

Uncovering heterogeneous cardiometabolic risk profiles in US adults: the role of social and behavioral determinants of health

Qinglan Ding ¹, Yuan Lu,² Jeph Herrin,³ Tianyi Zhang,⁴ David G Marrero^{5,6}

To cite: Ding Q, Lu Y, Herrin J, *et al.* Uncovering heterogeneous cardiometabolic risk profiles in US adults: the role of social and behavioral determinants of health. *BMJ Open Diab Res Care* 2023;**11**:e003558. doi:10.1136/bmjdr-2023-003558

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2023-003558>).

Received 30 May 2023
Accepted 15 August 2023



© Author(s) (or their employer(s)) 2023. 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 Qinglan Ding;
qinglanding@purdue.edu

ABSTRACT

Introduction Social and behavioral determinants of health (SBDH) have been linked to diabetes risk, but their role in explaining variations in cardiometabolic risk across race/ethnicity in US adults is unclear. This study aimed to classify adults into distinct cardiometabolic risk subgroups using SBDH and clinically measured metabolic risk factors, while comparing their associations with undiagnosed diabetes and pre-diabetes by race/ethnicity.

Research design and methods We analyzed data from 38,476 US adults without prior diabetes diagnosis from the National Health and Nutrition Examination Survey (NHANES) 1999–2018. The k-prototypes clustering algorithm was used to identify subgroups based on 16 SBDH and 13 metabolic risk factors. Each participant was classified as having no diabetes, pre-diabetes or undiagnosed diabetes using contemporaneous laboratory data. Logistic regression was used to assess associations between subgroups and diabetes status, focusing on differences by race/ethnicity.

Results Three subgroups were identified: cluster 1, primarily middle-aged adults with high rates of smoking, alcohol use, short sleep duration, and low diet quality; cluster 2, mostly young non-white adults with low income, low health insurance coverage, and limited healthcare access; and cluster 3, mostly older males who were the least physically active, but with high insurance coverage and healthcare access. Compared with cluster 2, adjusted ORs (95% CI) for undiagnosed diabetes were 14.9 (10.9, 20.2) in cluster 3 and 3.7 (2.8, 4.8) in cluster 1. Clusters 1 and 3 (vs cluster 2) had high odds of pre-diabetes, with ORs of 1.8 (1.6, 1.9) and 2.1 (1.8, 2.4), respectively. Race/ethnicity was found to modify the relationship between identified subgroups and pre-diabetes risk.

Conclusions Self-reported SBDH combined with metabolic factors can be used to classify adults into subgroups with distinct cardiometabolic risk profiles. This approach may help identify individuals who would benefit from screening for diabetes and pre-diabetes and potentially suggest effective prevention strategies.

INTRODUCTION

Diabetes constitutes a significant public health challenge in the USA, affecting approximately 1 in 10 adults, with 2% of cases going undiagnosed.¹ Pre-diabetes, a

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Social and behavioral determinants of health (SBDH) have been linked to diabetes risk.
- ⇒ Conventional models of diabetes assessment often overlook the influence of SBDH on diabetes risk.

WHAT THIS STUDY ADDS

- ⇒ This study demonstrates that integrating a wide range of SBDH and objectively measured metabolic risk factors through clustering analysis can identify distinct cardiometabolic risk subgroups in US adults without a prior diagnosis of diabetes.
- ⇒ Three distinct subgroups were identified, with 'Middle-aged adults with multiple metabolic risk factors' and 'Elderly adults with chronic conditions and low physical activity levels' showing significantly elevated risk for both pre-diabetes and undiagnosed diabetes compared with 'Healthy socioeconomically vulnerable young adults'.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ The findings highlight the importance of considering SBDH in addition to metabolic risk factors to accurately identify individuals at risk for pre-diabetes and undiagnosed diabetes.
- ⇒ A comprehensive approach incorporating SBDH has the potential to enhance risk stratification and improve the targeting of high-risk individuals.
- ⇒ This implication could impact research endeavors, guide informed clinical practice, and contribute to the development of policies aimed at preventing and effectively managing diabetes.

precursor to type 2 diabetes, affects a staggering 38% of the adult population in the USA, representing a leading risk factor for the development of diabetes.^{1,2} Poorly controlled diabetes is associated with elevated risk of microvascular and macrovascular complications and is also linked to a 75% increase in adult mortality rate.³ Given the substantial economic and social impact of diabetes, estimated to have cost

\$327 billion in 2017, addressing this disease is of utmost importance.⁴

Screening asymptomatic individuals for diabetes and pre-diabetes has been proposed as a strategy for earlier detection and treatment. However, the heterogeneity in cardiometabolic risk profiles among individuals with prediabetes⁵ and the lack of evidence supporting the effectiveness of such screening present challenges.⁶ Developing personalized screening strategies that effectively allocate resources to those at the highest risk is paramount.⁶ While heredity is known to play a significant role in diabetes susceptibility,⁷ recent studies have also highlighted the importance of social and behavioral determinants of health (SBDH) in contributing to variations of diabetes risk and clinical outcomes.^{8,9} Integrating SBDH into clinical profiling could yield a more comprehensive understanding of the critical factors influencing diabetes risk,^{10,11} thereby enhancing clinicians' abilities to identify, engage, and manage individuals at high risk for diabetes and its complications.¹²

Despite the potential benefits of incorporating SBDH into diabetes risk assessment, little is known about whether consideration of these factors can improve risk stratification in the general population and for racial/ethnic minorities. This knowledge gap may be attributed, in part, to the limited routine assessment of social and behavioral factors in clinical practice.¹³ Furthermore, methodological challenges arise when attempting to classify subjects into homogeneous groups due to the mixture of numeric and categorical attributes.^{14,15} Existing visual and statistical analytic tools often focus on a single data type or handle mixed data types by converting categorical and nominal attributes into numeric values, potentially overlooking the nuances of SBDH.¹⁵

To address this gap, we analyzed a nationally representative sample of adults without a history of diabetes. Employing the K-prototype cluster algorithm, we categorized participants based on their SBDH and metabolic risk factors. Subsequently, we compared the rates of undiagnosed diabetes and pre-diabetes across the distinct clusters, determining diabetes status using laboratory data. Furthermore, we examined whether the association between these clusters and undiagnosed diabetes or pre-diabetes varied by race/ethnicity. Our ultimate objective is to inform the development of a targeted screening approach for diabetes or pre-diabetes that incorporates SBDH to enhance the prediction of diabetes risk and facilitate interventions for prevention.

RESEARCH DESIGN AND METHODS

Data sources and study population

We used pooled data from the 1999–2018 cycles of the National Health and Nutrition Examination Survey (NHANES), a nationally representative, ongoing, cross-sectional program of surveys that assess the health and nutritional status of the non-institutionalized civilian population in the USA.¹⁶ A physical examination and

laboratory data are collected on a subset of NHANES respondents. Details of NHANES survey instruments and protocols have been described extensively.¹⁷

We included adults aged 18 and above with laboratory data and without known diabetes (defined as self-reported history of diabetes, self-reported use of diabetes medications, or self-reported dietary treatment due to diabetes at the time of being surveyed), and excluded participants missing information on diabetes status, hemoglobin A1c, or fasting blood glucose values. We selected variables representing social determinants of health, behavioral determinants of health, and metabolic risk factors based on previously documented associations with diabetes or cardiovascular health.^{8,18}

Social determinants of health variables

We used the policy report from the US Department of Health and Human Services to guide our selection of social determinants of health variables.¹⁹ We included measures of race/ethnicity, sex, family income, education, health insurance status, and access to healthcare. We categorized race/ethnicity as non-Hispanic whites, non-Hispanic blacks, Hispanic, and non-Hispanic others. Family income was categorized based on the established poverty-income ratio (PIR), calculated as family income divided by the federal poverty level (online supplemental method 1). Participants were considered to have health insurance if they reported coverage by a private, Medicare/Medicaid, or other government or state-sponsored insurance plan. Health insurance was further classified into no insurance, private insurance, government insurance alone, or both. We defined access to healthcare as reporting either a place to visit for routine health or at least one healthcare visit in the past year. We also included other potentially important demographic variables, such as age and marital status.

