
Neftali Eduardo Antonio-Villa MD1, *, Luisa Fernandez-Chirino2, * BSc, Arsenio Vargas-Vázquez MD1, Jessica Paola Bahena-Lopez1 MD, Carlos A. Aguilar-Salinas3 MD, PhD and Omar Yaxmehen Bello-Chavolla MD, PhD4

1MD/Ph.D. (PECEM), Faculty of Medicine, National Autonomous University of Mexico.
2Faculty of Chemistry, National Autonomous University of Mexico
3División de Nutrición, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubiran, Mexico
4Dirección de Investigación, Instituto Nacional de Geriatría.

*These authors contributed equally in the drafting of this manuscript

Correspondence: Omar Yaxmehen Bello-Chavolla. Dirección de Investigación. Instituto Nacional Geriatría. Anillo Perif. 2767, San Jerónimo Lídice, La Magdalena Contreras, 10200, Mexico City, Mexico. Phone: +52 (55) 5548486885. E-mail: oyaxbell@yahoo.com.mx

CONFLICT OF INTERESTS: The authors declare that they have no conflict of interests.

2,040 text words; 16 references; 2 figures.

Running headline: Trends in diabetes subgroups in NHANES 1988-2018

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT (248 WORDS)

AIMS: Data-driven diabetes subgroups have been proposed as an alternative to address diabetes heterogeneity; changes in trends for these subgroups have not previously been reported. Here, we analyzed trends of diabetes subgroups, stratified by sex, race, education level, and age categories in the U.S.

METHODS: We used data from consecutive NHANES cycles spanning the 1988-2018 period. Diabetes subgroups (mild obesity-related [MOD], severe-insulin deficient [SIID], severe-insulin resistant [SIRD], and age-related diabetes [MARD]) were classified using self-normalizing neural networks. SAID was assessed for NHANES-III cycles. Prevalence was estimated using examination sample weights considering 2-year cycles (biannual change [B.C.]) to evaluate trends.

RESULTS: Diabetes prevalence in the US increased from 7.5% (95% CI 7.1-7.9) in 1988-1989 to 13.9% (95% CI 13.4-14.4) in 2016-2018 (BC 1.09%, 95% CI 0.98-1.31, p<0.001). Non-Hispanic Blacks had the highest prevalence. Overall, MOD, MARD, and SIID had the highest increase during the studied period. Non-Hispanic Blacks had a sharp increase in MARD and SIID, Mexican Americans in SIID, and non-Hispanic Whites in MARD. Males, subjects with primary school/no-education, and adults aged 40-64 years had the highest increase in MOD prevalence. Trends in diabetes subgroups sustained after stratification by body-mass index categories.

CONCLUSIONS: The prevalence of diabetes and its data-driven subgroups in the U.S. has increased from 1988-2018. These trends were different across sex, ethnicities, education, and age categories, indicating significant heterogeneity in diabetes within the U.S. Sex-specific factors, population aging, and socioeconomic aspects, together with obesity prevalence increase, could be implicated in the uprising trends of diabetes in the U.S.
RESEARCH IN CONTEXT

What is already known about this subject?

- Diabetes is a heterogeneous disease with increasing prevalence worldwide.
- An approach to characterize diabetes heterogeneity uses data-driven subgroups, including mild obesity-related (MOD), severe-insulin deficient (SIID), severe-insulin resistant (SIRD), severe autoimmune (SAID) and age-related diabetes (MARD).
- Prevalence trends of data-driven diabetes subgroups in the U.S. remain as an area of opportunity to assess diabetes heterogeneity over time in the U.S.

What is the key question?

- What are the diabetes subgroup prevalence trends in adult subjects from the National Health and Nutrition Examination Survey in the U.S. from 1988 to 2018?

What are the new findings?

- Diabetes prevalence in the U.S. has steadily increased from 1988 to 2018, with Non-Hispanic Blacks and Mexican Americans having the highest diabetes prevalence.
- MOD, MARD, and SIID had the highest increase during the studied period; SIRD and SAID remained unchanged.
- Males, subjects with primary school/no-education, and adults aged 40-64 years had sharp increases in MOD.

How might this impact on clinical practice in the foreseeable future?

- The increase in specific diabetes subgroups could help to implement tailored preventive measures depending on age and ethnicity. Targeted treatment and public health measures should be informed using clinical characteristics for each subgroup.
INTRODUCTION

Diabetes is a highly heterogeneous disease with complex and various underlying etiologies, which result in diverse clinical courses, treatment response, and risk profiles for chronic micro and macrovascular complications. Recent studies have reported steady increases in diabetes prevalence worldwide due to various environmental, genetic, and lifestyle-related factors (1). As an approach to address diabetes heterogeneity, Alhqvist et al. proposed the consideration of five data-driven diabetes subgroups, including severe-autoimmune (SAID), mild age-related (MARD), mild obesity-related (MOD), severe-insulin deficient (SIID) and severe-insulin resistant (SIRD) (2). Data-driven diabetes subgroups have been reproduced independently in other populations and have shown have unique genetic, risk and clinical profiles which are likely to influence assessments of diabetes heterogeneity and clinical course differently across ethnicities and age groups (3).

