Novel Clusters of Newly-diagnosed Type 2 Diabetes and Their Association with Diabetic Retinopathy: a 3-year follow-up study

Key words: Type 2 diabetes; Cluster analysis; Diabetic retinopathy; Hyperglycemia
Running title: Novel Clusters of T2D and their Association with DR

Yu Liu[1*], Miaomiao Sang[1*], Yang Yuan[1], Ziwei Du[1], Wei Li [2], Hao Hu[1], Liang Wen[3], Fenghua Wang[4], Haijian Guo[5], Bei Wang [6], Duolao Wang[7], Zilin Sun[1#], Shanhu Qiu[8#]

[1] Department of Endocrinology, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, Nanjing, China

[2] Department of Endocrinology, Suzhou Hospital of Anhui Medical University (Suzhou Municipal Hospital of Anhui Province), Suzhou, China

[3] Department of Ophthalmology, Fushun Eye Hospital, Fushun, China

[4] Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University and Beijing Ophthalmology and Visual Science Key Lab, Beijing, China

[5] Jiangsu Provincial Centre for Disease Control and Prevention, Nanjing, China

[6] Key Laboratory of Environmental Medicine and Engineering of Ministry of Education, Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, China

[7] Liverpool School of Tropical Medicine, Liverpool, United Kingdom

[8] Department of General Practice, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, Nanjing, China

[*] Contributed equally.

[#]Corresponding authors:

Shanhu Qiu, Department of General Practice, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, No.87 Dingjiaqiao Street, Nanjing 210009, P.R. China. **E-mail**: tigershanhu@126.com

or Zilin Sun, Department of Endocrinology, Zhongda Hospital, Institute of Diabetes, School of Medicine, Southeast University, No.87 Dingjiaqiao Street, Nanjing 210009, P.R. China. **E-mail**: sunzilin1963@126.com

DECLARATIONS

Conflicts of interest

We declare that we have no conflicts of interest.

Founding

This study was supported by the National Key R&D Program of China (2016YFC1305700), the National Key Scientific Instrument and Equipment Development Project of China (No. 51627808), and the Nanjing Special Fund for Health Science and Technology Development (No. YKK18261). Shanhu Qiu has been supported by the "Best Young Scholars" Fellowship from Southeast University. The funders had no roles in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

AUTHORS' CONTRIBUTIONS

Shanhu Qiu, Zilin Sun, Yu Liu and Miaomiao Sang were responsible for the study design. Zilin Sun, Haijian Guo and Bei Wang were responsible for the recruitment of participants, data acquisition, and the quality control of the study. Yu Liu and Miaomiao Sang were responsible for data collection, data analysis, interpretation, and the writing of the manuscript. Yang Yuan, Ziwei Du, Wei Li and Hao Hu assisted with data interpretation and manuscript modification. Liang Wen and Fenghua Wang were responsible for visual acuity measurement and its interpretation. Duolao Wang provided assistance with statistical analysis and quality control. All authors have read the manuscript critically and approved the submitted version.

ABSTRACT

Background

Cluster analysis may assist in stratifying heterogeneous clinical presentations of type 2 diabetes (T2D). However, the association of cluster-based subgroups with diabetes-related outcomes such as diabetic retinopathy remains unclear. This study was aimed to address this issue with novel clusters of T2D derived from four simple parameters.

Method

We developed a k-means clustering model in participants with newly diagnosed T2D (N=1,910) from the SENSIBLE and SENSIBLE-Addition studies, based on body mass index (BMI), waist circumference (WC), mean arterial pressure (MAP), and hemoglobin A1c (HbA1c). Diabetic retinopathy was ascertained with the protocol from the Early Treatment of Diabetic Retinopathy Study. Participants (N=515) without diabetic retinopathy at baseline were followed-up for 3 years. Logistic regression analyses were performed to obtain the odds ratios (ORs) and 95% confidence intervals (CIs).

Results

Three clusters were identified, with cluster 0, 1 and 2 accounting for 48.2%, 8.9% and 42.9%, respectively. Participants with T2D were featured by the lowest BMI, WC, MAP, and HbA1c in cluster 0, poor glycemic condition in cluster 1, and the highest BMI, WC, and MAP in cluster 2. Compared with cluster 0, cluster 1 was associated with increased odds of diabetic retinopathy in both the cross-sectional study (OR 6.25, 95% CI: 3.19-12.23) and the cohort study (OR 9.16, 95% CI: 2.08-40.34), while cluster 2 was not. Moreover, most participants remained their clusters unchanged during follow-up.

Conclusions

Our cluster-based analysis showed that participants with poor glycemic condition rather than high blood pressure and obesity had higher risk of diabetic retinopathy.

INTRODUCTION

The prevalence of diabetes has been increased over the past decades in China, which is reported to be 12.8% in 2015 from 2.5% in 1994 [1, 2]. Type 2 diabetes (T2D) is a highly heterogeneous disease with substantial variations in clinical presentations and prognostic consequences [3-5], and accounts for more than 90% of all the cases with diabetes [6]. As a result, refining diabetes, in particular T2D, with different phenotypes, is of clinical importance to implement individualized preventive and/or therapeutic interventions.

