Temitope Olotu - APN 775 - Week 10

Methods

Study Design and Data Source

This study was a secondary analysis of publicly available data obtained from the National Health and Nutrition Examination Survey (NHANES) 2017–2018 cycle.[1-5] NHANES is a nationally representative, cross-sectional survey conducted by the National Center for Health Statistics using a complex, multistage probability sampling design.[1] The purpose of the analysis was to examine whether limited English proficiency is associated with glycemic control, as measured by HbA1c levels, and the use of diabetes education services among U.S. adults aged 20 years and older with diagnosed diabetes. Data files were downloaded from the Centers for Disease Control and Prevention (CDC) NHANES website and imported into Stata Version 19.

The analytic dataset included adults in the specified age group who were diagnosed with diabetes. It was created by merging the NHANES Demographics (DEMO_J), Diabetes Questionnaire (DIQ_J), and Glycohemoglobin Laboratory (GHB_J) datasets using the participant identification number (SEQN).[1-5] Limited English proficiency, defined using interview language and interpreter-use variables, served as the primary exposure. The outcomes were glycated hemoglobin (HbA1c), measured by the laboratory variable LBXGH, and utilization of diabetes education services, measured by the questionnaire variable DIQ230.[1-2] Age, sex, race/ethnicity, education level, income-to-poverty ratio, and health insurance status were included as covariates in the regression analyses.[2,5]

Study Population and Sample Size

The analytic sample included adults aged 20 years and older who participated in NHANES 2017–2018 and reported having been told by a doctor or other health professional that they had diabetes (DIQ010).[2] Participants were eligible if they were aged 20 years or older and had available data on interview language or interpreter use, which was used to define limited English proficiency, the primary exposure variable. Participants were restricted to adults aged 20 years and older with a self-reported physician diagnosis of diabetes (DIQ010 = 1) and available HbA1c laboratory measurements. Separate analytic samples were created for each study outcome. Participants included in the HbA1c analysis had complete data for HbA1c (LBXGH) and limited English proficiency, while participants included in the diabetes education analysis had complete data on diabetes service use (DIQ230) and limited English proficiency.

Participants with missing data for the outcome being analyzed or for limited English proficiency were excluded from that analysis. The adjusted regression analyses included only participants with complete data on all model variables, including age, sex, race/ethnicity, education level, income-to-poverty ratio, and health insurance status. The final analytic sample size for each outcome was determined after merging the NHANES datasets and applying the inclusion and exclusion criteria. It was reported separately for the analyses of HbA1c and the use of diabetes education services.

As a supplemental plan, an a priori power analysis was conducted using GPower version 3.1.9.7.[6-8] Using the primary outcome, mean HbA1c level, the analysis used an independent-samples t-test to compare adults with and without limited English proficiency. Assuming a medium effect size (Cohen's d = 0.50), a significance level of $\alpha = 0.05$, and 95% statistical power, the analysis indicated that a minimum total sample size of 210 participants (105 participants per group) was required. Because the study used existing NHANES data, no attrition adjustment was required. The initial analytic sample included 836 participants who met the study inclusion criteria. After excluding participants with missing outcome data, 804 participants remained for the HbA1c analyses and 825 participants remained for the diabetes education service utilization (DIQ230) analyses. For the adjusted regression models, complete-case analysis was used, resulting in final analytic samples of 698 participants for the HbA1c analyses and 716 participants for the diabetes education service utilization analyses after excluding participants with missing values for one or more covariates included in the models. Both final analytic samples exceeded the minimum sample size of 210 participants estimated by the a priori power analysis.

The GPower output for the a priori power analysis is provided in Appendix A.

Study Variables

The primary exposure variable in this study was the limited English proficiency (LEP), defined using NHANES interview language and interpreter-use variables. The primary outcome was glycemic control, measured by glycated hemoglobin (HbA1c). The secondary outcome was the utilization of diabetes education services, measured by the NHANES diabetes questionnaire variable DIQ230, which assesses how long ago the participant last saw a diabetes nurse educator, registered dietitian, or nutritionist for diabetes care. Additional demographic variables, including age, sex, race/ethnicity, education level, income-to-poverty ratio, and health insurance status, were included as covariates in the adjusted regression analyses.

