Weihao Wang (Orcid ID: 0000-0002-5896-2793)

Application of new international classification of adult-onset diabetes in Chinese inpatients with diabetes mellitus

Weihao Wang[#1] MD, Xiaobei Pei[#1] MBBS, Lina Zhang[#1] MD, Zhi Chen[2], PHD, Dong Lin[1] MD, Xiaoye Duan[1], MD, Jingwen Fan[1], MD, Qi Pan[*1] MD, Lixin Guo[*1] MD

1. Department of Endocrinology, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, P.R. China

2. School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane 4072, Australia.

Short Running Title: Classification of diabetes mellitus

#These authors contributed equally to the manuscript

***Corresponding Authors:**

1. Professor Lixin Guo

   [1]Department of Endocrinology, Beijing Hospital, National Center of Gerontology, Beijing, P.R. China

   E-mail address: glxwork2016@163.com

   Contact no.: +8601085136146

2. Professor Qi Pan,

[1]Department of Endocrinology, Beijing Hospital, National Center of Gerontology, Beijing, P.R. China

E-mail address: panqi621@126.com

ABSTRACT

Background: To determine whether the new international cluster-based classification method can be applied to Chinese inpatients with diabetes mellitus (DM).

Methods: Adult patients with DM hospitalized in our tertiary care center from January 2017 to December 2018 were included in the study. K-means cluster analysis was done in clusters based on glutamic acid decarboxylase antibodies, body mass index, glycosylated hemoglobin, homeostasis model-assessed beta cell function, insulin resistance index and age at diagnosis of DM. Chi-square test was used to analyze inter-subgroup differences in DM-related complications and family history of DM. $P<0.05$ was considered significant.

Results: A total of 1152 inpatients with DM were included in the study. Five subgroups were obtained by cluster analysis with highest proportion of population in mild obesity-related DM subgroup (34.55%), followed by mild age-related DM (21.55%), severe insulin deficiency DM (20.51%), severe insulin resistance DM (19.02%) and severe autoimmune DM subgroup (4.36%). The prevalence of diabetic retinopathy, diabetic peripheral vascular disease, diabetic ketosis, coronary heart disease, hypertension and family history of DM differed significantly among the subgroups ($P<0.05$ for all).

Conclusions: This cluster-based classification could be applied to hospitalized adult patients with DM in China. It might help in strategizing for DM patients, and hence, improve management of DM in these patients.

KEYWORDS

Diabetes, inpatients, classification, cluster analysis

1. BACKGROUND

Diabetes mellitus (DM) is a metabolic disorder characterized by hyperglycemia caused due to deficiency of insulin secretion and/or insulin action.[1] Its prevalence has risen globally due to the aging of the population and the changes in lifestyle.[2] In China, its prevalence has increased remarkedly in the past three decades,[3,4] that is, from 1% in 1980 to 10.9% in 2013,[5,6] which highlights the failure of existing prevention strategies. Classification of DM helps in determining therapeutic strategy and thus, plays major role in its management.[7,8] Conventionally, DM has been diagnosed based on glucose levels in blood. As DM is a heterogenous disease with varied patterns of clinical presentations and progression,[9,10] it is imperative to consider other relevant variables while classifying patients with DM.[9]

DM is conventionally classified into type 1 and type 2: type 1 DM occurs due to destruction of pancreatic beta-cells by autoantibodies, is predisposed to ketoacidosis, and is diagnosed at a younger age compared with type 2 DM,[11] while type 2 DM is caused by insulin deficiency and insulin resistance.[12] Further, slowly evolving, immune-mediated diabetes of adults(previously referred to as latent autoimmune diabetes of adults) is a subtype of type 2 DM characterized by presence of glutamic acid decarboxylase antibodies (GADA). Its manifestations are similar to type 2 DM at diagnosis, but switch to those of type 1 DM over time;[7,13] however, it progresses slower than type 1 DM.[14] Building on the 2020 European Association for the Study of Diabetes (EASD) and American Diabetes Association (ADA) consensus for type 2 DM and heterogeneity within autoimmune diabetes, C-peptide levels and proxy for β-cell

functions should be chosen to drive therapeutic decisions for LADA. [15]