Cluster analysis has emerged as an effective approach in stratifying diabetes into different phenotypes in recent years. In 2018, Ahlqvist *et al*. identified 5 clusters of adult-onset diabetes, using glutamate decarboxylase antibodies (GADA), age at diagnosis, body mass index (BMI), hemoglobin A1c (HbA1c), and homoeostatic model assessment for β-cell function and insulin resistance based on fasting plasma glucose (FPG) and C-peptide [7]. Their work, together with others, have consistently showed that clusters of diabetes including severe insulin deficiency diabetes and severe insulin resistant diabetes were related to increased risk of diabetic complications such as diabetic retinopathy and diabetic kidney disease [8-13]. However, GADA and C-peptide, the key parameters employed in these cluster analyses, were seldomly assessed in clinical practice or epidemiological surveys, limiting their wide use potentially.

In contrast, parameters such as BMI, waist circumference (WC), systolic blood pressure (SBP), diastolic blood pressure (DBP), and HbA1c, are routinely measured. However, no studies have investigated whether clusters of diabetes derived from these parameters would have any association with diabetes-related outcomes such as diabetic retinopathy, which is the leading cause of blindness in working-age population [14, 15]. To address this issue, we firstly conducted cluster analysis based on the Study on Evaluation of iNnovated Screening tools and determination of optimal diagnostic cut-off points for type 2 diaBetes in Chinese muLti-Ethnic (SENSIBLE) and SENSIBLE-Addition studies. Secondly, we assessed whether the derived clusters were associated with diabetic retinopathy ascertained by the gold-standard approach for visual acuity measurement – the Early Treatment of Diabetic Retinopathy Study (ETDRS) protocol [16], in SENSIBLE-cohort study.

METHODS

Study design and population

Data for cluster analysis in our study were driven from the SENSIBLE [17] and SENSIBLE-Addition studies [18]. The details of the sampling approaches and the designs of both studies were described elsewhere [17, 18]. Participants aged more than 70 years were also included in the present study to obtain a larger sample size. Among the 17,629 participants surveyed, 1,927 were identified as newly diagnosed T2D. After excluding participants with missing information on blood pressure, height, or body weight, or having data considered extreme outliers (> 99.9 or < 0.1 percentiles) (n = 17), a total of 1,910 participants were included for the cluster analysis (Supplementary figure 1A).

SENSIBLE-cohort study was a 3-year longitudinal survey, which was designed to determine the optimal cut-off value of advanced glycation end-products and HbA1c for diagnosing T2D in China. A total of 7,600 participants from the SENSIBLE [17] and SENSIBLE-Addition [18] studies were randomly selected as the baseline dataset, and invited to be followed-up at a 1.5-year interval around for 3 years. In this cohort study, the first follow-up survey was started in July 2018 (enrolling 5,664 participants with a response rate of 74.5%), and the second in May 2020 (enrolling 6,048 participants with a response rate of 79.6%). Upon the exclusion of participants without diabetes or with known diabetes at baseline (n = 5,290), 758 participants were included. After further exclusion of participants without fundus imaging (n = 187), with unqualified fundus imaging (n = 15), or having diabetic retinopathy at baseline (n = 41), 515 were finally eligible for the analysis on the risk of incident diabetic retinopathy in relation to the clusters of diabetes (Supplementary figure 1B).

The protocols of the studies were approved by the ethics committee of Zhongda Hospital, Southeast University, and the involved sub-center hospitals. Written informed consent was obtained from each participant.

Procedure

Participants at follow-ups underwent the same procedure as baseline [17]. Information on sociodemographic parameters, medical history, and lifestyle factors were obtained using questionnaires by trained interviewers. Body weight, height, and WC were measured using standardized protocols. SBP and DBP were measured trice at a seated position after 5-minute rest using an automated device (YE680E, *yuwell*, China). Two fundus images centered on the optic disc and macular area were taken for each eye from

each participant using Canon non-mydriatic digital fundus camera (model: CR-2 AF). However, tropicamide or phenylephrine was used to dilate pupil, in the case that fundus images cannot be well taken using the non-mydriatic digital fundus camera. All fundus images were sent to Beijing Tongren Hospital for ascertainment of diabetic retinopathy by two independent professional ophthalmologists in a blind manner.

All participants were asked to fast for at least 10 hours, and fasting blood samples were taken for the measurement of HbA1c, FPG, triglyceride (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), serum uric acid (SUA), and serum creatinine (SCr). Afterwards, all participants were instructed to swallow a standard 75 g glucose solution, with blood samples taken 2 hours later again, for the 2-hour postprandial glucose (2hPG). All blood samples were centrifuged within 30 minutes after collection, and shipped at 4 °C by air to the central laboratory in Nanjing Adicon Clinical Laboratories. FPG, 2hPG, TG, TC, HDL-C, LDL-C, SUA, and SCr were measured using an automatic chemistry analyzer (Synchron LX-20, Beckman Coulter Inc., CA, USA), and HbA1c was by high-performance liquid chromatography (D-10™ Hemoglobin Analyzer, Bio-Rad Inc., CA, USA).

