et al.* The many faces of diabetes: a disease with increasing heterogeneity. *Lancet* 2014;383:1084–94.
- 3 Banks MP, Casanova R, Gregg EW, *et al.* Epidemiology of diabetes phenotypes and prevalent cardiovascular risk factors and diabetes complications in the National health and nutrition examination survey 2003–2014. *Diabetes Res Clin Pract* 2019;158:107915.
- 4 Ahlqvist E, Storm P, Käräjämäki A, *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–9.
- 5 Zaharia OP, Strassburger K, Strom A, *et al.* Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 2019;7:684–94.
- 6 Dennis JM, Shields BM, Henley WE, *et al.* Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019;7:442–51.
- 7 Anjana RM, Baskar V, Nair ATN, *et al.* Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the inspired study. *BMJ Open Diabetes Res Care* 2020;8:e001506.
- 8 Tanabe H, Saito H, Kudo A, *et al.* Factors associated with risk of diabetic complications in novel Cluster-Based diabetes subgroups: a Japanese retrospective cohort study. *J Clin Med* 2020;9:2083.
- 9 Zou X, Zhou X, Zhu Z, *et al.* Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol* 2019;7:9–11.
- 10 Dutta D, Mukhopadhyay S. Novel diabetes subgroups. *Lancet Diabetes Endocrinol* 2018;6:438.
- 11 Kahkoska AR, Geybels MS, Klein KR, *et al.* Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, leader and SUSTAIN-6 cardiovascular outcomes trials. *Diabetes Obes Metab* 2020;22:1537–47.
- 12 Ahlqvist E, Prasad RB, Groop L. Subtypes of type 2 diabetes determined from clinical parameters. *Diabetes* 2020;69:2086–93.
- 13 Centers for Disease Control and Prevention. National health and nutrition examination survey. Available: <https://www.cdc.gov/nchs/nhanes/index.htm> [Accessed 14 Nov 2020].
- 14 Palta P, Huang ES, Kalyani RR, *et al.* Hemoglobin A1c and mortality in older adults with and without diabetes: results from the National

- health and nutrition examination surveys (1988–2011). *Diabetes Care* 2017;40:453–60.
- 15 Lloyd-Jones DM, Hong Y, Labarthe D, *et al.* Defining and setting national goals for cardiovascular health promotion and disease reduction: the American heart association's strategic impact goal through 2020 and beyond. *Circulation* 2010;121:586–613.
  - 16 Zheng Y, Wen TS, Shen Y, *et al.* Age at menarche and cardiovascular health: results from the NHANES 1999–2016. *Menopause* 2021;28:18–24.
  - 17 Inker LA, Schmid CH, Tighiouart H, *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* 2012;367:20–9.
  - 18 Naqvi SH, Nunes AP. Age-Stratified Analysis of Associations Between Participant's Characteristics and NAFLD. *Mapping Intimacies* 2021.
  - 19 Ruhl CE, Everhart JE. Fatty liver indices in the multiethnic United States National health and nutrition examination survey. *Aliment Pharmacol Ther* 2015;41:65–76.
  - 20 Cho IY, Chang Y, Sung E, *et al.* Depression and increased risk of non-alcoholic fatty liver disease in individuals with obesity. *Epidemiol Psychiatr Sci* 2021;30:e23.
  - 21 Angulo P, Hui JM, Marchesini G, *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;45:846–54.
  - 22 Shah AG, Lydecker A, Murray K, *et al.* Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009;7:1104–12.
  - 23 Kim K-S, Lee Y-M, Lee I-K, *et al.* Paradoxical associations of insulin resistance with total and cardiovascular mortality in humans. *J Gerontol A Biol Sci Med Sci* 2015;70:847–53.
  - 24 Bernau C, Riestler M, Boulesteix A-L, *et al.* Cross-study validation for the assessment of prediction algorithms. *Bioinformatics* 2014;30:i105–12.
  - 25 Zipf G, Chiappa M, Porter KS, *et al.* National health and nutrition examination survey: plan and operations, 1999–2010. *Vital Health Stat 1* 2013;56:1–37.
  - 26 Menke A, Casagrande S, Geiss L, *et al.* Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA* 2015;314:1021.
  - 27 Guerrero-Romero F, Simental-Mendía LE, González-Ortiz M, *et al.* The product of triglycerides and glucose, a simple measure of insulin sensitivity. Comparison with the euglycemic-hyperinsulinemic clamp. *J Clin Endocrinol Metab* 2010;95:3347–51.
  - 28 Vasques ACJ, Novaes FS, de Oliveira MdaS, *et al.* TyG index performs better than HOMA in a Brazilian population: a hyperglycemic clamp validated study. *Diabetes Res Clin Pract* 2011;93:e98–100.
  - 29 Er L-K, Wu S, Chou H-H, *et al.* Triglyceride Glucose-Body mass index is a simple and clinically useful surrogate marker for insulin resistance in nondiabetic individuals. *PLoS One* 2016;11:e0149731.
  - 30 Ren X, Chen ZA, Zheng S, *et al.* Association between triglyceride to HDL-C ratio (TG/HDL-C) and insulin resistance in Chinese patients with newly diagnosed type 2 diabetes mellitus. *PLoS One* 2016;11:e0154345.
  - 31 Gaye B, Canonico M, Perier M-C, *et al.* Ideal Cardiovascular Health, Mortality, and Vascular Events in Elderly Subjects. *J Am Coll Cardiol* 2017;69:3015–26.
  - 32 Guo L, Zhang S. Association between ideal cardiovascular health metrics and risk of cardiovascular events or mortality: a meta-analysis of prospective studies. *Clin Cardiol* 2017;40:1339–46.
  - 33 Samieri C, Perier M-C, Gaye B, *et al.* Association of cardiovascular health level in older age with cognitive decline and incident dementia. *JAMA* 2018;320:657.
  - 34 Rasmussen-Torvik LJ, Shay CM, Abramson JG, *et al.* Ideal cardiovascular health is inversely associated with incident cancer: the Atherosclerosis risk in Communities study. *Circulation* 2013;127:1270–5.
  - 35 Bardagjy AS, Steinberg FM. Relationship between HDL functional characteristics and cardiovascular health and potential impact of dietary patterns: a narrative review. *Nutrients* 2019;11:1231.
  - 36 Willis LH, Slentz CA, Bateman LA, *et al.* Effects of aerobic and/or resistance training on body mass and fat mass in overweight or obese adults. *J Appl Physiol* 2012;113:1831–7.
  - 37 Shih H-M, Chuang S-M, Lee C-C, *et al.* Addition of metabolic syndrome to albuminuria provides a new risk stratification model for diabetic kidney disease progression in elderly patients. *Sci Rep* 2020;10:6788.
  - 38 Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* 2016;65:1038–48.
  - 39 Ciardullo S, Monti T, Perseghin G. High prevalence of advanced liver fibrosis assessed by transient elastography among U.S. adults with type 2 diabetes. *Diabetes Care* 2021;44:519–25.