Independent Variable

The primary independent variable was limited English proficiency (LEP), derived using the NHANES Demographics variables SIALANG (language of the Sample Person interview) and SIAINTRP (whether an interpreter was used during the Sample Person interview).[2] Participants were classified as having limited English proficiency if the Sample Person interview was conducted in Spanish (SIALANG = Spanish) or if an interpreter was used to conduct the interview (SIAINTRP = Yes).[2] Participants whose interviews were conducted in English without an interpreter (SIALANG = English and SIAINTRP = No) were classified as non-LEP and served as the comparison group. The NHANES variables SIALANG and SIAINTRP did not contain "Refused," "Don't know," or other invalid response categories in the 2017–2018 dataset; therefore, these variables were used as originally coded to derive the LEP variable.

Dependent Variables

- **Glycemic Control**: This was the primary outcome. Glycemic control was measured using glycated hemoglobin (HbA1c) percentages, identified by the NHANES laboratory variable LBXGH – Glycohemoglobin (HbA1c %).[4]
- **Utilization of Diabetes Education Services:** This was the secondary outcome. Use of diabetes education services was measured using the NHANES diabetes questionnaire variable DIQ230, which assessed the time since the participant last saw a diabetes nurse educator, dietitian, or nutritionist by asking the respondents, “How long ago did you see a diabetes specialist?”[3] The following original five ordered response categories were retained for the ordinal logistic regression analysis:
    - 1 year ago or less
    - More than 1 year but no more than 2 years ago
    - More than 2 years but no more than 5 years ago

- o More than 5 years ago
- o Never

Descriptive Variables

Descriptive variables included age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH3), education level (DMDEDUC2), family income-to-poverty ratio (INDFMPIR), and health insurance status (HIQ011). [2,5] These variables were used to describe the characteristics of the study sample and to compare participants with limited English proficiency and English-proficient participants. In addition, age, sex, race/ethnicity, education level, family income-to-poverty ratio, and health insurance status were included as covariates in the adjusted survey-weighted linear and logistic regression models to account for potential confounding.

Data Management

The NHANES 2017–2018 datasets used in this study included the Demographics file DEMO_J, the Diabetes questionnaire file DIQ_J, and the Glycohemoglobin laboratory file GHB_J.[2-4] The statistical analyses were performed using Stata Version 19 (StataCorp LLC, College Station, TX). The datasets were downloaded from the NHANES website and imported into Stata in SAS transport (.XPT) format.[11] Each dataset was reviewed to verify variable names, coding, missing values, and the number of observations.[12,13] The datasets were merged using the unique participant identifier (SEQN).[12,13] Merge success was verified using the "_merge" indicator before the variable was removed from the final analytic dataset. Following the merge, the dataset was restricted to adults aged 20 years and older with diagnosed diabetes.[12,13]

The NHANES codebooks were carefully reviewed before data cleaning. Responses coded as "Refused", "Don't know," or other invalid responses were recoded to missing values according to the NHANES codebooks before analysis.[1-5] Variables with valid NHANES skip-pattern missing values were retained as missing. Variables that did not contain these response categories, including SIALANG and SIAINTRP, were retained as originally coded. A binary limited English Proficiency (LEP) variable was created from the interview language and interpreter-use variables. The final analytic sample for each outcome was confirmed after applying the inclusion and exclusion criteria, and the number of participants included in each analysis was reported.

Statistical Analysis and Survey Design

The NHANES 2017–2018 data were analyzed using survey procedures to account for the complex, multistage probability sampling design.[15] Before conducting any analyses, the survey design was specified in Stata using the command svyset SDMVPSU [pweight = WTMEC2YR], strata (SDMVSTRA). The NHANES survey design variables SDMVPSU (primary sampling unit), SDMVSTRA (sampling strata), and the two-year Mobile Examination Center sample weight (WTMEC2YR) were used to produce nationally representative estimates.

Descriptive statistics were used to summarize participant characteristics.[9,10] Survey-weighted means with 95% confidence intervals were calculated for continuous variables, including age, family income-to-poverty ratio, and HbA1c.[9,10] Survey-weighted frequencies and percentages were calculated for categorical variables, including sex, race/ethnicity, limited English proficiency status, health insurance, education level, and time since the participant last saw a diabetes nurse educator, dietitian, or nutritionist.[9,10]