The heterogeneity of DM makes it difficult to individualize treatment for patients with diabetes.[16] Previously, Ahlqvist et al. clustered the newly diagnosed patients with diabetes based on the presence of GADA, age at diagnosis, body mass index (BMI), glycosylated hemoglobin (HbA1c), homeostasis model-assessed beta cell function (HOMA2-β) and insulin resistance (HOMA2-IR) to: 1) severe autoimmune DM (SAID), 2) severe insulin deficiency DM (SIDD), 3) severe insulin resistance DM (SIRD), 4) mild obesity-related DM (MOD), and 5) mild age-related DM (MARD).[17] The study found that the characteristics of patients and risk of diabetic complications significantly varied across the clusters, and the sub-classification provided better prediction of disease progression.[17] A similar cluster analysis of new-onset diabetes patients in China and America divided patients into similar five clusters based on age, BMI, HbA1x, HOMA2-IR and HOMA2-β, and reported findings consistent with those of Ahlqvist et al.[17,18]

However, it is unclear whether the new clustering methods can be applied to classify hospitalized Chinese patients with diabetes. Hence, we implemented the new cluster-based classification to hospitalized Chinese patients with diabetes at our tertiary care center. Additionally, we determined prevalence of DM-associated complications, and family of history of diabetes among different DM subtypes.

2. METHODS

2.1 Study design

A total of 1585 Chinese patients with diabetes aged ≥18 years were hospitalized in our

medical center from January 1, 2017 to December 31, 2018. Among them, patients aged ≥18 years at the time of first diagnosis[19] of diabetes, and for whom, complete clinical data were available were included in the study. Patients were excluded if they aged <18 years at the time of initial diagnosis, had infection, or other complications, or other conditions requiring fasting or intravenous supply of nutrition, were receiving glucocorticoid, were diagnosed with pancreatogenic diabetes, gestational diabetes or other secondary diabetes, were previously diagnosed with or were hospitalized due to malignant tumors, or if they had incomplete relevant clinical data. The study was approved by the ethics committee of Beijing Hospital.

2.2 Measurements

The related indicators included fasting plasma glucose (FPG) and plasma glucose 2 hours after a steamed bread meal, fasting insulin (FINS), 2 hours postprandial insulin (2hINS), fasting C-peptide (FC-P), 2 hours postprandial C-peptide (2hC-P) and glycosylated hemoglobin (HbA1C). For the measurement of blood glucose levels, automatic biochemical analyzer (BECKMAN COULTER AU5400, USA) was used, whereas insulin and C-peptide levels were measured by automatic immunoassay (SIEMENS ADVIA Centaur XP, Germany). HbA1c was measured by Trinity Biotech Primus Hb9210, USA. Serum GADA was determined by enzyme-linked immunosorbent assay at a wavelength of 405nm.

Other biochemical indicators measured were serum creatinine (SCr), uric acid, triglyceride (TG), total cholesterol, low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), all determined by automatic biochemical

analyzer (BECKMAN COULTER AU5400, USA). For urine routine examination, morning urine samples were analyzed with AUTION MAXTMAX-4030, Japan, and urine sediment examination was performed using SYSMEX UF1000i, USA. Results were used to diagnose diabetic ketosis (DK), and to exclude factors such as hematuria and proteinuria that may affect the determination of urinary microalbumin. Urinary microalbumin excretion rate was recorded based on chemiluminescence immunoassay (SIEMENS IMMULITE 2000).

