Influences of Sugar and Vitamin D on Oral Health in US Population with Increased Proportion of Underrepresented Minority Groups for Ages 18-64

Hsin Ying (Angel) Tseng[1], Yuxuan Jiang,[1], Hanbei Xiong[1]
[1]University of California, Los Angeles

I. INTRODUCTION

Eating, smiling, and talking. Whether it allows us to perform vital physical functions or to socialize with others, oral health is closely connected to our everyday lives. Important as it is, oral health requires careful attention when it comes to ensuring adequate teeth and gum conditions. Out of many determinants, diet is one of the major factors in influencing oral health, as it prevents or catalyzes the development of oral diseases.

Prior research indicates that high sugar consumption and low vitamin D intake lead to a higher risk of tooth decay and related gum diseases (Moynihan, 2005). Vitamin D is a key factor in tooth and bone mineralization that prohibits tooth decay (Stein & Tipton, 2011). In particular, vitamin D2, sourced from plants, and vitamin D3, sourced from animals, both raise vitamin D levels in blood and are responsible for bone health (Alayed Albarri et al., 2022). Vitamin D2 and D3 would then be transformed by the liver into metabolites of 25-hydroxyvitamin D3, epi-25-hydroxyvitamin D3, and 25-hydroxyvitamin D2 for usage by the body (Holick, 2009). Since numerous oral diseases are not curable and need life-long management, it is essential for researchers to investigate the determinants of oral health to plan for early prevention of oral diseases (Zucoloto et al, 2016).

In this study, we are interested in exploring the relationships between diet involving sugar consumption and the three metabolites of vitamin D intake as predictors of the outcome measurement of oral condition. Different from previous research, we focus our target population on people of historically-underrepresented groups in STEM research ages 18-64. For this study, we have two hypotheses: high sugar consumption is positively associated with worse oral health and each of the three kinds of vitamin D intake is negatively associated with worse oral health. Investigation will involve using multiple logistic regression.

II. METHODS & MATERIALS

A. Data Set

The data used for this study is obtained from the National Health and Nutrition Examination Survey (NHANES) of the years 2015 and 2016 (https://www.cdc.gov/nchs/nhanes/index.htm). The participants are individuals of all ages in the United States. The original data from the NHANES website has 8,858 observations and over a hundred variables. Data was collected through surveys, health examinations, and laboratory examinations. Since our target population is underrepresented groups and low-income people, the sample was oversampled for Hispanic persons, specifically Mexican Americans, non-Hispanic black persons, non-Hispanic Asian persons, and those who lived below the Department of Health and Human Services poverty guidelines.

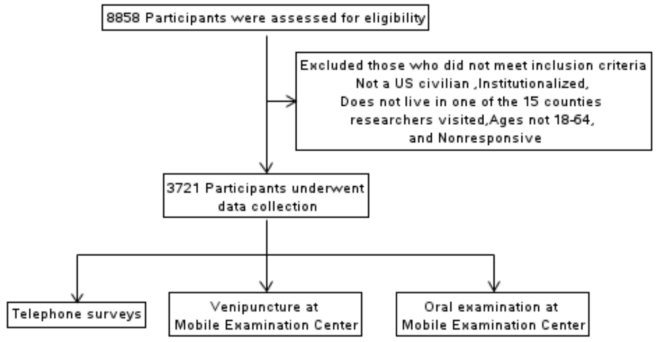

Figure 1: Sample Enrollment and Collection of Data

B. Study Population

In this study, the inclusion criteria to be in our sample is to be a civilian of the United States, noninstitutionalized, between ages 18-64, and lived in one of the 15 counties data was collected for all variables. Participants from the sample underwent three stages of data collection, the process can be seen in Figure 1. The sugar consumption data was calculated based on the amount of foods the participants consumed on two random days on a survey. The three vitamin D metabolites were

measured using blood samples from venipuncture of participants. Dental examinations by licensed dentists in at least one state in the United States were then done to evaluate oral health conditions at the mobile examination center. We then excluded participants with missing values in sugar consumption of both days, any of the three vitamin D intakes, and oral health conditions. Validation of data curation was done, where the corresponding codes were verified to produce the same results.