Definitions

Diabetes was defined according to the criteria from the American Diabetes Association: 1) FPG ≥ 7.0 mmol/L; 2) 2hPG ≥ 11.1 mmol/L; or 3) HbA1c ≥ 6.5% (48 mmol/mol) [19]. BMI was calculated as body weight (kg) divided by squared height (m). MAP was calculated as [SBP (mmHg) + 2 × DBP (mmHg)]/3. Kidney function was evaluated by the estimated glomerular filtration rate (eGFR) according to the Chronic Kidney Disease Epidemiology Collaboration equation [20, 21].

Diabetic retinopathy ascertainment

Diabetic retinopathy was ascertained based on the scoring system of ETDRS protocol. Participants were considered with diabetic retinopathy if ETDRS score was ≥ 20 for any eye. Moreover, minimal diabetic retinopathy was defined as EDTRS score as 20, mild as 31, moderate as 41, severe non-proliferative as 51, and proliferative as >60 [22].

Cluster analysis and visualization

The clusters of diabetes were generated using the cluster analysis by built-in k-means algorithm in the Scikit-learn library of Python 3. Specifically, we normalized the four numerical variables with Z-score method, and k-means function (max iteration =300,

initialization = 'k-means++') was then applied. The optimal number of clusters was determined based on silhouette width (Supplementary figure 2A), elbow method (Supplementary figure 2B), and Two-Step clustering method (Supplementary figure 2C). In this study, Two-Step clustering method was done in SPSS 26.0 (SPSS Inc., Chicago, IL, USA) from 2 to 15 clusters using log-likelihood as a distance measure and Schwarz's Bayesian criterion for clustering. The k-means algorithm was used again to assess the cluster redistribution patterns at 3-year follow-up.

Since the data from participants were distributed in a four-dimensional space, it was impractical to validate the clustering results from k-means intuitively. We therefore utilized the t-distributed stochastic neighbor embedding (tSNE) visualization technique to transform the normalized data into a two-dimensional space for better visualization. The tSNE analysis was conducted with the Scikit-learn library of Python, whereas the plot was visualized by the Matplotlib library.

Statistical analysis

Continuous data are presented as means and standard deviations (SDs) and compared using one-way analysis of variance, while categorical variables are presented as numbers and percentages and compared using χ^2 test. Odds ratios (ORs) and 95% confidence intervals (CIs) were obtained by logistic regression analyses to assess the association of different clusters with diabetic retinopathy, based on the following models: model 1 – crude model; model 2 – controlled for gender, age, and ethnic groups; and model 3 – additionally controlled for SUA, TC, TG, LDL-C, HDL-C, eGFR, and smoking and drinking status. These analyses were performed using SPSS 26.0 (SPSS Inc., Chicago, IL, USA). A two-tailed $P < 0.05$ was considered statistically significant.

RESULTS

Cluster distribution in cross-sectional study

The characteristics of 1,910 participants with newly diagnosed T2D from the SENSIBLE and SENSIBLE-Addition studies are shown in Table 1. Cluster analysis derived from variables including BMI, WC, MAP and HbA1c identified 3 clusters (Table 1 and Figure 1), with 48.2% in cluster 0, 8.9% in cluster 1, and 42.9% in cluster 2. Cluster 0 showed the lowest mean values of all the clustering variables, while cluster 1 was characterized by the highest mean value of HbA1c, and cluster 2 by the highest mean values of WC, BMI and MAP. Participants in cluster 1 were the youngest, and

had the lowest SUA levels, the highest eGFR, and the worst lipid metabolism, compared with cluster 0 and cluster 2. Details about the cluster centers which could be used for the stratification of different clusters, are provided in Supplementary table 1.

The distribution patterns of these 3 clusters in each ethnic group are presented in Figure 2. Han, Zhuang, Dai and Korean ethnic groups showed similar distribution patterns as the whole dataset, while Kazak and Uyghur groups had the largest proportions of cluster 2 but the lowest proportions of cluster 0.

The risk of diabetic retinopathy with different clusters in cross-sectional and cohort studies

In our cross-sectional survey, 1,059 participants provided eligible fundus imaging and were included for the present analysis (Supplementary figure 1A). Participants in cluster 1 showed the highest prevalence of diabetic retinopathy (26.3%) compared with those in cluster 0 (7.1%) and cluster 2 (6.5%). Using cluster 0 as the referent, the OR of diabetic retinopathy in cluster 1 was 4.70 (95% CI: 2.56 - 8.63) in the crude model (model 1), and 6.25 (95% CI: 3.19 - 12.23) after multivariable-adjustment (model 3, Table 3). No significant association between diabetic retinopathy and cluster 2 was observed in any models (Table 3).