2.3 Diagnostic criteria

DK was diagnosed according to an American Diabetes Association consensus statement.[19] Diabetic retinopathy (DR) was diagnosed based on fundus photographs by an ophthalmologist as described previously.[20] Diabetic nephropathy (DN) and its stages were detected by the presence of microalbuminuria, excluding other causes of proteinuria and microalbuminuria. Diagnosis of microalbuminuria was confirmed with ≥2 measurement of urinary albumin excretion rates of >200 μg/min, or urinary protein excretion of >300 mg per day, or ≥25 mg/mmol urinary albumin/creatinine in males and >35 mg/mmol in females. Diabetic peripheral vascular disease (DPVD) was detected by presence of atherosclerosis in the arteries of lower extremity of patients as suggested by Ahlqvist et al,[21] in whom diagnostic ankle brachial index was performed in combination with a history of claudication.[22] Diabetic distal symmetric polyneuropathy (DSPN) was diagnosed based on the Chinese guidelines for the prevention and treatment of type 2 diabetes (2017).[23] Hypertension was diagnosed as a systolic blood pressure (SBP) ≥140mmHg or diastolic blood pressure (DBP)

≥90mmHg[24]. In addition, levels of LDL-C and HDL-C were also recorded.

2.4 Cluster analysis

Variables used for cluster analysis were GADA, BMI, HbA1c, HOMA2-β, HOMA2-IR and age at diagnosis of DM. HOMA2-β and HOMA2-IR were calculated based on C-peptide concentrations instead of insulin concentration, to exclude the effect of exogenous insulin by the method described by Li et al. as per the formulae given below:[25]

HOMA2-β = 0.27*FC-p [pmol/L] ÷ (FPG [mmol/L]-3.5)

HOMA2-IR = 1.5 + (FPG [mmol/L] * FC-p [pmol/L] ÷ 2800)

The cluster analysis was performed with built-in k-means algorithm in the scikit-learn library of Python. Specifically, we normalized the five numerical variables with Z-score method. The variable GADA was a binary variable and thus we labelled the positive as numerical value 1 and negative as 0. Then, k-means clustering was applied upon the normalized data, yielding the 5 clusters. Each cluster corresponded to a type of diabetes among SAID, SIDD, SIRD, MOD and MARD.

2.5 t-Distributed Stochastic Neighbour Embedding(t-SNE) visualization

As the patient data was distributed in a 6-dimensional space, it was impractical to validate the clustering results from k-means intuitively. We thus leveraged t-SNE visualization technique to reduce the normalized data into a 2-dimensional space, so that we could gain intuition about how does the data look like, and which cluster does each patient converge to. Specifically, t-SNE was an unsupervised nonlinear dimension reduction technique, and mostly used for modelling high-dimensional data into a lower-

dimensional space for better visualization purpose. Even if the data was cast into a lower-dimensional space, t-SNE was capable of preserving the pair-wise distance among patients. The implementation of t-SNE analysis was also conducted with the scikit-learn library of Python, whereas the plot was visualized by the matplotlib library.

2.6 Statistical analysis

All the analyses were performed using SPSS 23.0 software. Chi-square test was used to analyze the differences in prevalence of DR, DN, DPVD, DSPN, DK, coronary heart disease (CHD) and hypertension (HTN). In addition, the risk of diabetes in the previous (parents), current (siblings), and in the next generation (children) among the groups, and the type of DM were assessed using Chi-square test. LDL-C and HDL-C were presented with means and standard deviations. *P* value <0.05 was considered significant.

3. RESULTS

Among the study population, the proportion of patients with type 2 DM, type 1 DM and slowly evolving, immune-mediated diabetes of adults were 95.45%, 2.54% and 2.01%, respectively (Figure 1A). Further, 34.55% of the total study population were clustered as MOD, whereas 21.55%, 20.51%, 19.02% and 4.36% of the patients were clustered into MARD, SIDD, SIRD and SAID, respectively (Figure 1B).