Although the application of diabetes subgroups within clinical practice could help identify the risk for associated complications, expected treatment response and clinical course towards its use as an approach to further personalized medicine in diabetes, its determinants at a population level and its potential usefulness to inform public health policies remains unclear.

Previous reports of data-driven diabetes subgroups in the US have largely focused on punctual prevalence estimations and its implications for mortality. However, the implications of these subgroups to understand diabetes heterogeneity and the evolution of diabetes trends according to known modifiers of diabetes prevalence have not previously been explored. Here, we analyzed trends of data-driven diabetes subgroups using data from the National Health and Nutrition Examination Surveys spanning a three-decade period in the U.S. from 1988-2018. To evaluate the influence of obesity, population aging and sociodemographic factors which likely influence diabetes prevalence trends over time, we further stratified trends of data-driven diabetes subgroup prevalence over time by ethnicity, sex, age, education level, and body-mass index (BMI) categories.
METHODS

Study subjects
We used U.S. citizen data from the 1988-2018 cycles of NHANES. NHANES had the approval of the NCHS Ethics Review Board. Details of response rate, medical questionnaires, anthropometrical and biochemical measurements for each cycle are reported elsewhere (4). We included data of adults ≥18 years with complete data to estimate diabetes subgroups. Diabetes was defined as glycated hemoglobin (HbA1c) ≥6.5%, fasting plasma glucose ≥7.0 mmol/L (≥126 mg/dl), or previous diabetes diagnosis following current ADA guidelines (5). A flowchart of data extraction is available in Supplementary Figure 1.

Diabetes subgroup estimation in NHANES
Prediction of data-driven diabetes subgroups was based on the algorithms previously developed by Bello-Chavolla et al. Subgroup classification is based on a supervised machine learning approach which uses self-normalizing artificial neural networks (SNNN) trained with surrogate metabolic measures to estimate insulin action related phenomena from population-based studies and which show higher concordance compared to results from k-means clustering. We used model 1 for the prediction of diabetes subgroups in those patients living with diabetes from NHANES 1999-2018 using C-peptide measures and model 2 for estimation using insulin-based measures where available. The algorithm uses as input variables HOMA2-IR, HOMA2-β, BMI, HbA1c, years since diabetes onset standardized using means from NHANES-III values. For the subgroup estimation, we used the electronic-based application, available at: https://uiem.shinyapps.io/diabetes_clusters_app/. Since the algorithm does not consider SAID classification, we defined SAID using anti-glutamic acid decarboxylase antibodies (anti-GAD65, index values >0.069), which were only available for the 1988-1991 NHANES cycles (NHANES-III); therefore, analyses including SAID were only included as a weighted sub-analysis.

Weighted prevalence and trend analyses
Prevalence of diabetes was estimated using examination sample weights from NHANES; for diabetes subgroups, we subset participants with available insulin measures; all estimations were conducted using the `survey` R package. We fitted linear regression analyses using 2-year cycles (biannual cycles, BC) as continuous variables was considered to evaluate trends over time (7). To evaluate the effect of aging, sociodemographic factors and obesity as modifiers of diabetes subgroup trends, we performed weighted subgroup analyses for trends stratified by age, sex, ethnicity (codified as Mexican Americans, Non-Hispanic Whites and Non-Hispanic Blacks), education level (codified as primary school/no-education, secondary/high-school/A.A. degree and college/higher). Statistical significance was established as a p<0.05 threshold. All statistical analyses were conducted using R software (v4.0.0).

RESULTS

Trends in diabetes subgroups

We used data from 73,594 adults. Diabetes prevalence in the US increased from 7.5% (95%CI 7.1-7.9) in 1988-1989 to 13.9% (95%CI 13.4-14.4) in the 2016-2018 cycle (BC 1.09%, 95%CI 0.87-1.31, p<0.001). Regarding diabetes subgroups, MOD had the greatest increase in prevalence over time (from 2.4%, [95%CI 2.2-2.6] to 6.4% [95%CI 3.9-4.6]; BC: 0.52% [95%CI 0.37-0.66], p<0.001), followed by MARD (from 1.2%, [95%CI 1.1-1.3] to 3.9% [95%CI 1.3-4.3]; BC: 0.44% [95%CI 0.27-0.62], p<0.01) and SIID (from 1.9%, [95%CI 1.7-2.1] to 2.9% [95%CI 2.7-3.1]; BC: 0.18% [95%CI 0.07-0.31], p<0.01). SIRD had non-significant increases in prevalence during the 1988-2018 period (Figure 1). SAID remained stable over the 6-year period (1988-1994) in which its prevalence could be assessed (Supplementary Figure 2).