In our cohort study, 515 participants without diabetic retinopathy at baseline were followed-up for 3 years (Supplementary figure 1). Of them, 24 developed incident diabetic retinopathy. The baseline characteristics of this subsample based on the clusters of diabetes are shown in Table 2. During follow-up, the incidence of diabetic retinopathy was the highest (14.8%) in cluster 1 compared to cluster 0 (3.4%) and cluster 2 (4.9%). Comparing with cluster 0, the OR of incident diabetic retinopathy was 9.16 (95% CI: 2.08 - 40.34) in cluster 1 after multivariable-adjustment (model 3, Table 3). However, there was no significant association between incident diabetic retinopathy with cluster 2 (Table 3).

Cluster redistribution with diabetic retinopathy in cohort study

The patterns of cluster redistribution from the 513 eligible participants during the 3-year follow-up are shown in Figure 3 and Supplementary table 2. Two participants from 515 at baseline were excluded because of missing information on clustering variables. The proportions of different clusters were generally similar to those at baseline. Most participants remained their clusters unchanged during follow-up. However, 123 (24.0%) participants switched their cluster allocation, with 50 (19.2%), 16 (61.5%), and 57

(21.8%) in cluster 0, 1 and 2, respectively.

The incidence of diabetic retinopathy at the 3-year follow-up in relation to cluster redistribution is shown in Supplementary figure 3 and Supplementary table 3. It appears that participants switched to cluster 2 from cluster 1 had the highest risk of diabetic retinopathy (33.3%) at follow-up compared with other switchers.

DISCUSSION

In this study we showed several findings in Chinese population with newly diagnosed T2D that: (i) 3 clusters were identified, with cluster 0 featured by mild metabolic condition, cluster 1 by poor glycemic condition, and cluster 2 by high blood pressure and obesity; (ii) participants in cluster 1 had higher risk of diabetic retinopathy compared with those in cluster 0; (iii) the majority of participants remained their cluster pattern unalerted during follow-up.

Previous studies that used cluster analysis for the stratification of diabetes subgroups based on GADA, age, BMI, HbA1c, FPG, and C-peptide were shown to be effective in predicting future health outcomes including microvascular and macrovascular diseases [7-13]. However, C-peptide included in these studies for T2D clusters was not routinely assessed in clinical practice, and the accuracy of the measurement approach for GADA using enzyme-linked immunosorbent assay has been questioned compared with radioligand binding assay[23]. These concerns may therefore limit the wide use of such clusters. Unlike C-peptide and GADA, parameters such as BMI, WC, and blood pressure could be easily acquired and their usage in cluster analysis was shown to be effective in the stratification for diabetes [24]. However, in our study MAP, instead of SBP and DBP, was employed for cluster analysis. This was done to simplify the cluster analysis. Moreover, to provide more accurate clusters of T2D, we included HbA1c for cluster analysis - a commonly measured parameter for newly diagnosed diabetes to reflect glycemic control during the past 2-3 months [25].

In our study 3 clusters were identified. Cluster 1 grouped participants with the highest HbA1c, which indicated that they might have diabetes undiagnosed for a long period. Cluster 0 was featured by the lowest BMI, WC, MAP and HbA1c, while cluster 2 ranked first with BMI, WC and MAP. There is evidence that T2D was developed in Chinese adults at a considerably lower BMI compared with western populations [26]: the average BMI of newly diagnosed diabetes in Chinese was around 23.7 kg/m^2, which

was more than 27 kg/m² in Americans [27]. This is consistent with our finding that the majority of participants were identified as cluster 0. However, there seemed to be a heterogenous distribution of clusters across different ethnic groups (e.g., participants from Uyghur and Kazak ethnicities were more likely to fall into cluster 2). This might be attributed to the differences in lifestyle factors and genetics (e.g., participants from Uyghur and Kazak ethnicities have admixed ancestral components of East Asian and European populations [28, 29]).

In this study we also found that about one-fourth of the participants switched their cluster allocation at follow-up. Of note, half of the participants in cluster 1 represented by poor glycemic condition switched to cluster 0 characterized by mild metabolic condition. This implies that achieving good glycemic control was generally emphasized in clinical practice, in particular for newly diagnosed participants with poor glycemic condition. However, some participants in cluster 1 remained their clusters unchanged at follow-up (that is, characterized by poor glycemic control). It is speculated that these participants did not pay enough attention to manage their diabetes, at least for glycemic control. Moreover, this may also reflect the possibility of their disease progression.

Our study showed that participants in cluster 2 represented by high blood pressure and obesity did not have significantly increased odds of diabetic retinopathy, while participants in cluster 1 characterized by poor glycemic condition had. This is in support of the finding that poor glycemic control may lead to ocular complication as suggested previously [14, 30]. Furthermore, participants remained in cluster 1 or switched from cluster 2 to cluster 1 during follow-up showed high incidence of diabetic retinopathy. These, taken together, support that hyperglycemia might play a more important role in the development of diabetic retinopathy than hypertension or obesity in adults with newly diagnosed T2D [31].