After data cleaning, we have 3,721 participants and nine variables. The demographic information could be seen in Table 1. For categorical variables, we represent the number of people and proportion of each group under the same oral health condition. For continuous variables, we used median and interquartile ranges to show the difference in subgroups of oral health condition since histograms of these variables showed strong skewness. Within 3,721 participants, there are 2,311 (62.1%) participants with oral health problems, and 1,140 (30.6%) participants who do not have oral health problems. We notice there are more male (53.33%) in the oral health problems group than in the group without oral health problems (45.66%). Compared with the group that do not have oral health problems, we notice there are more Mexican American (↑4.96%), significantly more non-hispanic black people (↑9.62%), less other hispanic (↓2.33%), and less non-hispanic white (↓3.13%), less non-hispanic Asian (↓8.61%), and less other racial group (↓0.51%) in the population with oral health problems. At last, there is a higher sugar intake and a lower or equal amount of vitamin D intake for people with oral health problems, aligning with our hypothesis.

Table 1: Baseline Characteristics of Data

	Having Oral Health Problem	
	No (n=2311)	Yes (n=1410)
Male (%)	1053 (45.56)	752 (53.33)
Female (%)	1258 (54.44)	658 (46.67)
Mexican American (%)	405 (17.52)	317 (22.48)
Other Hispanic (%)	329 (14.24)	168 (11.91)
Non-Hispanic White (%)	705 (30.51)	386 (27.38)
Non-Hispanic Black (%)	435 (18.82)	401 (28.44)
Non-Hispanic Asian (%)	340 (14.71)	86 (6.10)
Other Race - Including Multi-Racial(%)	97 (4.20)	52 (3.69)
Average Sugar consumption	88.3 (57.9–129.1)	95.8 (59.9–142.6)
25-hydroxyvitamin D3	56.2 (40.2–73.4)	50.4 (36.0–67.2)
epi-25-hydroxyvitamin D3	3.52 (2.4–5.0)	3.2 (2.2–4.5)
25-hydroxyvitamin D2	1.5 (1.5–1.5)	1.5 (1.5–1.5)

Categorical parameters are accompanied by % sign. Each entry represents the frequency and proportion. Continuous parameters do not have % sign, and we included median and IQR in the bracket.

C. Statistical Methods

For multiple predictor variables and one categorical outcome variable, we conducted multiple logistic regression analysis to determine the association among each predictive variable and oral health condition. For each independent variable in the multiple logistic regression model, average sugar consumption, density of 25-hydroxyvitamin D2, density of 25-hydroxyvitamin D3, and density of epimer of 25-hydroxyvitamin D3 in human serum, we carried hypothesis testing with Wald statistics test to evaluate the significance of associations. Validation of the multiple logistics regression procedures was done with separate code written. All statistical methods and procedures performed used the software SAS 9.4.

III. RESULTS

In order to perform multiple logistic regression analysis, we examined its three assumptions. The observations are independent since there is no duplicate of individual information. We examined the variance inflation factor (VIF) for each variable and they are much lower than 10, so we assume there is no multicollinearity based on Appendix Table 1. We also examined the linearity to logit of predictors and outcome by Box-Tidwell test in Appendix Table 2. We do not have enough information to reject the statement that there is no linear relationship between oral health and logit of vitamin D3, and between oral health and logit of epimer of vitamin D3 with threshold α = 0.05. We can conclude there is a linear relationship between oral health and logit of Sugar consumption, and between oral health and logit of Vitamin D2. However, we still included the variables which failed linearity check to run multiple logistic regression since we assume there should be a significant association based on past works.