Distribution of classification variables for data-driven diabetes subgroups, its overall frequency and stratification by ethnicity are available in detail within Supplementary Figures 3, 4 and 5, respectively.

Sex-specific trends in diabetes subgroups
We observed the largest increases in diabetes prevalence during the studied period for men (BC: 1.4%, 95%CI 1.10-1.64, p<0.001) when compared to women (BC: 0.83%, 95%CI 0.60-1.06, p<0.001). When evaluating diabetes subgroups, men had sustained increases in MOD, MARD, and SIID prevalence over time; in contrast, increases in subgroup prevalence for women were only observed for MOD and MARD (Figure 1; Supplementary Table 1).

*Diabetes subgroup prevalence trends are different across ethnicities.*

Ethnicity had meaningful effects on the trajectory of diabetes prevalence trends; notably, diabetes subgroup prevalence increased unequally across ethnicities. Non-Hispanic Blacks (from 11.1%, [95%CI 10.1-12.1] to 16.7% [95%CI 16.0-17.4]; BC: 1.0% [95%CI 0.6-1.4], p<0.01) and Mexican Americans (from 9.0%, [95%CI 8.3-9.7] to 14.5% [95%CI 13.4-15.6]; BC: 0.9% [95%CI 0.6-1.2], p<0.01) had the highest prevalence of diabetes compared with Non-Hispanic Whites (from 1.9%, [95%CI 6.5-7.5] to 13.3% [95%CI 12.7-13.9]; BC: 1.1% [95%CI 0.8-1.3], p<0.001). Regarding specific diabetes subgroups, non-Hispanic Blacks had marked increases in MOD, MARD, and SIID, Mexican Americans in MOD and SIID, and non-Hispanic Whites in MOD and MARD (Figure 2; Supplementary Table 2).

*Lower-educational level impacts on diabetes subgroup trends*

We observed significant differences in diabetes prevalence trends when stratifying by education level. Notably, adults with primary-school/no-education as their highest attained education level had the largest increase in the prevalence of diabetes compared to higher education attainment. When assessing diabetes subgroup prevalence, we observed an accentuated increase in MOD prevalence for adults with primary school/no-education and secondary/high-school/A.A. degree, in whom we also observed a large increase in SIID prevalence; in contrast, MARD prevalence increased in those with college or higher attainment (Figure 2; Supplementary Table 2). Assessing the effect of education and ethnicity, only MARD and MOD prevalence increased in non-Hispanic Whites with a college education or higher (Supplementary Figure 6).
Effect of age on diabetes subgroup prevalence trends is modified by ethnicity.

Subjects aged >65 years had the highest increase in diabetes prevalence during the studied period. In adults aged 18-39, MOD increased in prevalence over time. Both MOD and MARD prevalence increased for subjects aged 40-64 years; for older adults >65 years, MARD and MOD had the highest increases in prevalence over time (Figure 2; Supplementary Table 2). When assessing the differential effect of age by ethnicity, we identified marked differences in diabetes subgroup distribution. Of notice, non-Hispanic Blacks had an upward slopping trend for the 40-64 age category (Supplementary Figure 7).

Impact of BMI on diabetes subgroup prevalence trends

Adults with obesity had the highest increase in diabetes prevalence compared with overweight and normal-weight subjects. Normal-weight subjects had a significant upward trend in SIID prevalence; for overweight and obese subjects, these upwards trends increased for both MARD and MOD (Supplementary Table 3 and Supplementary Figure 7).

DISCUSSION

Here, we show an increase in the overall prevalence of diabetes and its data-driven subgroups using U.S. citizens data from consecutive NHANES cycles spanning a three-decade period from 1988 to 2018. Data-driven diabetes subgroup trends were different when considering age groups, ethnicities, and education level, establishing factors which influence the heterogeneity in diabetes within the U.S. and unequal distribution of the increased burden of diabetes overtime on underrepresented minorities, subjects with fewer years of education, and older adults. Overall, we observed that sex, ethnicity, education, and age all significantly impacted the increasing prevalence of diabetes subgroups, indicating that the contribution of socioeconomic factors, obesity, and population aging is partly responsible for the overall increase in diabetes prevalence heterogeneity. Interestingly, we also observed that men had sharper increases in diabetes prevalence, particularly for SIID, which could be attributed to sex-based differences in treatment adherence and glycemic control (8). Notably, we observed
that MOD and MARD prevalence consistently increased across most ethnic and age groups in both sexes, indicating that milder phenotypes had the largest increase over time.