Assumptions of linear regression, including linearity, normality of residuals, homoscedasticity, multicollinearity, and influential observations, were evaluated before fitting the final model.[9,10,16-19] Survey-weighted linear regression (svy: regress) was used to examine the association between limited English Proficiency and HbA1c.[9,10,16-19] Both unadjusted and adjusted models were estimated. The adjusted model controlled for age (RIDAGEYR), sex (RIAGENDR), race/ethnicity (RIDRETH3), education level (DMDEDUC2), family income-to-poverty ratio (INDFMPIR), and health insurance status (HIQ011).[1-5]

Survey-weighted ordinal logistic regression (svy: ologit) was used to examine the association between limited English proficiency and the time since participants last saw a diabetes nurse educator, dietitian, or nutritionist (DIQ230).[9,10,16-18] Both unadjusted and adjusted models were estimated using the same covariates as the HbA1c model.

Participants with missing HbA1c measurements (n = 32) were excluded only from the HbA1c analyses, resulting in an outcome-specific analytic sample of 804 participants. Participants with missing DIQ230 responses (n = 11) were excluded only from the diabetes education service utilization analyses, resulting in an outcome-specific analytic sample of 825 participants. For the adjusted regression analyses, complete-case analysis was used, resulting in final analytic samples of 698 participants for the HbA1c model and 716 participants for the diabetes education service utilization model after excluding participants with missing data for one or more covariates.

Statistical significance was assessed using the two-sided tests with an alpha level of 0.05. Regression coefficients (β), 95% confidence intervals, and p-values were reported, as appropriate.

Ethical Considerations

This study used publicly available, de-identified NHANES data. Because the data contain no personally identifiable information, this secondary data analysis did not require Institutional Review Board (IRB) approval.

Results

Participant Characteristics

The final analytic sample included 836 U.S. adults aged 20 years and older with diagnosed diabetes. Participant characteristics are presented in Table 1. Participants had a survey-weighted mean age of 61.68 years and a survey-weighted mean HbA1c of 7.28%. Approximately 9.92% of participants were classified as having limited English proficiency, whereas 90.08% were classified as English proficient. Most participants had health insurance coverage, and the largest racial/ethnic group was non-Hispanic White.

Additional demographic, socioeconomic, and clinical characteristics are presented in Table 1.

Table 1. Characteristics of U.S. Adults Aged ≥20 Years with Diagnosed Diabetes, NHANES 2017–2018

Characteristic	Unweighted n	Survey-weighted estimate
Continuous variables		
Age, years	836	61.68 (95% CI, 59.84–63.52)
HbA1c, %	804	7.28 (95% CI, 7.12–7.44)
Family income-to-poverty ratio	724	2.99 (95% CI, 2.75–3.23)
Limited English proficiency		
Non-LEP	753	90.08%
LEP	83	9.92%
Sex		
Male	454	54.23%
Female	382	45.77%
Race/Ethnicity		
Mexican American	123	8.76%
Other Hispanic	74	5.38%
Non-Hispanic White	286	62.47%
Non-Hispanic Black	194	11.31%
Non-Hispanic Asian	112	6.25%
Other Race/Multiracial	47	5.82%
Education level		
Less than 9th grade	121	7.29%
9th–11th grade	105	9.31%
High school graduate/GED	194	29.09%
Some college/AA degree	259	31.97%
College graduate or above	155	22.34%
Health insurance status		
Yes	762	92.74%
No	74	7.26%

Abbreviations: *HbA1c = glycated hemoglobin; LEP = limited English proficiency.*
Note. *Total unweighted sample = 836. Unweighted sample sizes vary across variables because of missing data. Continuous variables are presented as survey-weighted means (95% confidence intervals). Categorical variables are presented as survey-weighted percentages. Estimates account for the complex NHANES sampling design.*

Regression Analyses

Complete-case analyses resulted in final analytic samples of 698 participants for the HbA1c regression analyses and 716 participants for the diabetes education service utilization regression analyses after excluding participants with missing outcome or covariate data.

The results of the crude and adjusted survey-weighted regression analyses are presented in Table 2. In the crude linear regression model, limited English proficiency (LEP) was significantly associated with higher HbA1c levels ($\beta = 0.54$, 95% CI: 0.13–0.95, $p = 0.013$). However, after adjustment for age, sex, race/ethnicity, education level, family income-to-poverty ratio, and health insurance status, this association was no longer statistically significant ($\beta = -0.20$, 95% CI: −0.62 to 0.22, $p = 0.323$).