The odd ratio estimate, 95% confidence interval, standard error, and *p*-value for each predictor variable can be seen in Table 2. If we kept one parameter changeable and the rest of parameters fixed, the odds of having oral health condition is expected to change by coefficient 1.002 for one unit increase in average sugar consumption, by coefficient 0.985 with one unit increase in 25-hydroxyvitamin D2, by coefficient 0.986 with one unit increase in 25-hydroxyvitamin D3, by coefficient 1.033 with one unit increase in epimer of

25-hydroxyvitamin D3. Taking α-level as 0.05, all predictors are statistically significant except for epimer of 25-hydroxyvitamin D3. All predictors showed low estimated standard error derived by Fisher information, indicating low uncertainty of population parameter.

Table 2: Logistic Regression Results

Parameter	Odds Ratio	Standard Error	95% Confidence Interval	Pr > ChiSq
Sugar Consumption	1.002	0.000528	1.001-1.003	<.0001
Vitamin D2	0.985	0.00410	0.977-0.992	0.0001
Vitamin D3	0.986	0.00230	0.982-0.991	<.0001
Epi-Vitamin D3	1.033	0.0216	0.990-1.078	0.1336

Each row includes a variable which stands for average sugar consumption, 25-hydroxyvitamin D2, 25-hydroxyvitamin D3, and epimer of 25-hydroxyvitamin D3 density respectively. Odds Ratios stands for Odds Ratio estimates. We used Wald statistics to derive the p value and confidence interval.

Figure 2 shows the ROC curve for our logistic regression model. The area under the curve is 0.5892 which is obviously above the baseline model which contains area 0.5. It validates the result from our Table 1 that there is an association between our predictors and outcome.

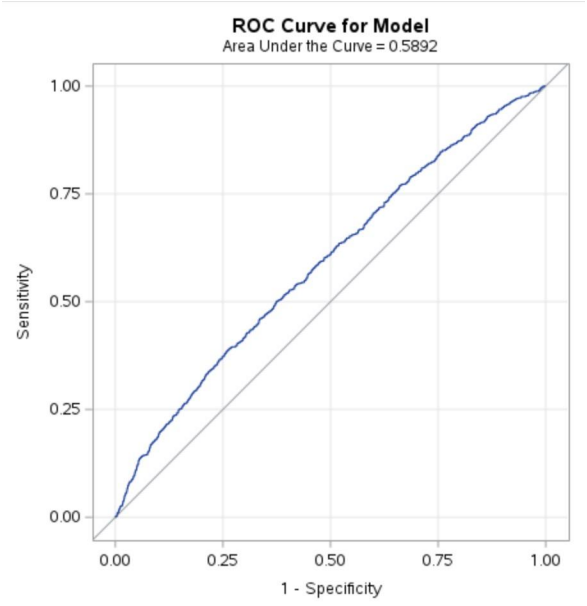

Figure 2: ROC Curve for Logistic Regression Model

IV. CONCLUSIONS

In conclusion, with means to investigate whether sugar consumption and vitamin D intake are significant determinants of oral health for underrepresented minority groups between ages 18 and 64, this study implemented a multiple logistic regression model to test for associations between the predictors and the outcome of oral health condition. Results show that the the model with variables average sugar consumption, the density of 25-hydroxyvitamin D2, the density of 25-hydroxyvitamin D3, and the density of epimer of 25-hydroxyvitamin D3 are better than the baseline model in predicting oral health condition.

With the support of hypothesis testing results, we accept our hypothesis which conclude that there is positive association between worse oral health and sugar consumption, negative association between worse oral health and 25-hydroxyvitamin D2, and negative association between worse oral health and 25-hydroxyvitamin D3. Although the density of epimer of 25-hydroxyvitamin D3 showed a positive association with the worse oral health condition in figure 2c, we do not have enough evidence to support this finding. Hence, we reject our hypothesis that epimer of vitamin D3 intake is negatively associated with worse oral health. Specifically, higher sugar consumption is related to worse oral health condition and higher vitamin D2 and vitamin D3 intake correlates to better oral health condition.