3.1 Characteristics of patients in different subgroups

Cluster analysis revealed that each of the five subtypes had specific characteristics (Figure 2). SAID was characterized by poor metabolic control (Figure 2A), relatively lower BMI (Figure 2B), onset in youth (Figure 2C), insulin deficiency due to impaired

beta cell function (HOMA2-β; Figure 2D) and relatively lower insulin resistance (HOMA2-IR; Figure 2E). Characteristics of SIDD were similar to that of SAID; except for GADA which was negative in SIDD, and insulin deficiency due to non-autoimmune diseases. SIRD group patients were overweight (Figure 2B), had islet cells producing insulin, but were suffering of insulin resistance (Figure 2E). MOD was prevalent in overweight patients who had better metabolic control compared with those of SIRD group (Figure 2B). MARD patients had poor metabolic control and were older at onset compared with patients in other groups (Figure 2A and 2C).

3.2 t-SNE visualization of patient data and the k-means results

As shown in Fig 3, we adopted t-SNE to visualize the data with the k-means clustering results. It was particularly noticeable that the five clusters that formed from the k-means results were similarly distributed in such a 2-dimensional space. Notably, as all patients of SAID were positive with GADA, more discrimination could be observed in Cluster 1 compared with other clusters. For the remaining data representations, the clusters formed by k-means clearly separated them as four groups corresponding to the rest of four types of diabetes. According to the observation above, we further validated our clustering outcomes from the k-means algorithm.

3.3 Risk of complications in different subgroups

Table 1 presents data of complications and family history of DM in among the clusters. The prevalence of majority of complications and family history of DM (except DSPN, smoking, and next generation suffering from diabetes) significantly differed among the groups (*P* values: <0.0001 for DPVD, DK, HTN and family history of

diabetes; 0.0003 for DR and CHD; <0.01 for LDL-C, HDL-C and siblings with DM; 0.0059 for alcohol consumption; Table 1).

SIDD group had the highest prevalence of DR, DK, LDL-C, HDL-C and had parents suffering of DM (0.39%, 0.22%, 3.36 ± 9.08 mg/dL, 1.57 ± 8.33 mg/dL and 0.59, respectively; Table 1). Prevalence of DN and CHD was highest in the SIRD group (0.32% and 0.27%, respectively), whereas alcohol consumption and DM in siblings was highest in the MOD group (0.41% and 0.35%, respectively). The SIRD group had the lowest levels of HDL-C (0.97 ± 0.20 mg/dL). There was no significance difference in the prevalence of diabetes in children of DM patients among different subtypes.

4. DISCUSSION

The new cluster-based classification method has previously been used to classify Nordic and Chinese patients with diabetes.[17,18] To the best of our knowledge, this is the first study that implemented this typing method to hospitalized Chinese patients with DM. We included an additional variable, GADA, which was not used in the previous study in Chinese population.[18] The results showed that this classification method could also be generalized to hospitalized patients with DM in China. Additionally, this study, for the first time, revealed difference in the prevalence of diabetes-related complications and family history of diabetes among the classified clusters in Chinese inpatients with DM. Together with the findings of previous studies, the current study confirms that this cluster-based classification method could be generalized to adult patients with DM from varied ethnicities.[17,18,26]

This cluster-based method is especially important to be implemented for type 2 DM

which is highly heterogeneous.[9] By subgrouping patients based on six DM-related variables, we can determine the risk of different DM-associated complications. Additionally, this would enable physicians to tailor individualized treatment for patients based on their clustering characteristic.[7,8]

Previously, Ahlqvist et al. found that majority of the patients with DM had MARD (39.1%) followed by MOD, SIDD, SIRD and SAID (21.6%, 17.5%, 15.3% and 6.4%, respectively).[17] Similar distribution was also observed by Zou et al.; MARD was the most common subtype (45.1%), followed by MOD, SIDD and SIRD (32.7%, 13.5% and 8.6%, respectively).[18] Our results differ from these previous reports as MOD was the most common subtype in the current study, followed by MARD, SIRD, SIDD and SAID. This might be due to the fact DM patients included in this study were hospitalized, whereas those included in previous studies were newly diagnosed with DM.[17,18] Moreover, Ahlqvist et al. had reported that the metabolic control in patients in the MARD group was better than that of MOD patients, which might have reduced the possibility of hospitalization in MARD patients as observed in the current study.[17] However, there is no epidemiological evidence to support this hypothesis at present. Ahlqvist et al. also reported that 4.5% of the patients had either GADA positive and/or slowly evolving, immune-mediated diabetes of adults.[27] In China, the incidence of GADA positive slowly evolving, immune-mediated diabetes of adults was 5.9%, with higher prevalence in the northern region compared with southern region.[28] At the same time, phenotypical and genetic characteristics of LADA were associated with gender and age. [29,30] However, there was only GADA that was evaluated in Ahlqvist's study