The strengths of our study included a prospective cohort design and the use of ETDRS protocol - the gold standard approach for accurate visual acuity measurement to assess diabetic retinopathy [16]. However, our study has several limitations. Firstly, the diagnosis of T2D in our study was clinically judged based on age, BMI and disease history; however, we could not exclude the possibility that some participants with type 1 diabetes might be misclassified as T2D, although there is evidence the prevalence of type 1 diabetes is only about 5.49% in China [6]. Secondly, the sample size of our study was small, in particular for the cohort analysis. Future studies with larger sample sizes

are required to confirm our findings. Thirdly, our cluster analysis was conducted based on 4 simple variables. However, we did not assess whether the use of some other easy-obtained parameters such as physical activity would add any incremental benefit in stratifying the subgroups of T2D. Finally, the clinical significance of our refined T2D subgroups was only evaluated based on diabetic retinopathy, it remains unclear whether it is related to other diabetes-related outcomes such as cardiovascular events or diabetic kidney disease.

In conclusion, our study suggested that novel cluster-based subgroups of diabetes based on simple and cost-effective indicators could be employed for risk stratification for diabetic retinopathy in Chinese adults with newly diagnosed T2D. Participants characterized by poor glycemic condition rather than high blood pressure and obesity showed higher risk for diabetic retinopathy. Future studies with cohort design and larger sample sizes are needed to confirm our findings.

ACKNOWLEDGEMENT

None.

REFERENCE

1. Pan XR, Yang WY, Li GW, Liu J (1997) Prevalence of diabetes and its risk factors in China, 1994. National Diabetes Prevention and Control Cooperative Group. Diabetes Care 20(11):1664-1669
2. Li Y, Teng D, Shi X, Qin G, Qin Y, Quan H, Shi B, Sun H, Ba J, Chen B et al (2020) Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. Bmj 369:m997
3. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L (2014) The many faces of diabetes: a disease with increasing heterogeneity. Lancet 383(9922):1084-1094
4. Fitipaldi H, McCarthy MI, Florez JC, Franks PW (2018) A Global Overview of Precision Medicine in Type 2 Diabetes. Diabetes 67(10):1911-1922
5. Bowman P, Flanagan SE, Hattersley AT (2018) Future Roadmaps for Precision Medicine Applied to Diabetes: Rising to the Challenge of Heterogeneity. J Diabetes Res 2018:3061620

6. Tang X, Yan X, Zhou H, Yang X, Niu X, Liu J, Ji Q, Ji L, Li X, Zhou Z (2019) Prevalence and identification of type 1 diabetes in Chinese adults with newly diagnosed diabetes. Diabetes Metab Syndr Obes 12:1527-1541
7. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, Vikman P, Prasad RB, Aly DM, Almgren P et al (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 6(5):361-369
8. Zou X, Zhou X, Zhu Z, Ji L (2019) Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. Lancet Diabetes Endocrinol 7(1):9-11
9. Zaharia OP, Strassburger K, Strom A, Bönhof GJ, Karusheva Y, Antoniou S, Bódis K, Markgraf DF, Burkart V, Müssig K et al (2019) Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol 7(9):684-694
10. Anjana RM, Baskar V, Nair ATN, Jebarani S, Siddiqui MK, Pradeepa R, Unnikrishnan R, Palmer C, Pearson E, Mohan V (2020) 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 8(1):e001506
11. Tanabe H, Saito H, Kudo A, Machii N, Hirai H, Maimaituxun G, Tanaka K, Masuzaki H, Watanabe T, Asahi K et al (2020) Factors Associated with Risk of Diabetic Complications in Novel Cluster-Based Diabetes Subgroups: A Japanese Retrospective Cohort Study. J Clin Med 9(7):2083
12. Xiong XF, Yang Y, Wei L, Xiao Y, Li L, Sun L (2021) Identification of two novel subgroups in patients with diabetes mellitus and their association with clinical outcomes: A two-step cluster analysis. J Diabetes Investig 12(8):1346-1358
13. Wang W, Pei X, Zhang L, Chen Z, Lin D, Duan X, Fan J, Pan Q, Guo L (2021) Application of new international classification of adult-onset diabetes in Chinese inpatients with diabetes mellitus. Diabetes Metab Res Rev 37(7):e3427
14. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, Chen SJ, Dekker JM, Fletcher A, Grauslund J et al (2012) Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 35(3):556-564