V. DISCUSSION

Oral health accounts for a crucial part of our physical and mental health, where without proper care, consequences like difficulties in eating can eventually damage our overall physical health and the inability to speak comfortably weakens one's self-confidence. Hence, this study examines the main factors notable for impacting oral health: sugar consumption and vitamin D intake. Understanding the importance of reducing health inequities in public health, we focus our target population on underrepresented minority groups in the United States between ages 18 to 46.

Our findings are consistent with most prior research findings, where sugar consumption is negatively associated with oral health and vitamin D2 and D3 are positively associated with oral health. Such findings also allowed us to confirm our hypotheses. However, we did not obtain a significant association between epimer vitamin D3, even though it is expected of vitamin D metabolites to be related to oral health. This finding contrasts

with findings of some prior research (Holick, 2009). Such an inconsistency could be a result of the limitations of our study.

Some limitations of this study include the inability to use our findings to comment on any interaction effects between the predictors, if any, as we only tested for associations between each individual predictors with the outcome of oral health condition. Additionally, data of sugar consumption was calculated from the foods the participants reported in the survey, which may be susceptible to recall bias, where the participants might not remember all the foods they had that day. We also recognize sugar consumption and vitamin D intake do not encompass all the factors influencing oral health, where access to teeth cleanings and healthcare may be potential confounders. We thus limit our study in the data available to us. Moreover, since variable 25-hydroxyvitamin D3 and epimer of 25-hydroxyvitamin D3 failed the linearity assumption in the Box-Tidwell test, there is a possible nonlinear relationship between these variables and our outcome, which can explain poorness of our model performance.

Future research extended from this study's findings may be done on subdividing the population into different ethnic groups and examining the correlation between oral health condition and racial ethnic background, since we observe that non-hispanic black and Mexican American constitute a high proportion in the population with oral health problems. Adopting a non-linear model could also be a good research direction to explore the relationship between oral health and vitamin D3, and between oral health and epimer of vitamin D3 due to our failed linearity assumption before logistic regression.

ACKNOWLEDGEMENTS

We would like to acknowledge the Centers for Disease Control and Prevention (CDC) for providing the data set of the National Health and Nutrition Examination Survey (NHANES) in the cycle of 2015 to 2016. Additional acknowledgments would be dedicated to Xiufang Zhang and Xiang Wang from FMD K&L Inc. that provided the code template to generate a flowchart in SAS.

REFERENCES

1. Alayed Albarri, E. M., Sameer Alnuaimi, A., & Abdelghani, D., "Effectiveness of vitamin D2 compared with vitamin D3 replacement therapy in a primary healthcare setting: a retrospective cohort study," *Qatar medical journal*, Vol. 3, No. 29, 2022.
2. Holick M. F., "Vitamin D status: measurement, interpretation, and clinical application," *Annals of Epidemiology*, Vol. 19, No. 2, pp. 73-78, 2009.
3. Moynihan, P., "The interrelationship between diet and oral health," *Proceedings of the Nutrition Society,* Vol. 64, No. 4, pp. 571-580, 2005.
4. Stein, S. H., Tipton, D. A., "Vitamin D and its impact on oral health—An update," *Journal of the Tennessee Dental Association*, Vol. 91, No. 2, pp. 30, 2011.
5. Zhang, X. and Wang, X., "Flowcharts Drawn with SAS/GTL," Paper Paper DV-043, 2019.
6. Zucoloto, M. L., Maroco, J., & Campos, J. A., "Impact of oral health on health-related quality of life: a cross-sectional study," *BMC Oral Health*, Vol. 16, No. 55, 2016.

APPENDIX

I. Source Data File

The source of the public data set is from the National Health and Nutrition Examination Survey (NHANES) of the years 2015 and 2016 (https://www.cdc.gov/nchs/nhanes/index.htm). We merged five data sets: the demographic data set, the oral health examination data set, two diet data sets for calculating sugar consumption, and the laboratory data set for vitamin D intake. Data dictionaries for all five data sets are linked onto a document titled “NHANES Data Dictionary” and submitted to CCLE along with the source data files.