Behavioral determinants of health variables

We included measures of dietary intake, physical activity, sleep duration, smoking status, and alcohol use. We evaluated diet quality as a dietary pattern score ranging from 0 to 5, derived from the American Heart Association's national goals for cardiovascular health promotion and disease prevention.^{20,21} Participants' diets were evaluated on frequency and portion intake of fruit and vegetables, fish, whole grains, sugar-sweetened beverages, and daily sodium (online supplemental method 2).²¹

Physical activity levels were assessed using the Centers for Disease Control and Prevention's Physical Activity Guidelines for Americans,²² classifying participants into inactive and active (online supplemental method 3). Sleep duration was assessed using self-report hours of sleep the participants usually got. Smoking status was assessed using serum cotinine levels, and we defined active smokers as those with a cotinine level ≥ 10 ng/mL.²³ Current alcohol user was characterized by having at least 12 alcoholic drinks in the past year and having consumed alcohol on at least 1 day in the past year.²⁴

Metabolic risk factors

Metabolic risk factors were assessed through physical examination and laboratory testing in mobile examination centers. Metabolic risk factors included in the study were body mass index (BMI), waist circumference, blood pressure (BP), fasting glucose, hemoglobin A1c (HbA1c), cholesterol, blood urea nitrogen (BUN), and uric acid levels. Cholesterol included total cholesterol levels, triglycerides, and high-density and low-density lipoprotein (HDL and LDL). Participants' blood samples were collected after 8 hours of overnight fasting. We also included participants' history of hypertension and their family history of diabetes as additional metabolic risk factors. Participants were considered to have a history of hypertension if they had an average systolic BP ≥ 140 mm Hg or average diastolic BP ≥ 90 mm Hg of the three BP measurements; or had reported current use of antihypertensive medication.²⁵

Undiagnosed diabetes and pre-diabetes definitions

In our cohort of adults without prior diagnosis of diabetes, undiagnosed diabetes was defined by having an HbA1c level $\geq 6.5\%$ or a fasting blood glucose level ≥ 126 mg/dL.²⁶ Adults who did not have a prior diabetes diagnosis or undiagnosed diabetes but had a HbA1c level between 5.7% and 6.4% or a fasting blood glucose level between 100 and 125 mg/dL were defined as having pre-diabetes.²⁶

Variable selection

To address multicollinearity,¹⁵ we conducted principal components analysis²⁷ to reduce the number of variables before the cluster analysis (online supplemental methods 4). Numeric variables were standardized with a mean of 0 and an SD of 1. Within each domain, we selected numeric variables that yielded the smallest $1 - R^2$ value. Categorical variables with redundancy and those with $>10\%$ missing data were excluded. As a result, 21 independent variables were chosen from the initial 29 variables, explaining 81%–100% of the total data variance (online supplemental tables 1–5).

Statistical analyses

We applied the K-prototype clustering algorithm to the analytic dataset, which comprised 29 variables from three domains: social determinants of health, behavioral determinants of health, and metabolic risk factors. The K-prototype algorithm, a combination of k-means and k-modes algorithms, was selected for its suitability in clustering mixed data types and large datasets (online supplemental method 5).^{14 15} A seven-step approach was employed to assign individuals with similarity in variable profiles to specific clusters, identifying distinct population clusters (online supplemental method 6).

To address missing data, we employed multivariate imputation with fully conditional specification to create 10 imputed datasets. The K-prototype algorithm was then applied to generate clusters on each of the

10 imputed datasets. The optimal number of clusters was determined using the Silhouette score and Elbow method (online supplemental tables 6 and 7).²⁸ Once the optimal number of clusters was established, the K-prototype algorithm was re-run on all 10 datasets to identify clusters based on volume. Hierarchical aggregation was applied to combine cluster assignments across the imputed datasets, ensuring robust results by assigning individuals to their most commonly assigned cluster.²⁹

Following cluster identification, respondent characteristic and laboratory features were summarized by cluster, and differences were assessed using χ^2 tests for categorical variables and one-way analysis of variance for continuous variables. Stacked bar charts were used to visualize differences in SBDH frequencies across clusters (figure 1). Furthermore, bivariate and multinomial logistic regression models were fitted to examine the relationship between risk factors and cluster membership, with cluster assignment as the dependent variable and all classifier variables as independent variables (online supplemental table 8).

Logistic regression models were employed to evaluate the association between cluster and undiagnosed diabetes and pre-diabetes. Race-specific adjusted ORs and 95% CI for undiagnosed diabetes and pre-diabetes risk were calculated by repeating the analysis in different race/ethnicity groups.

All statistical analyses were conducted using Stata V.17³⁰ (StataCorp, Texas, USA) and R³¹ (R Foundation for Statistical Computing, Austria). The institutional review board of Purdue University West Lafayette campus determined that this study does not require review as human subjects research as defined under the US Department of Health and Human Services regulations 45 CFR 46.

RESULTS

There were 42,606 participants aged 18 years or older in NHANES who were non-pregnant and had information on self-report SBDH, physical examination, and clinical laboratory data. After excluding adults with known diabetes, 38,476 adults with no prior diagnosis of diabetes were included in these analyses. K-prototype analyses of the 10 imputed datasets identified three clusters as providing the best clustering solution across all 10 datasets. The identified clusters differed significantly in SBDH and metabolic risk factors ($p < 0.01$ for all characteristics, table 1). When examining the bivariate associations between cluster membership and individual risk factors, 27 of the 29 variables representing SBDH and metabolic risks showed a significant association with cluster membership ($p < 0.05$, online supplemental table 8).

The three clusters were named according to their prominent features and described below (figures 1 and 2).