Figure 1 presents the adjusted predicted HbA1c values by LEP status from the fully adjusted survey-weighted regression model. After adjustment for demographic and socioeconomic characteristics, adults with LEP had a slightly lower predicted mean HbA1c (7.08%) than English-proficient adults (7.28%). The substantial overlap in the 95% confidence intervals is consistent with the adjusted regression results, indicating no statistically significant difference in HbA1c between the two groups after accounting for potential confounding variables.

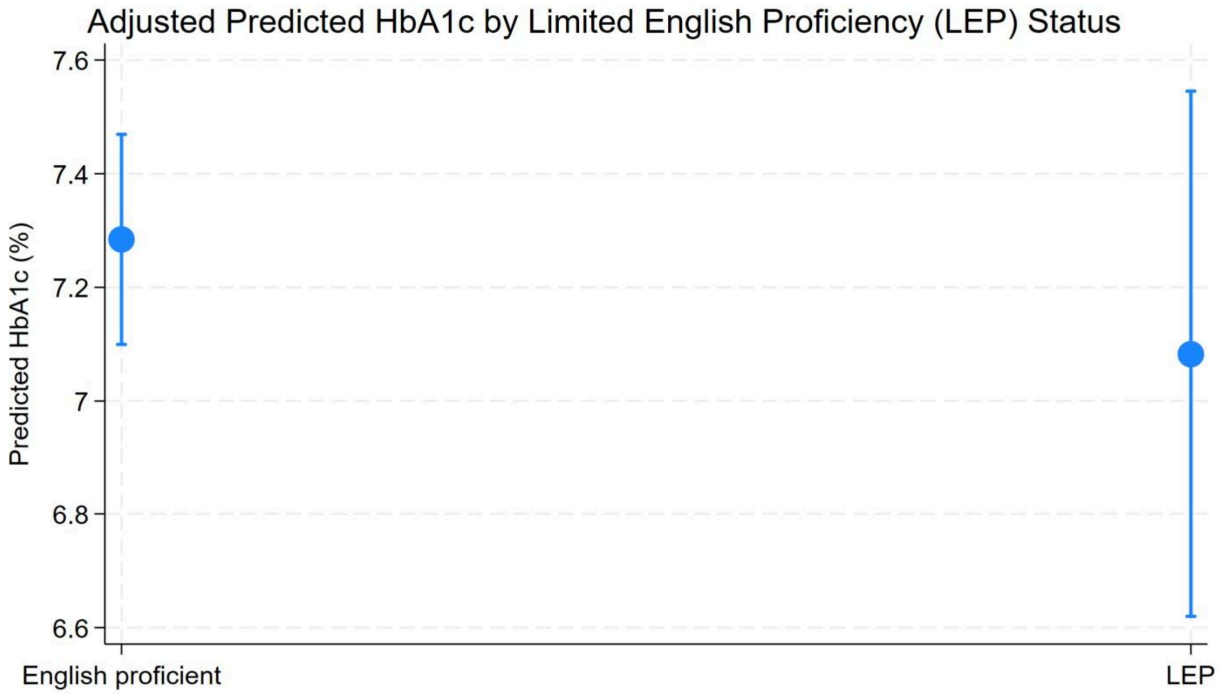
Similarly, LEP was not significantly associated with the time since participants last saw a diabetes nurse educator, dietitian, or nutritionist in either the crude model ($\beta = 0.15$, 95% CI: −0.33 to 0.63, $p = 0.512$) or the adjusted model ($\beta = 0.37$, 95% CI: −0.49 to 1.24, $p = 0.369$).

Table 2. Survey-Weighted Regression Analyses Examining the Association Between Limited English Proficiency and Study Outcomes Among U.S. Adults With Diagnosed Diabetes, NHANES 2017–2018

Outcome	Model	Regression Coefficient (β)	95% CI	P-value
HbA1c (%)	Crude	0.54	0.13-0.95	0.013
	Adjusted	-0.20	-0.62-0.22	0.323
Time Since Last Diabetes Education Visit (DIQ230)	Crude	0.15	-0.33-0.63	0.512
	Adjusted	0.37	-0.49-1.24	0.369

Note. *HbA1c regression analyses included 698 participants, and DIQ230 regression analyses included 716 participants after complete-case analysis. Adjusted models controlled for age, sex, race/ethnicity, education level, family income-to-poverty ratio, and health insurance status. For clarity, only the regression coefficients for the primary exposure variable (limited English proficiency) are presented. β = regression coefficient; CI = confidence interval.*

Figure 1. Adjusted predicted HbA1c by limited English proficiency status among U.S. adults aged ≥20 years with diagnosed diabetes, NHANES 2017–2018.