As previously reported, diabetes prevalence varies significantly by ethnicity. When stratifying by diabetes subgroups, the differences in the evolution of diabetes trends and its heterogeneity becomes more apparent. Notably, Non-Hispanic Blacks and Mexican Americans had marked increases in MOD, MARD, and SIID prevalence, while non-Hispanic Whites had increases predominantly on MOD and MARD, reflecting disparities underlying cardio-metabolic burden and glycemic control influenced by ethnicity (9). Our results are consistent with the reported rise in obesity prevalence in the U.S., particularly in Mexican-Americans and non-Hispanic Blacks (10). Previous research has shown that non-Hispanic Blacks and Mexican-Americans develop diabetes at lower BMI and have on average, an earlier age of diabetes onset compared to non-Hispanic Whites, as observed in MOD (11). A previous diabetes subgroup study in Mexicans identified a higher incidence of SIRD and MOD in cases with higher rates of metabolic syndrome traits and a lower incidence of SIID and MARD, which has reported a higher prevalence in Europeans (6,11). These differences are likely attributable to ethnic-specific genetic differences in metabolic function, differences in age structures across different populations, and ethnic disparities in management and occurrence of cardiometabolic risk factors (12,13). Our study provides relevant insights on ethnic-specific differences in diabetes heterogeneity within the U.S., consistent with previous studies using NHANES-III data, and identifies ethnic groups that may be particularly vulnerable to the rise in obesity and population aging (10). Furthermore, our study shows that education level plays a significant role in modifying diabetes prevalence and heterogeneity over time; most notably, this became apparent for non-Hispanic Whites. The role of education level and ethnicity in modifying the clinical course of diabetes had been previously explored, which could partially explain the differences observed in our study (14). Further pathophysiological, genetic, and socioeconomic determinants in diabetes phenotypes across
ethnic groups in the U.S. should be assessed to understand their diabetes subgroup incidence role.

In summary, the prevalence of diabetes and its subgroups have increased from 1988-2018 amongst all evaluated groups. Limitations include that risk factors related to subgroup incidence cannot be evaluated in a cross-sectional survey, opening further study opportunities. Similarly, we only assessed SAID prevalence in the 1988 to 1994 cycles due to lack of auto-antibody measurements for other NHANES cycles, limiting our ability to study changes in the prevalence of autoimmune diabetes over time. We also acknowledge that criteria for the definition of diabetes have varied over time within the study period, which could have influenced the estimation of diabetes prevalence, particularly for earlier NHANES cycles. Finally, since data required for the identification of data-driven diabetes subgroups are only available for a subset of NHANES participants, we could not explore all possible category combinations to assess the combined effect of all evaluated covariates, which remains an area of opportunity for further assessment. Obesity is an essential determinant of the increases in diabetes prevalence over time (15); however, our results suggest that this is true primarily for MOD and SIRD, but not MARD or SIID. Along with increases in diabetes prevalence trends, our results also reflect increases in diabetes heterogeneity, which are likely to complicate treatment standardization and suggest differential effects of public health measures to reduce diabetes burden, which should take into account some of the modifiers we explored in this study. Increases in diabetes heterogeneity and subgroup prevalence over time are likely attributable to a multifactorial cause, including but not restricted to increasing obesity rates, population aging, socioeconomic disparities, and lifestyle, which should be assessed longitudinally and inform public policy (16).

ACKNOWLEDGMENTS: NEAV, AVV and JPBL are enrolled at the PECEM program of the Faculty of Medicine at UNAM and are supported by CONACyT.

CONFLICT OF INTERESTS: The authors declare that they have no conflict of interests.
DATA AVAILABILITY: All data sources and R code are available for reproducibility of results at: https://github.com/oyaxbell/clusters_nhanes

AUTHORS' CONTRIBUTIONS: Research idea and study design NEAV, LFC, OYBC, AVV; data acquisition: NEAV, LFC; data analysis/interpretation: NEAV, LFC, OYBC, AVV, JPBL, CAAS; statistical analysis: NEAV, LFC, OYBC; manuscript drafting: NEAV, LFC, OYBC, AVV, JPBL, CAAS; supervision or mentorship: OYBC, CAAS. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions about the accuracy or integrity of any portion of the work are appropriately investigated and resolved.
**FIGURE LEGENDS**

**Figure 1:** Trends on diabetes and its subgroups (A) and its stratification by sex (B) in the National Health and Nutrition Examination Survey from 1988-2018. Prevalence estimations are weighed according to survey design. **Abbreviations:** MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.
Figure 2: Trends on diabetes and its subgroups stratified by ethnicity (A) codified as Mexican Americans, Non-Hispanic Whites and Non-Hispanic Blacks, an education level (B) codified as primary school/no-education, secondary/high-school/A.A. degree and college/higher, and age categories (C) in the National Health and Nutrition Examination Survey from 1988-2018. Prevalence estimations are weighed according to survey design. Abbreviations: MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes.
REFERENCES