15. Cheung N, Mitchell P, Wong TY (2010) Diabetic retinopathy. Lancet 376(9735):124-136
16. Pang Y, Sparschu L, Nylin E (2020) Validation of an automated-ETDRS near and intermediate visual acuity measurement. Clin Exp Optom 103(5):663-667
17. Li W, Xie B, Qiu S, Huang X, Chen J, Wang X, Li H, Chen Q, Wang Q, Tu P et al (2018) Non-lab and semi-lab algorithms for screening undiagnosed diabetes: A cross-sectional study. EBioMedicine 35:307-316
18. Qiu S, Du Z, Li W, Chen J, Wu H, Liu J, Cai M, Wang B, Guo H, Sun Z (2020) Exploration and Validation of the Performance of Hemoglobin A1c in Detecting Diabetes in Community-Dwellers With Hypertension. Ann Lab Med 40(6):457-465
19. American Diabetes Association (2021) 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2021. Diabetes Care 44(Suppl 1):S15-s33
20. Inker LA, Astor BC, Fox CH, Isakova T, Lash JP, Peralta CA, Kurella Tamura M, Feldman HI (2014) KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. Am J Kidney Dis 63(5):713-735
21. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T et al (2009) A new equation to estimate glomerular filtration rate. Ann Intern Med 150(9):604-612
22. Early Treatment Diabetic Retinopathy Study Research Group (1991) Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Ophthalmology 98(5 Suppl):786-806
23. Törn C, Mueller PW, Schlosser M, Bonifacio E, Bingley PJ (2008) Diabetes Antibody Standardization Program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2. Diabetologia 51(5):846-852
24. Carrillo-Larco RM, Castillo-Cara M, Anza-Ramirez C, Bernabé-Ortiz A (2021) Clusters of people with type 2 diabetes in the general population: unsupervised machine learning approach using national surveys in Latin America and the Caribbean. BMJ Open Diabetes Res Care 9(1):e001889
25. Campbell L, Pepper T, Shipman K (2019) HbA1c: a review of non-glycaemic

variables. J Clin Pathol 72(1):12-19

26. Ma RC, Chan JC (2013) Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci 1281(1):64-91

27. Ma RC, Lin X, Jia W (2014) Causes of type 2 diabetes in China. Lancet Diabetes Endocrinol 2(12):980-991

28. Xu S, Huang W, Qian J, Jin L (2008) Analysis of genomic admixture in Uyghur and its implication in mapping strategy. Am J Hum Genet 82(4):883-894

29. Xie T, Shen C, Liu C, Fang Y, Guo Y, Lan Q, Wang L, Ge J, Zhou Y, Wen S et al (2020) Ancestry inference and admixture component estimations of Chinese Kazak group based on 165 AIM-SNPs via NGS platform. J Hum Genet 65(5):461-468

30. Lee R, Wong TY, Sabanayagam C (2015) Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. Eye Vis (Lond) 2:17

31. Song P, Yu J, Chan KY, Theodoratou E, Rudan I (2018) Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis. J Glob Health 8(1):010803

Tables

Table 1. Characteristics of cluster-based subgroups of diabetes in cross-sectional study [a].

	Cluster 0 [b]	**Cluster 1 [c]**	**Cluster 2 [d]**	***P* [e]**
N (%)	921(48.2)	170(8.9)	819(42.9)	
Gender (male) , N (%)	333(36.2)	75(44.1)	396(48.4)	<0.001
Age, years	56 ± 9	54 ± 9	56 ± 9	0.001
Height, cm	158.5 ± 8.1	160.4 ± 9.0	161.0 ± 8.3	<0.001
Weight, kg	60.6 ± 8.5	66.8 ± 12.5	77.4 ± 10.8	<0.001
Smoking status (smoker), N (%)	173(18.9)	43(25.6)	174(21.4)	0.10
Drinking status (drinker), N (%)	233(25.4)	52(31.0)	254(31.2)	0.02
SBP, mmHg	137 ± 19	139 ± 19	149 ± 20	<0.001
DBP, mmHg	82 ± 11	86 ± 13	90 ± 12	<0.001
WC, cm	80.8 ± 7.4	87.6 ± 9.7	96.5 ± 8.0	<0.001
BMI, kg/m^2	24.1 ± 2.5	25.9 ± 3.6	29.8 ± 3.3	<0.001
MAP, mmHg	100 ± 12	103 ± 14	110 ± 13	<0.001
HbA1c, %	6.2 ± 0.8	10.8 ± 1.8	6.5 ± 0.9	<0.001
HbA1c, mmol/mol	45 ± 9	94 ± 19	47 ± 10	<0.001
FPG, mmol/L	6.9 ± 1.3	13.5 ± 3.7	7.3 ± 1.6	<0.001
2hPG, mmol/L	12.0 ± 3.6	24.1 ± 6.3	12.5 ± 3.6	<0.001
SUA, μmol/L	318 ± 85	280 ± 79	349 ± 96	<0.001
SCr, mmol/L	62.5 ± 18.5	59.8 ± 16.5	66.7 ± 16.9	<0.001
eGFR, ml/min/1.73m^2	98 ± 15	103 ± 14	97 ± 16	<0.001
TC, mmol/L	5.4 ± 1.2	5.6 ± 1.3	5.4 ± 1.2	0.03
LDL-C, mmol/L	3.0 ± 0.9	3.2 ± 0.9	3.1 ± 0.9	0.04
TG, mmol/L	2.0 ± 1.8	3.0 ± 3.7	2.5 ± 2.3	<0.001
HDL-C, mmol/L	1.6 ± 0.4	1.4 ± 0.4	1.4 ± 0.3	<0.001

SBP, systolic blood pressure; DBP, diastolic blood pressure; WC, waist circumference; BMI, body mass index; MAP, mean arterial pressure; HbA1c, hemoglobin A1c; FPG, fasting plasma glucose; 2hPG, 2-hour postprandial glucose; SUA, serum uric acid; Scr, serum creatine; eGFR, estimated glomerular filtration rate; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol

[a] It was analysed based on 1,910 participants.