and no other autoimmune antibodies were present in LADA. In LADA subjects, high GADA titer was related to a severe autoimmunity profile and routine screening for other antibodies was recommended. [29]

Further, in the NHANES III study, level of HbA1c was highest in the SIDD group, whereas in the current study, it was highest in the MARD group.[18] This could be due to the fact that 98.2% of the MARD patients in this study had type 2 DM and the inpatients in our study center were also mostly elderly patients with DM. For elderly DM patients with poor health or short life expectancy, the control target of HbA1c can be relaxed to 8.0% or even 8.5% [31–33] while the target HbA1c should be <7.5% for healthy elderly DM patients who have relatively longer life expectancy.[32]

In the Nordic study, each subtype was associated with different complications (mainly DR, DN, DK).[17] DR was most common in the SIDD subgroup (Odds ratio [OR]: 1.6; 95% CI: 1.3, 1.9), which is consistent with our findings.[17] Risk of DN was also highest in the SIRD subgroup (Risk ratio: 4.89; 95% 2.68, 8.93; P<0.0001) which was in accordance to our current study. However, the inter-subgroup difference observed in the current study was insignificant (P = 0.0508), which might be due to the small sample size. Our results showed highest prevalence of DK in the SIDD subgroup which in accordance to the findings of Ahlqvist et al. Interestingly, the Nordic study suggested HbA1c as a predictor of risk for DK (OR: 2.73; 95% CI: 2.47, 3.03; P<0.0001).[17] In contrast, in the present study, MARD subgroup had highest HbA1c, but highest prevalence of DK was found in the SIDD. This might be due to the clinical differences between newly diagnosed DM patients and inpatients with DM. Moreover,

SAID, SIDD and MOD subtypes were associated with a lower risk of coronary heart disease in the previous study.[17] On the contrary, our study found highest prevalence of CHD in SIRD subgroup.

Furthermore, the Nordic study did not evaluate blood lipids in the included patients.[17] We found that SIDD and SIRD subgroups had the highest risk of elevated LDL and reduced HDL, respectively. Additionally, highest prevalence of HTN was observed in SAID subgroup in the current study, which was not assessed in the previous study.[17]

The German Diabetes Study reassessed the cluster results after a 5-year follow-up. The prevalence of SAID was slightly higher in this study, whereas the distribution of the other clusters was similar with the Swedish study. They found that patients in cluster membership could change as the disease progression. It was worth noting that the repeatability of the clustering algorithm applied to patients with 5-year duration was only 77%. At the same time, testing of more than one islet-related antibody could result in more patients being classified with autoimmune diabetes. This study also interpreted the relationship between clusteration and diabetes-related complications. Patients in SIRD cluster had decreased eGFR and increased cystatin-C levels, suggesting a significant correlation between insulin resistance and early progression of diabetic nephropathy[34] .

Also another study confirmed that the data-driven clusters of Ahlqvist's group were reproducible in randomized controlled trial data. They had a new discovery that there would be better clinical utility from modelling clinical features, rather than from

clustering patients with different clinical features into different subgroups. According to the analysis, SAID, SIDD and MOD had an increased risk of glycaemic progression. SIRD and MARD had a faster progression of renal disease. Another difference between this study and Ahlqvist's study is the cluster differ in response to diverse hypoglycemic treatment. There was a particular benefit for SIRD with using thiazolidinediones, and for MARD with using sulfonylureas[35]. These results raise the possibility that hypoglycemic treatment of the diverse drugs might be identifiable through combining clinical features in a model for drug selection.