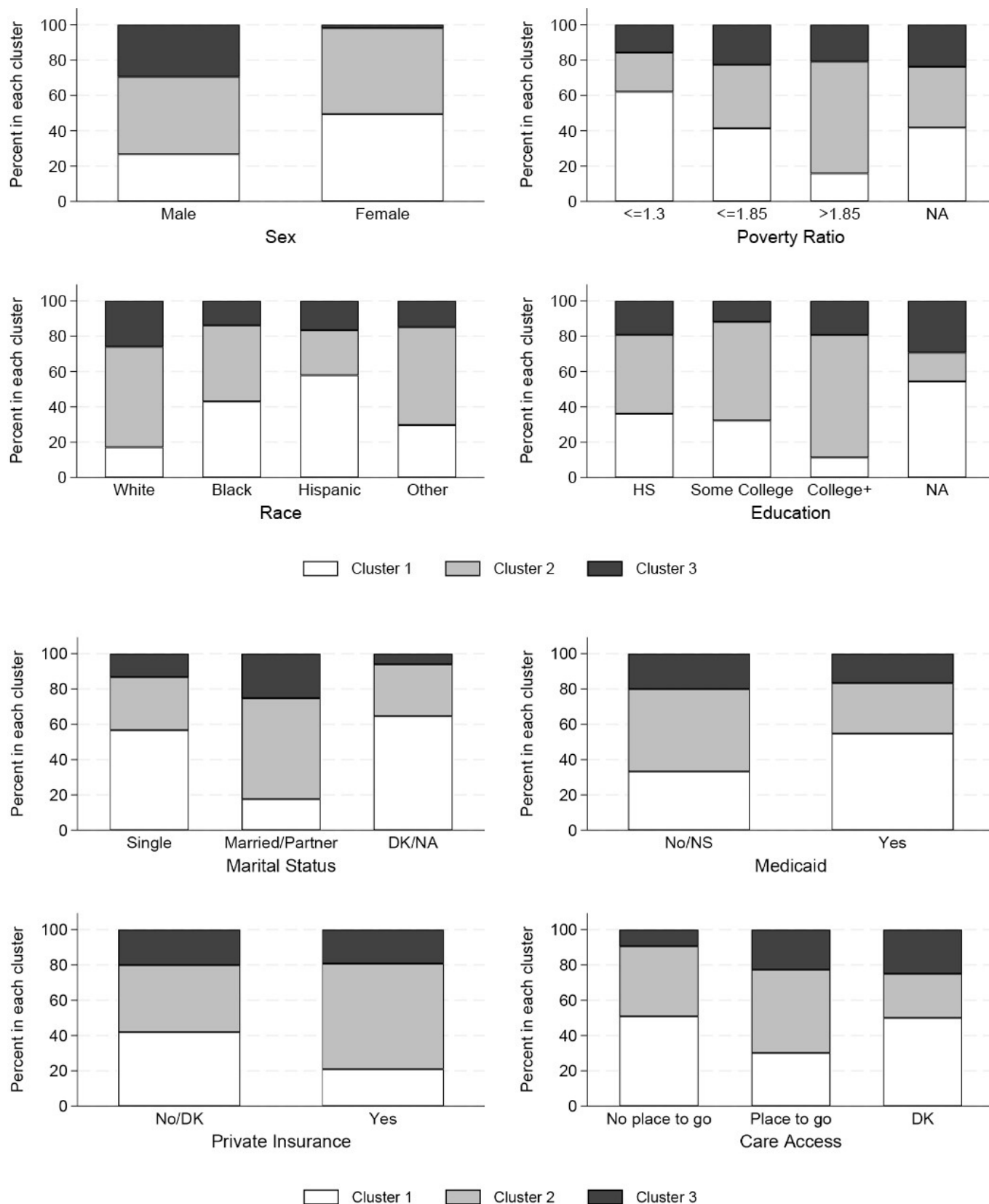


Figure 1 Differences in social determinants of health characteristics across the three clusters. BMI, body mass index; DK, response was unknown; HS, high school or less; NA, value not available.

Cluster 1: middle-aged adults with multiple metabolic risk factors

Cluster 1 (n=18,823 (48.9% of all adults), mean

age=42.4SD (14.0), 22% female) was the largest cluster, comprising middle-aged adults with multiple metabolic risk factors. This group had the highest poverty-income

Table 1 Social determinants of health, behavioral determinants of health and cardiometabolic risk characteristics by clusters

Variables	Cluster 1	Cluster 2	Cluster 3	P value
	n (%) or mean (SD)	n (%) or mean (SD)	n (%) or mean (SD)	
	18,823 (48.9%)	13,401 (34.8%)	6252 (16.3%)	
Social determinants of health				
Female	4150 (22.0)	9080 (67.8)	180 (2.9)	<0.001
Age, mean, years	42.4 (14.0)	29.4 (10.1)	65.5 (15.1)	<0.001
Poverty ratio				
<1.30	3611 (19.2)	5886 (43.9)	1653 (26.4)	<0.001
>1.85	11,908 (63.3)	4341 (32.4)	3024 (48.4)	
Race/ethnicity				
Non-Hispanic white	9809 (52.1)	3349 (25.0)	3261 (52.2)	<0.001
Non-Hispanic black	3715 (19.7)	3525 (26.3)	954 (15.3)	
Hispanic	3437 (18.3)	5199 (38.8)	1595 (25.5)	
Other	1862 (9.9)	1328 (9.9)	442 (7.1)	
Education				
High school or less	4621 (24.5)	3230 (24.1)	1314 (21.0)	<0.001
Married/with partner	13,383 (71.1)	4432 (33.1)	4148 (66.3)	<0.001
Health insurance				
Yes	14,463 (76.8)	8495 (63.4)	5399 (86.4)	<0.001
Healthcare access				
No place to go	3960 (21.0)	3660 (27.3)	767 (12.3)	<0.001
Behavioral determinants of health				
Smoking status				
Significant	2096 (11.1)	1973 (14.7)	481 (7.7)	<0.001
Alcohol use				
Yes	12,352 (65.6)	5685 (42.4)	3587 (57.4)	<0.001
Sleep duration (hours)	6.8 (1.4)	7.2 (1.5)	7.4 (1.5)	<0.001
Diet score	1.0 (0.7)	1.3 (0.7)	1.1 (0.7)	<0.001
Physical activity				
Inactive	10,819 (57.5)	7759 (57.9)	4194 (67.1)	<0.001
Metabolic risk factors				
BMI (kg/m ²)	29.4 (6.8)	26.2 (6.2)	27.9 (4.9)	<0.001
Waist circumference (cm)	100.4 (15.8)	88.2 (14.5)	102.0 (12.7)	<0.001
Total cholesterol (mg/dL)	202.1 (40.6)	172.8 (32.9)	191.3 (41.3)	<0.001
Triglycerides (mg/dL)	139.6 (101.2)	92.9 (57.7)	154.5 (197.5)	<0.001
HDL (mg/dL)	49.5 (15.0)	54.4 (14.7)	48.6 (14.5)	<0.001
LDL (mg/dL)	123.9 (34.8)	98.4 (28.5)	113.8 (34.9)	<0.001
Hemoglobin A1c (%)	5.4 (0.5)	5.2 (0.4)	5.8 (1.1)	<0.001
Mean SBP (mmHg)	123.6 (16.1)	112.7 (12.3)	131.1 (20.1)	<0.001
Mean DBP (mm Hg)	75.0 (11.4)	65.6 (11.1)	68.4 (13.9)	<0.001
Uric acid (mg/dL)	5.9 (1.3)	4.7 (1.1)	6.3 (1.4)	<0.001
BUN (mg/dL)	12.6 (3.6)	10.7 (3.3)	18.3 (7.3)	<0.001
History of HTN				
Yes	4681 (24.9)	1179 (8.8)	2901 (46.4)	<0.001
Family history of DM	7434 (39.5)	4370 (32.6)	1948 (31.2)	<0.001

BMI, body mass index; BUN, blood urea nitrogen; DBP, diastolic blood pressure; DM, diabetes; HDL, high-density lipoprotein; HTN, hypertension; LDL, low-density lipoprotein; n, number of participants; SBP, systolic blood pressure.