[b] Featured by the lowest blood glucose, blood pressure, waist circumference and body mass index.

[c] Featured by poor glycemic condition.

[d] Featured by the highest blood pressure, waist circumference and body mass index.

[e] Continuous variables (presented as means ± standard deviations) across groups were compared by one-way analysis of variance, and categorical variables [presented as numbers (percentages)] were compared by χ^2 test.

Table 2. Baseline characteristics of cluster-based subgroups of diabetes in cohort study [a].

	Cluster 0 [b]	Cluster 1 [c]	Cluster 2 [d]	*P* [e]
N, %	262(50.9)	27(5.2)	226(43.9)	
Gender (male) , N (%)	94(35.9)	13(48.1)	106(46.9)	0.04
Age, years	56 ± 8	53 ± 7	55 ± 8	0.13
Height, cm	159.0 ± 7.6	160.3 ± 9.8	161.1 ± 7.7	0.02
Weight, kg	62.2 ± 8.2	65.0 ± 9.7	77.3 ± 10.2	<0.001
Smoking status (smoker), N (%)	53(20.3)	6(22.2)	59(26.1)	0.30
Drinking status (drinker), N (%)	69(26.3)	6(22.2)	79(35.0)	0.08
SBP, mmHg	136 ± 18	135 ± 13	149 ± 19	<0.001
DBP, mmHg	82 ± 11	88 ± 10	90 ± 11	<0.001
WC, cm	81.3 ± 7.3	84.7 ± 9.2	95.4 ± 7.4	<0.001
BMI, kg/m^2	24.6 ± 2.4	25.2 ± 2.9	29.8 ± 3.3	<0.001
MAP, mmHg	100 ± 12	103 ± 10	110 ± 12	<0.001
HbA1c, %	6.2 ± 0.8	10.3 ± 1.6	6.4 ± 0.8	<0.001
HbA1c, mmol/mol	44 ± 8	89 ± 17	46 ± 9	<0.001
FPG, mmol/L	6.8 ± 1.1	12.3 ± 2.9	7.2 ± 1.4	<0.001
2hPG, mmol/L	11.8 ± 3.4	24.1 ± 4.4	12.3 ± 3.4	<0.001
SUA, μmol/L	308 ± 79	273 ± 74	334 ± 95	<0.001
SCr, mmol/L	61.6 ± 16.7	62.1 ± 20.0	64.5 ± 15.7	0.61
eGFR, ml/min/1.73m^2	99 ± 13	102 ± 14	99 ± 14	0.51
TC, mmol/L	5.2 ± 1.0	5.6 ± 1.1	5.3 ± 1.1	0.07
LDL-C, mmol/L	2.9 ± 0.8	3.3 ± 0.9	3.1 ± 0.9	0.001
TG, mmol/L	2.0 ± 2.0	2.6 ± 3.1	2.4 ± 1.7	0.02
HDL-C, mmol/L	1.6 ± 0.4	1.3 ± 0.4	1.4 ± 0.3	<0.001

SBP, systolic blood pressure; DBP, diastolic blood pressure; WC, waist circumference; BMI, body mass index; MAP, mean arterial pressure; HbA1c, hemoglobin A1c; FPG, fasting plasma glucose; 2hPG, 2-hour postprandial glucose; SUA, serum uric acid; Scr, serum creatine; eGFR, estimated glomerular filtration rate; TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol

[a] It was analysed based on 515 participants included at baseline.

[b] Featured by the lowest blood glucose, blood pressure, waist circumference and body mass index.

[c] Featured by poor glycemic condition.

[d] Featured by the highest blood pressure, waist circumference and body mass index.

[e] Continuous variables (presented as means ± standard deviations) across groups were compared by one-way analysis of variance, and categorical variables [presented as numbers (percentages)] were compared by χ^2 test.

Table 3. The association of different clusters with diabetic retinopathy in cross-sectional and cohort studies.

	Cases, n(%)	OR[a] (95%CI)	OR[b] (95%CI)	OR[c] (95%CI)
Odds of diabetic retinopathy in cross-sectional study based on 1,059 participants				
Cluster 0	38(7.1)	1	1	1
Cluster 1	20(26.3)	4.70(2.56-8.63)	5.28(2.80-9.94)	6.25(3.19-12.23)
Cluster 2	29(6.5)	0.92(0.56-1.51)	0.88(0.52-1.48)	0.97(0.56-1.68)
Odds of diabetic retinopathy in cohort study based on 515 participants				
Cluster 0	9(3.4)	1	1	1
Cluster 1	4(14.8)	4.89(1.40-17.11)	5.58(1.52-20.47)	9.16(2.08-40.34)
Cluster 2	11(4.9)	1.44(0.59-3.54)	1.45(0.57-3.65)	1.79(0.66-4.85)