We included GADA, age at diagnosis, BMI, HbA1c, HOMA2-β and HOMA2-IR as variables for cluster analysis. In the study by Safai et al., the duration of disease was also included in addition to the aforementioned variables.[26] After cluster analysis, the researchers categorized DM patients into five subtypes: autoimmune islet failure DM, short-term insulin resistance DM, non-autoimmune islet failure DM, long-term insulin resistance DM and DM with metabolic syndrome. This shows that the choice of variables directly influences the type of resultant subgroups. Therefore, future studies using different variables might further improve the method of clustering DM patients.

With the continuous development of artificial intelligence, the clustering analysis method can be further optimized by continuous integration and exploration of the two disciplines.[36] Application of the new cluster-based classification method in clinical practice will help physicians to individualize DM treatment.[7] However, large-scale, prospective, multi-center and multi-parameter studies are warranted to support this assertion.

This study had few limitations. Firstly, the sample size of this study is small. A future study with larger sample size would re-affirm the findings of this study. Secondly, as only hospitalized patients were included in the study, data on the six assessed at initial diagnosis and subsequent follow-up was not available. Thirdly, as we did not have data from the initial diagnosis, we could not evaluate the development of each subtype. Finally, we did not subclassify patients based on treatment regimen, which might have revealed the pattern of treatment prescription for different subtypes.

CONCLUSIONS

The new cluster-based classification of diabetes can be implemented in Chinese inpatients with DM. GADA, which was not included in previous study on Chinese patients with DM, is an important variable that should be considered while classifying Chinese patients with DM. Future studies are warranted to explore differences among different races, nationalities or even between genders that should be taken into account while classifying and treating DM patients.

List of abbreviations

DM: Diabetes mellitus; LADA: latent autoimmune diabetes in adults; GADA: glutamic acid decarboxylase antibodies; BMI: body mass index; HbA1c: glycosylated hemoglobin; HOMA2-β : homeostasis model-assessed beta cell function; HOMA2-IR: homeostasis model-assessed insulin resistance; SAID: severe autoimmune diabetes; SIDD: severe insulin-deficient diabetes; SIRD: severe insulin-resistant diabetes; MOD: mild obesity-related diabetes; MARD: mild age-related diabetes; FPG: fasting plasma glucose; FINS: fasting insulin; 2hINS: 2 hours postprandial insulin; FC-P: fasting C-

peptide; 2hC-P: 2 hours postprandial C-peptide; SCr: serum creatinine; TG: triglyceride; LDL-C: low density lipoprotein cholesterol; HDL-C: high density lipoprotein cholesterol; DK: diabetic ketosis; DR: Diabetic retinopathy; DN: Diabetic nephropathy; DPVD: Diabetic peripheral vascular disease; DSPN: Diabetic distal symmetric polyneuropathy; CHD: coronary heart disease; HTN: hypertension; OR: Odds ratio.

Declarations

Ethics approval

This study was approved by the Ethical Committee of Beijing Hospital.

Consent for publication

Not applicable

Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Funding

This study was funded by National Natural Science Foundation of China (Grant No. 81670763 and 81471050).

Acknowledgements

The authors declare that they have no competing interests.

Author contributions

Professor Lixin Guo and Qi Pan made substantial contributions to conception and design and revised it critically for important intellectual content. Weihao Wang,

Xiaobei Pei and Lina Zhang made substantial contributions to acquisition of data, analysis and interpretation of data. Dong Lin, Xiaoye Duan and Jingwen Fan have been involved in drafting the manuscript.

REFERENCES

1. Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus.

2. Cho NH, Shaw JE, Karuranga S, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract*. 2018;138:271-281.

3. Hu C, Jia W. Diabetes in China: Epidemiology and Genetic Risk Factors and Their Clinical Utility in Personalized Medication. *Diabetes*. 2018;67(1):3-11.