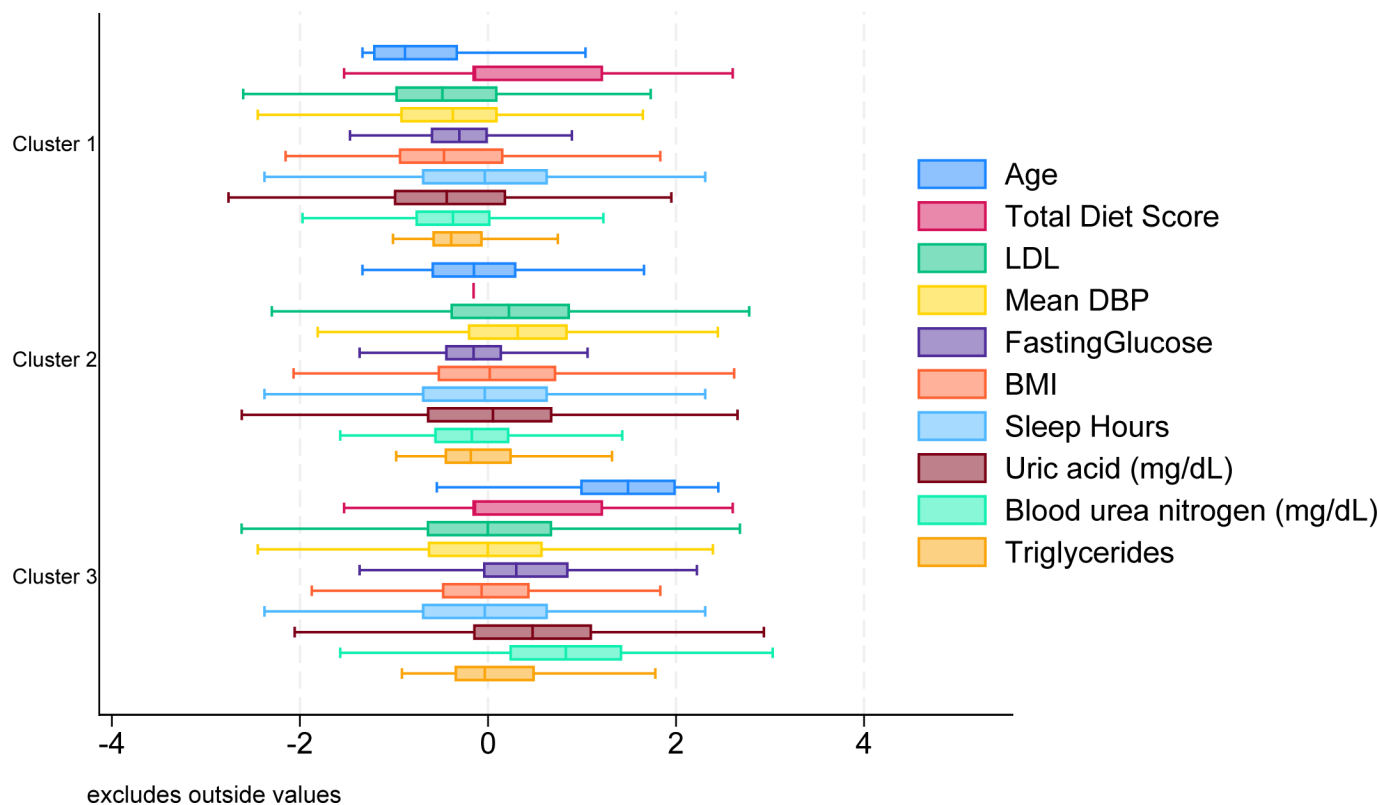


Figure 2 Differences in behavioral determinants of health characteristics and metabolic risk factors across the three clusters. The above continuous variables were standardized to have a mean of 0 and SD of 1 for plotting and easy comparison in one graph. BMI, body mass index; LDL, low-density lipoprotein cholesterol; mean DBP, mean/average diastolic blood pressure.

ratio (63.3%, PIR >1.85), tended to be married or with a long-term partner (71.1%), and had the highest levels of education (25.3% held a college degree and above) and private insurance coverage (40.3%) among the three clusters. However, they also had the highest prevalence of heavy smokers (21.6%) and alcohol users (65.6%), the lowest sleep duration (6.8 hours/night), and the lowest diet pattern score (1.0). Furthermore, they had the highest BMI (29.4) and higher measured metabolic risk factors, including total cholesterol, LDL, and diastolic BP, and the highest prevalence of family history of diabetes (39.5%).

Cluster 2: healthy socioeconomically vulnerable young adults

Cluster 2 (n=13,401 (34.8% of all adults), mean age=29.4SD (10.1), 67.8% female) comprised younger socioeconomically vulnerable adults. They were more likely to be single (61.7%), poor (PIR index <1.30), and racial/ethnic minorities (blacks: 26.3%; Hispanic 38.8%). Compared with the other clusters, cluster 2 had the most participants with no health insurance coverage (35.7%) and no healthcare access (27.3%). Cluster 2 participants had the highest diet pattern scores (1.3), the lowest alcohol use (42.4%), and more favorable metabolic risk profile, including a lower BMI, waist circumference, cholesterol levels, HbA1c, systolic/diastolic BP, uric acid, and BUN.

Cluster 3: elderly adults with chronic conditions and low physical activity levels

Cluster 3 (n=6252 (16.3% of all adults), mean age=65.5SD (15.1), 2.9% female) comprised physically inactive elderly adults with chronic conditions. Notably, most participants reported having health insurance coverage (86.4%) and access to healthcare (87.7%). Cluster 3 participants were the least likely to smoke (heavy smoker: 14.7%), had the longest sleep duration (7.4 hours), and were least physically active (inactive: 67.1%) among the three clusters. They had the highest waist circumference and triglyceride levels, HbA1c, systolic BP, uric acid, and BUN, and the most participants with a history of hypertension (46.4%).

Differences in risk of undiagnosed diabetes and pre-diabetes across clusters

In this study, we found that 2.7% (representing 4,269,098 persons in the USA) of non-pregnant adults with no prior diabetes diagnosis had undiagnosed diabetes, and 22.6% (representing 35,637,895 US adults) had pre-diabetes (online supplemental table 9). Cluster 3 had the highest prevalence of undiagnosed diabetes (11.0%) and pre-diabetes (40.9%), while Cluster 2 had the lowest in undiagnosed diabetes (0.5%) and pre-diabetes (11.4%) (online supplemental table 9). Additionally, undiagnosed diabetes was more prevalent in non-Hispanic blacks (3.4%), followed by non-Hispanic individuals of