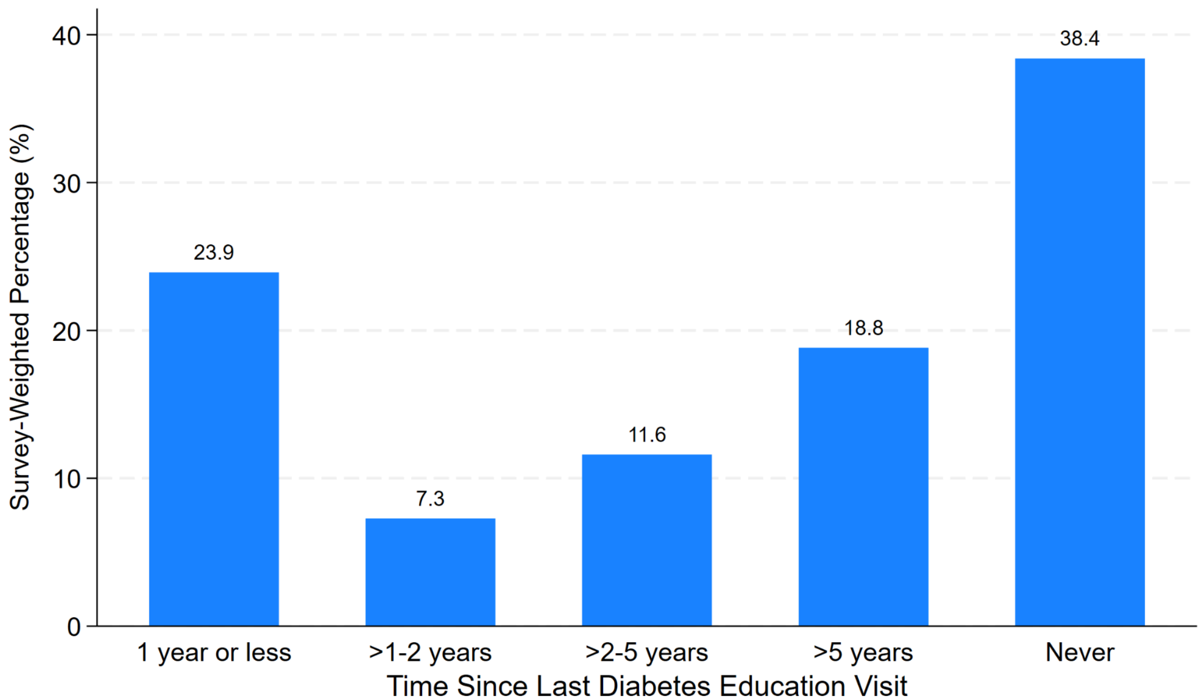


Note. Figure 1 illustrates the adjusted predicted HbA1c values by limited English proficiency (LEP) status. After adjustment for age, sex, race/ethnicity, education level, family income-to-poverty ratio, and health insurance status, the predicted mean HbA1c was 7.28% among English-proficient adults and 7.08% among adults with LEP. The substantial overlap in the 95% confidence intervals is consistent with the adjusted regression analysis, which found no statistically significant association between LEP and HbA1c after accounting for demographic and socioeconomic characteristics ($\beta = -0.20$; 95% CI, −0.62 to 0.22; $P = 0.323$).

Diabetes Education Service Utilization

The survey-weighted distribution of the time since participants last saw a diabetes nurse educator, dietitian, or nutritionist is presented in Figure 2. The largest proportion of adults diagnosed with diabetes (38.4%) reported that they had never seen a diabetes educator, dietitian, or nutritionist. Approximately one-quarter reported having seen one within the previous year, while smaller proportions reported their most recent visit occurred more than one year previously.

Figure 2. Survey-Weighted Distribution of Time Since Adults with Diagnosed Type 2 Diabetes Last Received Diabetes Education Services from a Diabetes Nurse Educator, Dietitian, or Nutritionist (NHANES 2017–2018).