II. SAS data set

The SAS data set used for our study can be obtained by running our SAS code provided in Appendix.

Proc Contents of the data set for our study:

The CONTENTS Procedure

Data Set Name	PROJECT.CLEANEDDATA	Observations	3721
Member Type	DATA	Variables	9
Engine	V9	Indexes	0
Created	10/31/2023 14:48:37	Observation Length	72
Last Modified	10/31/2023 14:48:37	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Label
8	AVGSUGAR	Num	8		Sugar intake average on both days
5	LBXVD2MS	Num	8		25-hydroxyvitamin D2
6	LBXVD3MS	Num	8		25-hydroxyvitamin D3
7	LBXVE3MS	Num	8		epi-25-hydroxyvitamin D3
9	ORALHEALTH	Num	3	OHEALTHFORMAT.	Generalized oral health condition
2	RIAGENDR	Num	8	GENDERFORMAT.	Gender
3	RIDAGEYR	Num	8	AGEFORMAT.	Age in years
4	RIDRETH3	Num	8	RACEFORMAT.	Race
1	SEQN	Num	8		ID

III. Additional Appendix Tables

Appendix Table 1: Examining Variance Inflation and Tolerance for Multicollinearity

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	Intercept	1	0.48442	0.02441	19.84	<.0001	.	0
AVGSUGAR	Sugar intake average on both days	1	0.00049347	0.00012264	4.02	<.0001	0.99567	1.00435
LBXVD2MS	25-hydroxyvitamin D2	1	-0.00318	0.00077744	-4.09	<.0001	0.95576	1.04629
LBXVD3MS	25-hydroxyvitamin D3	1	-0.00308	0.00051048	-6.04	<.0001	0.37928	2.63654
LBXVE3MS	epi-25-hydroxyvitamin D3	1	0.00743	0.00487	1.53	0.1272	0.38882	2.57191

Appendix Table 2: Box-Tidwell test for Linearity to log odds of predictors

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.5752	0.9395	7.5136	0.0061
AVGSUGAR	1	0.00471	0.00117	16.2497	<.0001
LBXVD2MS	1	0.00316	0.00662	0.2278	0.6332
LBXVD3MS	1	-0.00551	0.00566	0.9491	0.3299
LBXVE3MS	1	-0.00564	0.0382	0.0219	0.8824
ln_AVGSUGAR	1	-0.2846	0.1112	6.5536	0.0105
ln_LBXVD2MS	1	-0.3524	0.0946	13.8880	0.0002
ln_LBXVD3MS	1	-0.5056	0.3024	2.7958	0.0945
ln_LBXVE3MS	1	0.1887	0.1856	1.0337	0.3093