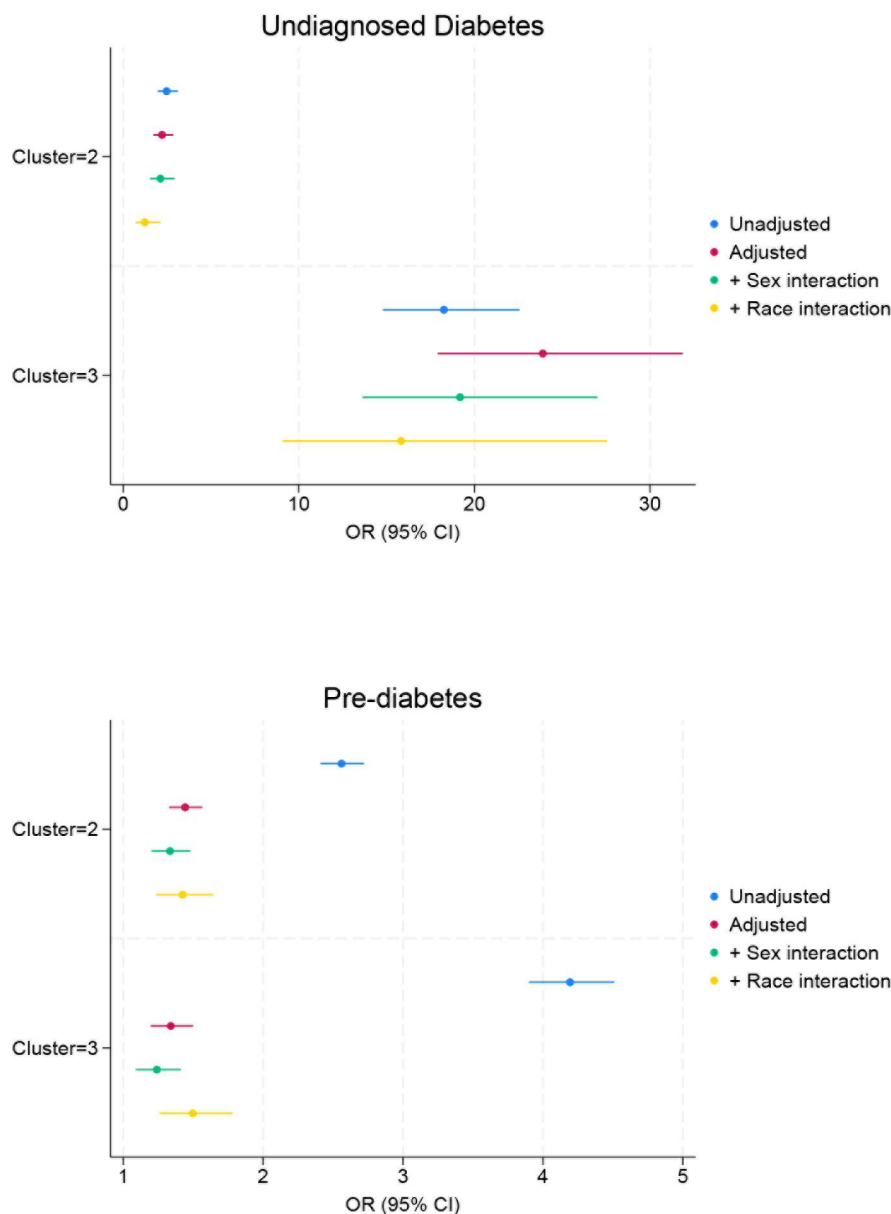


Figure 3 ORs for the association between clusters and undiagnosed diabetes or pre-diabetes. Covariates adjusted in the logistic models: age, sex, race, alcohol use, smoking status, diet score, BMI, history of hypertension, physical activity, education, and PIR. BMI, body mass index; PIR, poverty-income ratio.

other races (3.3%), Hispanics (3.2%), and non-Hispanic white (2.4%) (online supplemental table 10). On the other hand, pre-diabetes was most prevalent among individuals of other races (25.6%) and least prevalent among non-Hispanic whites (22.0%) (online supplemental table 11).

Logistic regression analyses showed that cluster membership was associated with differential risk of undiagnosed diabetes and pre-diabetes (figure 3 and online supplemental table 12). Cluster 3 had more than 25 times greater risk of undiagnosed diabetes (OR 22.5 (95% CI 17.7, 28.6)) and over seven times greater risk of pre-diabetes (OR 7.4 (95% CI 6.8, 8.1)) compared with cluster 2. Adults in cluster 1 were also more likely to have undiagnosed diabetes and pre-diabetes compared with those in cluster 2 (OR 5.2 (95% CI 4.1, 6.6) and 3.8 (95%

CI 3.6, 4.1), respectively). These associations remained significant after adjusting for potential confounders, and there were no significant differences in the risk of undiagnosed diabetes across clusters by race/ethnicity (figure 3 and online supplemental tables 13 and 14). The adjusted ORs for risk of undiagnosed diabetes associated with clusters 1 and 3 were 3.7 and 14.9, respectively ($p < 0.001$ for trend).

However, the association between clusters and pre-diabetes varied by race/ethnicity (p value interaction term=0.003, online supplemental table 13). Table 2 presents the results of racial/ethnic group-specific, fully adjusted regression models assessing the association between cluster membership and pre-diabetes risk. In the fully adjusted models, there was an increased risk of pre-diabetes associated with cluster 3 and cluster 1 for

Table 2 ORs for pre-diabetes for cardiometabolic risk clusters; stratified by race and ethnicity

Model	Cluster 2	Cluster 1	Cluster 3	P value
	Reference	OR (95% CI)	OR (95% CI)	
White				
Unadjusted	1	3.00 (2.63, 3.41)	6.52 (5.67, 7.50)	<0.001
Adjusted	1	1.31 (1.12, 1.53)	1.42 (1.16, 1.73)	0.002
Black				
Unadjusted	1	2.95 (2.62, 3.33)	3.78 (3.21, 4.46)	<0.001
Adjusted	1	1.59 (1.35, 1.88)	1.30 (1.02, 1.65)	<0.001
Hispanic				
Unadjusted	1	2.83 (2.54, 3.14)	3.50 (3.08, 3.98)	<0.001
Adjusted	1	1.54 (1.33, 1.80)	1.36 (1.12, 1.67)	<0.001
Other races*				
Unadjusted	1	3.26 (2.70, 3.94)	3.60 (2.79, 4.64)	<0.001
Adjusted	1	1.66 (1.28, 2.16)	1.06 (0.73, 1.55)	<0.001

*Other races included non-Hispanic Asian and multiracial.
CI, Confidence interval.

non-Hispanic whites, non-Hispanic blacks, and Hispanics. The magnitude of the ORs for cluster 3 was higher in non-Hispanic whites and Hispanics compared with non-Hispanic blacks (table 2). Furthermore, there were no significant differences in pre-diabetes risk between cluster 3 and cluster 2 among non-Hispanic other races.

DISCUSSION

In our study, we employed a novel unsupervised clustering technique and used data from a multiethnic national representative cohort to identify distinct subgroups of US adults with varying levels of cardiometabolic risk. We found heterogeneity in cardiometabolic risk profiles, with different clustering patterns of SBDH and metabolic risk factors. Specifically, we identified three distinct subpopulations with different degrees of risks for undiagnosed diabetes and pre-diabetes. The subpopulations labeled as 'Middle-aged adults with multiple metabolic risk factors' and 'Elderly adults with chronic conditions and low physical activity levels' exhibited significantly elevated risk compared with 'Healthy socioeconomically vulnerable young adults' group. Furthermore, we observed significant variations in the association between pre-diabetes and subpopulations by racial/ethnic group. These findings underscore the importance of comprehensive assessments of SBDH and objective measures of metabolic risk factors in evaluating diabetes risk. This approach could inform the development of more precise and effective screening and prevention strategies that are tailored to the unique needs of at-risk subpopulations.

As the availability of large-scale health data continues to expand, advanced data analytic techniques become increasingly crucial for uncovering complex patterns and relationships among multiple risk factors.³² In this regard, our approach, which incorporates social, behavioral and

metabolic risk factors, offers a promising new tool for improving the accuracy of cardiometabolic risk classification, particularly in populations characterized by diverse social determinants of health. However, further research is needed to elucidate the underlying mechanisms, including potential genetic, environmental, and behavioral factors, that contribute to these clustering patterns. By leveraging large-scale health data and advanced analytical techniques, we can gain deeper insights into the complex interplay between these factors and ultimately develop more effective strategies for preventing cardiometabolic disease.