***Note**: Percentages represent survey-weighted estimates among adults aged 20 years and older with diagnosed diabetes who had valid responses to the diabetes utilization question DIQ230 (n = 825).*

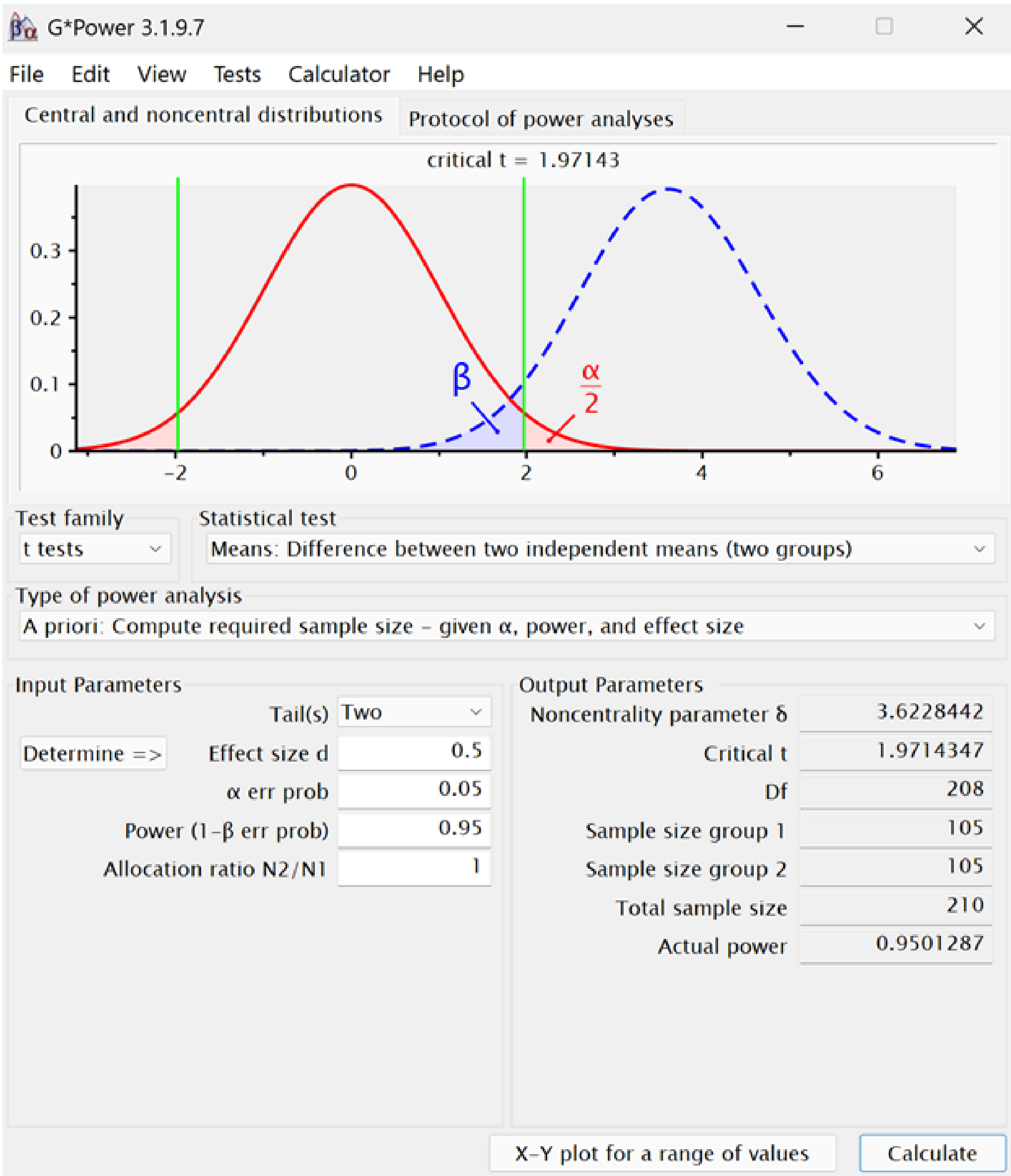
These findings describe the overall utilization of diabetes education services among adults with diagnosed diabetes in the study population.

Summary of Findings

Overall, limited English proficiency was significantly associated with higher HbA1c levels in the crude analysis but not after adjustment for demographic and socioeconomic characteristics. No statistically significant association was observed between limited English proficiency and the time since participants last saw a diabetes nurse educator, dietitian, or nutritionist in either the crude or adjusted analyses. After adjustment for potential confounding variables, the findings did not support an independent association between limited English proficiency and either HbA1c or diabetes education service utilization among U.S. adults with diagnosed diabetes.