4. Genchev D. Rate of diabetes in China "explosive." Published online November 20, 2018.

5. Wang L, Gao P, Zhang M, et al. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. *JAMA*. 2017;317(24):2515-2523.

6. [A mass survey of diabetes mellitus in a population of 300,000 in 14 provinces and municipalities in China (author's transl)]. *Zhonghua Nei Ke Za Zhi*. 1981;20(11):678-683.

7. Prasad RB, Groop L. Precision medicine in type 2 diabetes. *J Intern Med*. 2019;285(1):40-48.

8. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2019*. *Diabetes Care*. 2019;42(Supplement 1):S13-S28.

9. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: a disease with increasing heterogeneity. *Lancet Lond Engl*. 2014;383(9922):1084-1094.

10. Fitipaldi H, McCarthy MI, Florez JC, Franks PW. A Global Overview of Precision Medicine in Type 2 Diabetes. *Diabetes*. 2018;67(10):1911-1922.

11. World Health Organization. *Classification of Diabetes Mellitus*.

12. Punthakee Z, Goldenberg R, Katz P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Can J Diabetes*. 2018;42:S10-S15.

13. Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR. Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. *Diabetes*. 1993;42(2):359-362.

14. Tuomi T, Carlsson A, Li H, et al. Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes*. 1999;48(1):150-157.

15. Buzzetti R, Tuomi T, Mauricio D, et al. Management of Latent Autoimmune Diabetes in Adults: A Consensus Statement From an International Expert Panel. *Diabetes*. 2020;69(10):2037-2047.

16. Bowman P, Flanagan SE, Hattersley AT. Future Roadmaps for Precision Medicine Applied to Diabetes: Rising to the Challenge of Heterogeneity. Journal of Diabetes Research.

17. 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(5):361-369.

18. Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol*. 2019;7(1):9-11.

19. World Health Organization. (1999). Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. Published online 1999.

20. Martinell M, Dorkhan M, Stålhammar J, Storm P, Groop L, Gustavsson C. Prevalence and risk factors for diabetic retinopathy at diagnosis (DRAD) in patients recently diagnosed with type 2 diabetes (T2D) or latent autoimmune diabetes in the adult (LADA). *J Diabetes Complications*. 2016;30(8):1456-1461.

21. Association AD. Peripheral Arterial Disease in People With Diabetes. *Diabetes Care*. 2003;26(12):3333-3341.

22. Association AD. 11. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes—2019. *Diabetes Care*. 2019;42(Supplement 1):S124-S138.

23. WP J, Lu J. 2017 Chinese guidelines for the prevention and treatment of type 2 diabetes mellitus. *Chin J Diabetes*. 2018;10(1):2-3.

24. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertens Dallas Tex 1979*. 2018;71(6):1269-1324.

25. Li X, Zhou Z-G, Qi H-Y, Chen X-Y, Huang G. [Replacement of insulin by fasting C-peptide in modified homeostasis model assessment to evaluate insulin resistance and islet beta cell function]. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2004;29(4):419-423.

26. Safai N, Ali A, Rossing P, Ridderstråle M. Stratification of type 2 diabetes based on routine clinical markers. *Diabetes Res Clin Pract*. 2018;141:275-283.

27. Buzzetti R, Pietro SD, Giaccari A, et al. High Titer of Autoantibodies to GAD Identifies a Specific Phenotype of Adult-Onset Autoimmune Diabetes. *Diabetes Care*. 2007;30(4):932-938.

28. Zhou Z, Xiang Y, Ji L, et al. Frequency, immunogenetics, and clinical characteristics of latent

autoimmune diabetes in China (LADA China study): a nationwide, multicenter, clinic-based cross-sectional study. *Diabetes*. 2013;62(2):543-550.

29. Zampetti S, Capizzi M, Spoletini M, et al. GADA titer-related risk for organ-specific autoimmunity in LADA subjects subdivided according to gender (NIRAD study 6). *J Clin Endocrinol Metab*. 2012;97(10):3759-3765.

30. Niu X, Luo S, Li X, et al. Identification of a distinct phenotype of elderly latent autoimmune diabetes in adults: LADA China Study 8. *Diabetes Metab Res Rev*. 2019;35(1):e3068.