Our study reveals the variability in cardiometabolic risk profiles among US adults without prior diagnosis of diabetes, highlighting the critical role of SBDH in the development of pre-diabetes and undiagnosed diabetes. While the importance of SBDH has been recognized in previous research,^{33 34} our study confirms their essential role in defining the subpopulation at highest risk for developing pre-diabetes and diabetes. This incorporation of SBDH could improve the phenotyping of pre-diabetes³⁵ and the stratification of cardiometabolic risk at the population level, with significant implications for the prevention and management of diabetes and other chronic diseases.³⁶ Future research should focus on elucidating the underlying biological, social, and behavioral mechanisms contributing to the observed clustering patterns and differential risks of pre-diabetes and undiagnosed diabetes in these subpopulations.

Our study identified two subpopulations at high risk of undiagnosed diabetes: 'Cluster 1: Middle-aged adults with multiple metabolic risk factors' and 'Cluster 3: Elderly adults with chronic conditions and low physical activity levels'. These findings reveal a gradient of increasing risk, with younger participants in cluster 2

being at lower risk compared with middle-aged individuals in cluster 1, who in turn exhibit a lower risk than older adults in cluster 3. This trend is consistent with the well-established association between advanced age and increased risk of diabetes.^{1 37 38} The differences in age across the three clusters may contribute to variations in sociodemographic and metabolic characteristics. Therefore, elucidating the interaction and interplay between age and other factors is crucial for public health organizations to develop precise prevention and control programs for pre-diabetes and diabetes.³⁷ Our analysis extends beyond age as the sole risk factor, highlighting the additional significance of factors such as physical activity levels and comorbidity burden in identifying subpopulation at an even higher risk of undiagnosed diabetes. Moreover, access to healthcare and health insurance coverage emerged as important factors in detecting undiagnosed diabetes. These findings underscore the need for improved diabetes awareness, education, and preventive healthcare, particularly for less physically active older men who may be at risk of suboptimal diabetes testing despite being eligible for screening according to current guidelines.

In our study, we observed that older adults in cluster 3 displayed the longest sleep duration among the three clusters, which may contribute to the heightened risk of undiagnosed diabetes and pre-diabetes in this group. Notably, self-reported long and short sleep durations have been associated with an increased risk of type 2 diabetes.^{39 40} While existing research has primarily focused on the detrimental effects of short sleep duration on cardiometabolic health,^{41 42} emerging evidence indicates that long sleep duration is also related to an elevated risk of diabetes.³⁹ The increased diabetes risk among older adults in cluster 3 could be attributed to potential factors such as poor sleep quality, where the proportion of stage N1 and N2 sleep increases, while stage N3 deep sleep slow wave sleep decreases with age, and time awake after sleep onset tends to rise.⁴² Moreover, obstructive sleep apnea, which is more prevalent in individuals with long sleep duration, is known to be associated with an increased risk of incident diabetes.^{43 44} Additionally, altered levels of leptin and ghrelin, and their impact on appetite and glycemic control, may contribute to the heightened type 2 diabetes risk in cluster 3, particularly as these older adults were the least physically active.^{43 45} While randomized controlled trials are essential to elucidate the mechanisms linking long sleep to diabetes risk, our findings further support the significance of encouraging appropriate sleep duration in delaying or preventing diabetes.

The identification of a specific subgroup characterized by clustered behavioral determinants of health, labeled as 'Middle-aged adults with multiple metabolic risk factors', provides valuable insights into the relationship between health-related behaviors and diabetes risk. Emerging evidence suggests that multiple unhealthy behaviors have a significant impact on the incidence of diabetes.³⁹ In our study, after controlling for baseline metabolic risk,

we found that 'Middle-aged adults with multiple metabolic risk factors' had a nearly 1.5-fold increase in pre-diabetes risk compared with 'Healthy socioeconomically vulnerable young adults'. Similar observations have been made in other studies, including a population-based cohort study of Chinese adults, where a distinct cluster characterized by smoking, heavy drinking, physical inactivity, and insufficient sleep was associated with a higher likelihood of diabetes.⁴⁶ These findings suggest that the identification of adults with multiple unhealthy behaviors holds potential for predicting an increased risk of cardiometabolic disease. Furthermore, lifestyle modification or medications have been shown to prevent the progression from pre-diabetes to diabetes.⁴⁷ Therefore, cluster analyses capturing health-related behavior patterns in the adult population could provide a more effective way for identifying subgroups at risk of developing diabetes compared with conventional prediction methods.

Interestingly, our study found that the 'Healthy socioeconomically vulnerable young adults' subpopulation had a lower risk for both pre-diabetes and undiagnosed diabetes, which contrasts previous research suggesting that individuals with lower socioeconomic status are more susceptible to cardiometabolic disease.^{8 48} This discrepancy in results may be attributed to the younger age of this subpopulation in our study, as the prevalence of pre-diabetes and diabetes typically increases with age.⁴⁹ Moreover, inconsistencies in previous studies examining the inverse association between socioeconomic status and cardiometabolic disease^{50 51} may stem from variations in how socioeconomic status is measured.⁵² Additional factors, such as behavioral and psychosocial differences among specific racial and ethnic groups living in disadvantaged neighborhoods, may account for these disparities.⁵¹ In our study, the 'Healthy socioeconomically vulnerable young adults' subpopulation exhibited the lowest BMI and waist circumference, as well as the highest diet quality and physical activity levels, which likely contributed to their lower risk of pre-diabetes and diabetes. The high proportion of Hispanics in this subpopulation could also be a contributing factor, as Hispanics in the USA have been found to have comparable or better health outcomes than their non-Hispanic white counterparts despite facing higher rates of poverty and lower health insurance coverage.⁵³ Migration history, including the duration of time spent in the US, may play a role in the cardiometabolic risk profile of Hispanics.⁵⁴ Our findings suggest that incorporating SBDHs and migration factors into traditional metabolic risk assessments could help provide a more nuanced understanding of how these factors intersect with race/ethnicity in the development of cardiometabolic disease.

This study offers promising implications for advancing opportunities for cardiometabolic health improvement by addressing SBDH factors. Our findings suggest the potential use of unsupervised cluster analysis to enhance risk stratification at the population level by incorporating

SBDH into the classification of cardiometabolic risk. Clustering analyses have been demonstrated to improve the accuracy of risk prediction compared with traditional risk models.⁵⁵ Although we have not yet tested whether this approach improves the classification of cardiometabolic risk, our results indicate clinically relevant prognostic SBDH differences between subpopulations, providing a framework for developing machine learning algorithms to automatically identify individuals at increased risk for developing diabetes and cardiovascular disease. This approach may be particularly useful for developing tailored, cost-effective preventive strategies based on individuals' SBDH profiles. Our study highlights the importance of addressing health-related social needs in addition to clinical factors when promoting health equity.⁵⁶ It also provides guidance for the recruitment of participants into clinical trials involving diabetes screening, testing, or lifestyle interventions to reduce cardiometabolic risk, as certain subpopulations of adults are at greater risk of pre-diabetes and undiagnosed diabetes and may derive the greatest benefit from such interventions. These findings have implications for national organizations such as the American Diabetes Association⁸ and American Heart Association,⁵⁷ as they work toward improving understanding and addressing social determinants of health in promoting cardiometabolic health.