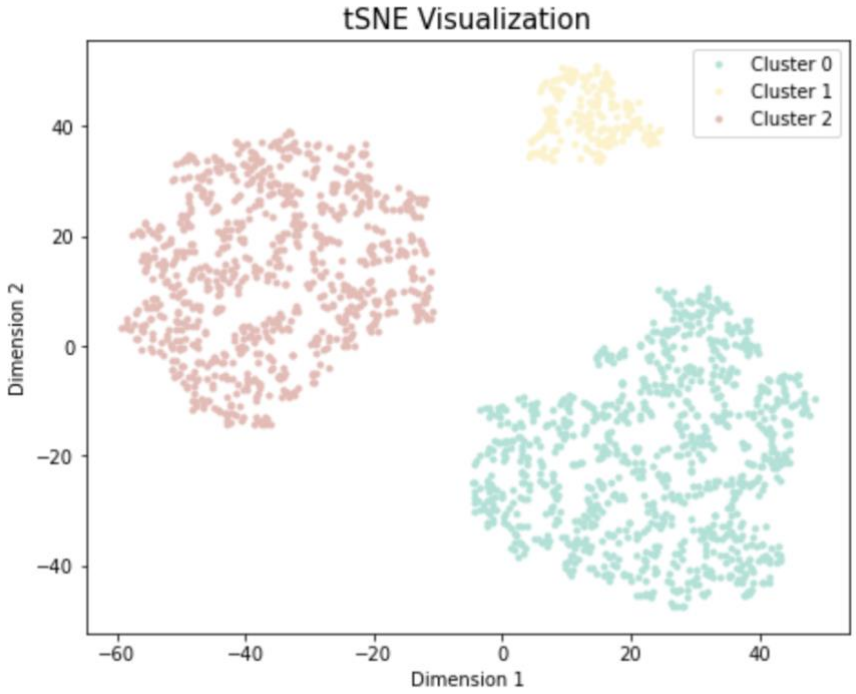
[a] Model 1: crude model.

[b] Model 2: adjusted by gender, age, and ethnic groups.

[c] Model 3: adjusted by gender, age, ethnic groups, serum uric acid, total cholesterol, triglyceride, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, estimated glomerular filtration rate, and smoking and drinking status.

Figures

Figure 1. t-SNE visualization of participant data in cross-sectional study [a].

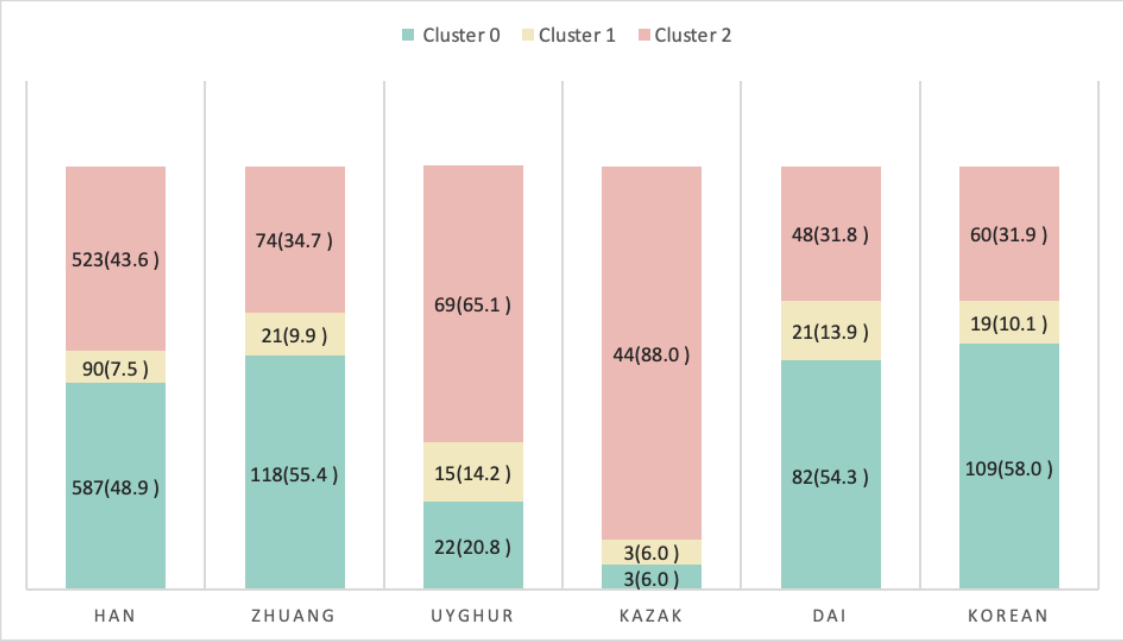


tSNE, t-distributed stochastic neighbor embedding

All data points are labelled with the clustering results from k-means.

[a] It was analysed based on 1,910 participants.

Figure 2. Cluster configuration in the different ethnic groups in cross-sectional study [a].



[a] It was analysed based on 1,908 participants (2 from the 1,910 participants were excluded as they were from other ethnic groups).

Data are presented as numbers (percentages).

Figure 3. Cluster redistribution from baseline to 3-year follow-up in cohort study [a].

[a] It was analysed based on 513 participants included at baseline (2 from the 515 participants were excluded due to the lack of complete information on clustering variables).