31. Snow R, Sandall J, Humphrey C. Use of clinical targets in diabetes patient education: qualitative analysis of the expectations and impact of a structured self-management programme in Type 1 diabetes. *Diabet Med J Br Diabet Assoc*. 2014;31(6):733-738.

32. Association AD. 12. Older Adults: Standards of Medical Care in Diabetes—2019. *Diabetes Care*. 2019;42(Supplement 1):S139-S147.

33. Qaseem A, Wilt TJ, Kansagara D, et al. Hemoglobin A $_{1c}$ Targets for Glycemic Control With Pharmacologic Therapy for Nonpregnant Adults With Type 2 Diabetes Mellitus: A Guidance Statement Update From the American College of Physicians. *Ann Intern Med*. 2018;168(8):569.

34. 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(9):684-694.

35. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. 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(6):442-451.

36. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. *J Med Internet Res*. 2018;20(5).

TABLE 1 Prevalence complications and family history of DM in different cluster

Variables	Cluster1 (SAID; n=50), %	Cluster2 (SIDD; n=235), %	Cluster3 (SIRD; n=218), %	Cluster4 (MOD; n=395), %	Cluster5 (MARD; n=245), %	Total (n = 1143)	*P* value
DR	0.35	0.39	0.29	0.28	0.20	0.29	0.0003
DN	0.16	0.27	0.32	0.23	0.26	0.26	0.0508
DPVD	0.73	0.67	0.80	0.83	0.89	0.61	＜0.0001
DSPN	0.39	0.48	0.50	0.49	0.56	0.50	0.1539
DK	0.18	0.22	0.11	0.05	0.16	0.12	＜0.0001
CHD	0.18	0.11	0.27	0.24	0.23	0.21	0.0003
HTN	0.63	0.60	0.26	0.44	0.35	0.35	＜0.0001
LDL-C (mg/dL), mean±SD	2.49±0.82	3.36±9.08	2.44±0.87	2.82±8.34	2.61±0.94	2.80±6.43	＜0.01
HDL-C (mg/dL), mean±SD	1.18±0.31	1.57±8.33	0.97±0.20	1.08±0.30	1.06±0.28	1.16±3.79	＜0.01
Smoking	0.41	0.45	0.48	0.46	0.37	0.36	0.1082
Alcohol consumption	0.33	0.39	0.40	0.41	0.28	0.31	0.0195
FH1	0.37	0.59	0.52	0.44	0.31	0.45	＜0.0001
FH2	0.14	0.27	0.31	0.35	0.25	0.29	0.0059
FH3	0.02	0.03	0.02	0.03	0.05	0.03	0.2155

DR, Diabetic Retinopathy; DN, Diabetic Nephropathy; DPVD, Diabetic Peripheral Vascular Disease; DPSN, Diabetic Distal Symmetric Polyneuropathy; DK, Diabetic Ketosis; CHD, Coronary Heart Disease; HTN, Hypertension; LDL-C, Low Density Lipoprotein Cholesterol; HDL-C, High Density Lipoprotein Cholesterol; FH1(family history), previous generation (parents) suffered from diabetes; FH2, siblings suffered from diabetes; FH3, next generation (children) suffers from diabetes.

FIGURE LEGENDS:

Figure 1. Proportion of each type in the classical typing method and clustering analysis typing method. A) Patient distribution using WHO typing method; B) Patient distribution using cluster analysis method. SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes; MOD, mild obesity-related diabetes; MARD, mild age-related diabetes.

Figure 2. Classification characteristics of diabetes mellitus. Distribution of HbA1c, BMI, age at diagnosis, HOMA2-β and HOMA2-IR in patients admitted to the study group after stratification. SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes; MOD, mild obesity-related diabetes; MARD, mild age-related diabetes.

Figure 3. t-SNE visualization of the patient data in a 2-dimensional space. All data points are labelled with the clustering results from k-means.