Limitations

Our study has several limitations that need to be considered. First, using the K-prototype clustering algorithm to categorize the population into discrete clusters may have overlooked the continuous spectrum of health and disease progression. Additionally, the SBDH characteristics of the population may change over time, highlighting the need for further research on the longitudinal change of cluster membership to provide additional information. Despite this, our analysis supports the hypothesis that subpopulations with distinct combinations of SBDH and metabolic risk factors can be revealed using cluster analysis. Second, our SBDH characterizations were based on self-report, which may introduce subjectivity into the analysis, and other variables may be of greater significance in developing meaningful cardiometabolic risk clusters. Moreover, we did not include genetic information in our clustering, which limits our ability to assess the impact of Mendelian pathogenic variants predisposing individuals to cardiometabolic conditions.⁵⁸ Nonetheless, our study benefits from a large multiethnic national representative sample of US adults, and our cardiometabolic risk clustering patterns overlap with those identified by another population study conducted in China.⁴⁶ Third, while our decisions on the number of clusters and variable selection were informed by established methods, there are no well-validated techniques for finding optimal numbers of clusters for K-prototype analysis. Replicating our findings in other population-based datasets is necessary, and our study results should be considered hypothesis-generating only. Last, non-Hispanic Asians and participants of mixed

or other race/ethnicity were grouped together as 'Others' due to the limited sample size available in the dataset. Consequently, we were unable to investigate potential effect modification by these subgroups. Performing further stratification analyses specifically focusing on non-Hispanic Asians could offer valuable insights into the observed absence of differences in pre-diabetes risk between cluster 3 and clusters among the Other races. Existing evidence suggests that individuals of South Asian descent, in particular, may demonstrate a higher risk cardiometabolic risk profile at lower BMI levels when compared with non-Hispanic whites.^{59 60} Nevertheless, our study provides valuable insights into the potential for unsupervised cluster analysis to improve risk stratification and identify subpopulations at increased risk for developing diabetes.

CONCLUSIONS

Our study used an unsupervised machine learning algorithm to reveal the heterogeneity of cardiometabolic risk profiles among US adults without a prior diabetes diagnosis. By identifying three distinct population clusters with unique combinations of SBDH and metabolic risk factors, we have demonstrated that each cluster is associated with different levels of risks for pre-diabetes and undiagnosed diabetes. These findings underscore the potential benefits of utilizing multivariate statistical techniques for comprehensive cardiometabolic risk classification and emphasize the importance of incorporating SBDHs into conventional risk models for clinical screening purposes. Consequently, our results have profound implications for the development of more targeted diabetes screening and prevention strategies. Future studies that consider these distinctive clustering patterns of cardiometabolic risk may contribute to the enhancement of the effectiveness of diabetes screening and prevention efforts.

Author affiliations

¹College of Health and Human Sciences, Purdue University, West Lafayette, Indiana, USA

²Division of Cardiology, Yale School of Medicine, New Haven, Connecticut, USA

³Division of Cardiology, Yale University, New Haven, Connecticut, USA

⁴Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

⁵School of Public Health, Indiana University, Bloomington, Indiana, USA

⁶Department of Medicine, The University of Arizona, Tucson, Arizona, USA

Acknowledgements We are grateful to the participants and staff of the NHANES study. We thank Dr Daniel Robertson, PhD, and Dr Kasia Lipska, MD, MHS, for reviewing the manuscript and providing valuable feedback.

Contributors QD and YL were involved in the conception, design, and conduct of the study, as well as the interpretation of the results. QD and JH had full access to all the data in the study and performed the data analyses. QD and YL drafted the manuscript. JH, TZ, and DGM edited, reviewed, and approved the final version of the manuscript. QD (the guarantor) accepts full responsibility for the study's work and conduct, had access to the data, and controlled the decision to publish.

Funding The study was funded by the Clifford B. Kinley Trust through an award to QD and YL. The funder had no role in the design and conduct of the study; management, analysis, and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer The institutional review board of Purdue University West Lafayette campus determined that this study does not require review as human subjects research as defined under United States Department of Health and Human Services regulations 45 CFR 46.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Individual laboratory measurements and survey responses are publicly released by the Centers for Disease Control and Prevention, National Center for Health Statistics.