Appendix A - G*Power A Priori Sample Size Analysis

Figure A1: GPower output showing the a priori sample size calculation for an independent-samples t-test (α = 0.05, power = 0.95, Cohen's d = 0.50)



REFERENCES

1. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. Questionnaires, datasets, and related documentation. National Health and Nutrition Examination Survey (NHANES). NHANES 2017-2018. First Published February 2020. Accessed June 20, 2026. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017.
2. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). 2017-2018 Data Documentation, Codebook, and Frequencies. Demographic Variables and Sample Weights (DEMO_J). Data File: DEMO_J.xpt. First Published February 2020. Accessed June 20, 2026. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/DEMO_J.htm
3. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). 2017-2018 Data Documentation, Codebook, and Frequencies. Diabetes (DIQ_J). Data File: DIQ_J.xpt. First Published February 2020. Accessed June 20, 2026. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/DIQ_J.htm
4. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). 2017-2018 Data Documentation, Codebook, and Frequencies. Glycohemoglobin (GHB_J). Data File: GHB_J.xpt. First Published: February 2020. Accessed June 20, 2026. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/GHB_J.htm
5. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). 2017-2018 Data Documentation, Codebook, and Frequencies. Health Insurance (HIQ_J). Data File: HIQ_J.xpt. First Published February 2020. Accessed June 20, 2026. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2017/DataFiles/HIQ_J.htm
6. Kang H. Sample size determination and power analysis using the G*Power software. *J Educ Eval Health Prof.* 2021;18:17. doi:10.3352/jeehp.2021.18.17
7. UCLA Advanced Research Computing. Statistical Methods and Data Analytics. Advanced Research Computing. Power analysis for two-group independent sample t-test.

G*Power Analysis Examples.
https://stats.oarc.ucla.edu/other/gpower/power-analysis-for-two-group-independent-sample-t-test/

8. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191. doi:10.3758/bf03193146
9. Jacobsen KH. Introduction to Health Research Methods. 3rd ed. *Jones & Bartlett Learning*; 2021.
10. Drummond KE, Murphy-Reyes A. Nutrition research: Concepts and applications. *Jones & Bartlett Learning;* 2018.
11. StataCorp. Stata Statistical Software: Release 19. College Station, TX: *StataCorp* LLC; 2025
12. StataCorp LLC. How to download, import, and prepare data from the NHANES website for analysis in Stata®. Published Jul 16, 2013. Accessed June 20, 2026. https://youtu.be/lRTl8GKsZTE?si=hmFcDe7cGrJjKi8i
13. StataCorp LLC. How to download, import, and merge multiple datasets from the NHANES website using Stata®. Published Jul 16, 2013. Accessed June 20, 2026. https://youtu.be/CUFr3CDM-4g?si=b3Wdf_bwehWTKQud
14. StataCorp LLC. Introduction to Stata basics. Accessed June 20, 2026. https://www.stata.com/links/stata-basics/
15. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). NHANES Tutorials. Accessed June 20, 2026. https://wwwn.cdc.gov/nchs/nhanes/tutorials/default.aspx
16. Dey D, Haque MS, Islam MM, et al. The proper application of logistic regression model in complex survey data: a systematic review. *BMC Med Res Methodol*. 2025;25(1):15. Published 2025 Jan 22. doi:10.1186/s12874-024-02454-54.
17. Mitani AA, Espin-Garcia O, Fernández D, Landsman V. Applying survey weights to ordinal regression models for improved inference in outcome-dependent samples with ordinal outcomes. *Stat Methods Med Res.* 2024;33(11-12):2007-2026. doi:10.1177/09622802241282091

18. Pew Research Center. Decoded. A short intro to linear regression analysis using survey data. https://www.pewresearch.org/decoded/2019/01/15/a-short-intro-to-linear-regression-analysis-using-survey-data/ . Published January 15, 2019. Accessed June 28, 2026.
19. Johnson CL, Paulose-Ram R, Ogden CL, et al. National health and nutrition examination survey: analytic guidelines, 1999-2010. *Vital Health Stat 2*. 2013;(161):1-24.

AI Disclosure

Google Docs was used for this assignment. The AI/auto-correct feature in Google Docs assisted with grammar editing. All research content, analysis, and final writing were completed and verified by the author.