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

Qinglan Ding <http://orcid.org/0000-0001-7431-0288>

REFERENCES

- Fang M, Wang D, Coresh J, *et al*. Undiagnosed diabetes in U.S. adults: prevalence and trends. *Diabetes Care* 2022;45:1994–2002.
- Center for Disease Control and Prevention. National diabetes Statistics report: center for disease control and prevention; 2022.
- Dal Canto E, Ceriello A, Rydén L, *et al*. Diabetes as a cardiovascular risk factor: an overview of global trends of macro and micro vascular complications. *Eur J Prev Cardiol* 2019;26:25–32.
- American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. *Diabetes Care* 2018;41:917–28.
- McLaughlin T, Abbasi F, Lamendola C, *et al*. Heterogeneity in the prevalence of risk factors for cardiovascular disease and type 2 diabetes mellitus in obese individuals: effect of differences in insulin sensitivity. *Arch Intern Med* 2007;167:642–8.
- American Diabetes Association. 2. classification and diagnosis of diabetes: standards of medical care in diabetes-2020. *Diabetes Care* 2020;43:S14–31.
- Ali O. Genetics of type 2 diabetes. *World J Diabetes* 2013;4:114–23.
- Hill-Briggs F, Adler NE, Berkowitz SA, *et al*. Social determinants of health and diabetes: a scientific review. *Diabetes Care* 2020;44:258–79.
- Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. *Am J Med Sci* 2016;351:366–73.
- Kim O, Ovbiagele B, Valle N, *et al*. Race-ethnic disparities in cardiometabolic risk profiles among stroke survivors with undiagnosed diabetes and prediabetes in the United States. *J Stroke Cerebrovasc Dis* 2017;26:2727–33.
- Cowie CC, Casagrande SS, Geiss LS, *et al*. Prevalence and incidence of type 2 diabetes and prediabetes. In: Cowie CC, Casagrande SS, Menke A, eds. *Diabetes in America*. Bethesda: National Institute of Diabetes and Digestive and Kidney Diseases (US), 2018.
- Hill J, Nielsen M, Fox MH. Understanding the social factors that contribute to diabetes: a means to informing health care and social policies for the chronically ill. *Perm J* 2013;17:67–72.
- Giuse NB, Koonce TY, Kusnoor SV, *et al*. Institute of medicine measures of social and behavioral determinants of health: a feasibility study. *Am J Prev Med* 2017;52:199–206.
- Huang Z. Extensions to the K-means algorithm for clustering large data sets with categorical values. *Data Min Knowl Discov* 1998;2:283–304.
- Nooraeni R, Arsa MI, Kusumo Projo NW. Fuzzy Centroid and genetic algorithms: solutions for numeric and categorical mixed data clustering. *Procedia Computer Science* 2021;179:677–84.
- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National health and nutrition examination survey data. Hyattsville, MD U.S. Department of Health and Human Services; 2022. Available: <https://wwwn.cdc.gov/nchs/nhanes/>
- Centers for Disease Control and Prevention (CDC). NHANES survey methods and analytic guidelines. Hyattsville, MD U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2020. Available: <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx> [Accessed 12 Dec 2021].
- Kyrou I, Tsigos C, Mavrogianni C, *et al*. Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe. *BMC Endocr Disord* 2020;20:134.
- Whitman A, Lew ND, Chappel A, *et al*. *Addressing social determinants of health: Examples of successful evidence-based strategies and current federal efforts: Assistant Secretary for Planning and Evaluation*. U.S. Department of Health and Human Services, 2022.
- Assistant Secretary for Planning and Evaluation. *U.S. federal poverty guidelines used to determine financial eligibility for certain federal programs*. Washington, D.C: U.S. Department of Health and Human Services, 2021.
- 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.
- Centers for Disease Control and Prevention (CDC). *Physical activity guidelines for Americans, 2nd ed*. Physical activity, 2022.
- Hukkanen J, Jacob P III, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev* 2005;57:79–115.
- Butler L, Popkin BM, Poti JM. Associations of alcoholic beverage consumption with dietary intake, waist circumference, and body mass index in US adults: national health and nutrition examination survey 2003–2012. *J Acad Nutr Diet* 2018;118:409–20.
- Cutler JA, Sorlie PD, Wolz M, *et al*. Trends in hypertension prevalence, awareness, treatment, and control rates in United States adults between 1988–1994 and 1999–2004. *Hypertension* 2008;52:818–27.
- American Diabetes Association. 2. classification and diagnosis of diabetes: standards of medical care in diabetes-2021. *Diabetes Care* 2021;44:S15–33.
- Black MH, Watanabe RM. A principal components-based clustering method to identify variants associated with complex traits. *Hum Hered* 2011;71:50–8.
- Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc B* 2001;63:411–23.
- Aschenbruck R, Szepannek G, Wilhelm AFX. Imputation strategies for clustering mixed-type data with missing values. *J Classif* 2023;40:2–24.
- Stata statistical software: release 17 (program). College Station, TX StataCorp LLC; 2021.
- Rstudio: integrated development for R (program). Boston, MA RStudio, PBC; 2020.
- Big data techniques for public health: a case study. 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); IEEE, 2017
- Schmittiel JA, Dyer WT, Marshall CJ, *et al*. Using neighborhood-level census data to predict diabetes progression in patients with laboratory-defined prediabetes. *Perm J* 2018;22:18–096.
- Galaviz KI, Narayan KMV, Lobelo F, *et al*. Lifestyle and the prevention of type 2 diabetes: a status report. *Am J Lifestyle Med* 2018;12:4–20.
- Færch K, Witte DR, Tabák AG, *et al*. Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol* 2013;1:43–51.
- Stefan N, Fritsche A, Schick F, *et al*. Phenotypes of prediabetes and stratification of cardiometabolic risk. *Lancet Diabetes Endocrinol* 2016;4:789–98.
- Yan Z, Cai M, Han X, *et al*. The interaction between age and risk factors for diabetes and prediabetes: a community-based cross-sectional study. *Diabetes Metab Syndr Obes* 2023;16:85–93.
- Xia M, Liu K, Feng J, *et al*. Prevalence and risk factors of type 2 diabetes and prediabetes among 593,288 middle-aged and elderly adults in China: a cross-sectional study. *Diabetes Metab Syndr Obes* 2021;14:1975–85.
- Shan Z, Li Y, Zong G, *et al*. Rotating night shift work and adherence to unhealthy lifestyle in predicting risk of type 2 diabetes: results from two large US cohorts of female nurses. *BMJ* 2018;363:k4641.

- 40 Leng Y, Cappuccio FP, Surtees PG, *et al.* Daytime napping, sleep duration and increased 8-year risk of type 2 diabetes in a British population. *Nutr Metab Cardiovasc Dis* 2016;26:996–1003.
- 41 Tobaldini E, Fiorelli EM, Solbiati M, *et al.* Short sleep duration and cardiometabolic risk: from pathophysiology to clinical evidence. *Nat Rev Cardiol* 2019;16:213–24.
- 42 St-Onge M-P, Grandner MA, Brown D, *et al.* Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. *Circulation* 2016;134:e367–86.
- 43 Han X, Liu B, Wang J, *et al.* Long sleep duration and afternoon napping are associated with higher risk of incident diabetes in middle-aged and older Chinese: the Dongfeng-Tongji cohort study. *Ann Med* 2016;48:216–23.
- 44 Ding Q, Qin L, Wojeck B, *et al.* Polysomnographic phenotypes of obstructive sleep apnea and incident type 2 diabetes: results from the DREAM study. *Ann Am Thorac Soc* 2021;18:2067–78.
- 45 Lucassen EA, Rother KI, Cizza G. Interacting epidemics? Sleep curtailment, insulin resistance, and obesity. *Ann N Y Acad Sci* 2012;1264:110–34.
- 46 Zhang G, Luo C, Cui Y, *et al.* Clustering of multiple health risk behaviors and its association with diabetes in a Southern Chinese adult population: a cross-sectional study. *PeerJ* 2020;8:e9025.
- 47 Knowler WC, Barrett-Connor E, Fowler SE, *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002;346:393–403.
- 48 Seligman HK, Jacobs EA, López A, *et al.* Food insecurity and glycemic control among low-income patients with type 2 diabetes. *Diabetes Care* 2012;35:233–8.
- 49 Richter B, Hemmingsen B, Metzendorf M-I, *et al.* Development of type 2 diabetes mellitus in people with intermediate hyperglycaemia. *Cochrane Database Syst Rev* 2018;10:CD012661.
- 50 Agardh E, Allebeck P, Hallqvist J, *et al.* Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int J Epidemiol* 2011;40:804–18.
- 51 Diez Roux AV, Jacobs DR, Kiefe CI, *et al.* Neighborhood characteristics and components of the insulin resistance syndrome in young adults: the coronary artery risk development in young adults (CARDIA) study. *Diabetes Care* 2002;25:1976–82.
- 52 Pollitt RA, Rose KM, Kaufman JS. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. *BMC Public Health* 2005;5:7.
- 53 Hunt KJ, Resendez RG, Williams K, *et al.* All-cause and cardiovascular mortality among Mexican-American and non-Hispanic white older participants in the San Antonio heart study- evidence against the "Hispanic paradox" *Am J Epidemiol* 2003;158:1048–57.
- 54 Bostean G. Does selective migration explain the hispanic paradox? A comparative analysis of Mexicans in the U.S. and Mexico. *J Immigr Minor Health* 2013;15:624–35.
- 55 Sharma A, Sun J-L, Likhnygina Y, *et al.* Patient phenotypes, cardiovascular risk, and Ezetimibe treatment in patients after acute coronary syndromes (from IMPROVE-IT). *Am J Cardiol* 2019;123:1193–201.
- 56 Thornton PL, Kumanyika SK, Gregg EW, *et al.* New research directions on disparities in obesity and type 2 diabetes. *Ann N Y Acad Sci* 2020;1461:5–24.
- 57 Havranek EP, Mujahid MS, Barr DA, *et al.* Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* 2015;132:873–98.
- 58 O'Sullivan JW, Raghavan S, Marquez-Luna C, *et al.* Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* 2022;146:e93–118.
- 59 Misra A. Ethnic-specific criteria for classification of body mass index: a perspective for Asian Indians and American diabetes association position statement. *Diabetes Technol Ther* 2015;17:667–71.
- 60 Zhu L, Yang WJ, Spence CB, *et al.* Lean yet unhealthy: Asian American adults had higher risks for metabolic syndrome than non-Hispanic white adults with the same body mass index: evidence from NHANES 2011–2016. *Healthcare (Basel)* 2021;9:1518.