IV. SAS Code

```
/* import the oral health data set from the website */
filename oral
"/home/u63620464/BIOSTAS203A/MidtermProject/OHXREF_I.XPT";

proc http
url="https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/OHXREF_I.XPT"
          out=oral;
run;

libname oral xport
"/home/u63620464/BIOSTAS203A/MidtermProject/OHXREF_I.XPT";

/* import the dietary data sets from the website */
filename diet1
"/home/u63620464/BIOSTAS203A/MidtermProject/DR1TOT_I.XPT";

proc http
url="https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR1TOT_I.XPT"
          out=diet1;
run;

libname diet1 xport
"/home/u63620464/BIOSTAS203A/MidtermProject/DR1TOT_I.XPT";

filename diet2
"/home/u63620464/BIOSTAS203A/MidtermProject/DR2TOT_I.XPT";
proc http
url="https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR2TOT_I.XPT"
          out=diet2;
run;

libname diet2 xport
"/home/u63620464/BIOSTAS203A/MidtermProject/DR2TOT_I.XPT";

/*import the demographic data set from the website */
filename demo
"/home/u63620464/BIOSTAS203A/MidtermProject/DEMO_I.XPT";

proc http
url="https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DEMO_I.XPT"
out=demo;
```

```
run;

libname demo xport
"/home/u63620464/BIOSTAS203A/MidtermProject/DEMO_I.XPT";

/*import the vitaminD data set from the website */
filename vd
"/home/u63620464/BIOSTAS203A/MidtermProject/VID_I.XPT";

proc http
url="https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/VID_I.XPT" out=vd;
run;

libname vd xport
"/home/u63620464/BIOSTAS203A/MidtermProject/VID_I.XPT";

/* clean the oral health data set accoring to the meaning provided
on the website*/
libname project "/home/u63620464/BIOSTAS203A/MidtermProject";

proc format;
     value oralformat 1="See a dentist immediately"
          2="See a dentist within the next 2 weeks"
          3="See a dentist at your earliest convenience"
          4="Continue your regular routine care";
run;

data project.cleanedoral;
     set oral.OHXREF_I;
     keep seqn OHAREC;
     where OHAREC is not missing;
     format OHAREC oralformat.;
     label seqn="ID" OHAREC="Oral health condition";
run;

proc sort data=project.cleanedoral;
     by seqn;
run;

/* clean the diet data sets */
data project.cleandiet1;
     set diet1.DR1TOT_I;
```

```
     keep seqn DR1TSUGR;
     label seqn="ID" DR1TSUGR="Sugar intake on day 1";
run;

proc sort data=project.cleandiet1;
     by seqn;
run;

data project.cleandiet2;
     set diet2.DR2TOT_I;
     keep seqn DR2TSUGR;
     label seqn="ID" DR2TSUGR="Sugar intake on day 2";
run;

proc sort data=project.cleandiet2;
     by seqn;
run;

/*clean the demo data sets*/
data project.cleandemo;
     set demo.DEMO_I;
     keep seqn RIDAGEYR RIAGENDR RIDRETH3;
     label seqn="ID" RIDAGEYR="Age in years" RIAGENDR="Gender"
RIDRETH3="Race";
run;

proc sort data=project.cleandemo;
     by seqn;
run;

/*clean the vitamin d data sets*/
data project.cleandvd;
     set vd.VID_I;
     keep seqn LBXVD2MS LBXVD3MS LBXVE3MS;
     label seqn="ID" LBXVD2MS="25-hydroxyvitamin D2"
          LBXVD3MS="25-hydroxyvitamin D3"
LBXVE3MS="epi-25-hydroxyvitamin D3";
run;

proc sort data=project.cleandvd;
     by seqn;
run;
```

```
/* merge 5 data sets into 1 and calculaate new columns*/
data project.mergeddata;
     merge project.cleanedoral project.cleandiet1
project.cleandiet2
          project.cleandemo project.cleandvd;
     by seqn;
run;

proc format;
     value ohealthformat 0="Continueroutine care" 1="Need medical
attetion";
run;

proc format;
     value ageformat 0-1="Infants" 1-12="Children"
13-17="Adolescent" 18-64="Adult"
          65-high="Older Adult";
run;

proc format;
     value raceformat 1="Mexican American" 2="Other Hispanic"
          3="Non-Hispanic White" 4="Non-Hispanic Black"
6="Non-Hispanic Asian"
          7="Other Race - Including Multi-Racial";
run;

proc format;
     value genderformat 1="Male" 2="Female";
run;

data project.mergeddata;
     set project.mergeddata;
     AVGSUGAR=mean(DR1TSUGR, DR2TSUGR);
     length ORALHEALTH 3;
     if OHAREC=1 or OHAREC=2 or OHAREC=3 then
          ORALHEALTH=1;
     else if OHAREC=4 then
          ORALHEALTH=0;
     format ORALHEALTH ohealthformat. RIDAGEYR ageformat. RIDRETH3
          raceformat. RIAGENDR genderformat.;
     label seqn="ID" AVGSUGAR="Sugar intake average on both days"
          OHAREC="Oral health condition"
```

```
          ORALHEALTH="Generalized oral health condition ";
run;

/*clean missing values, and add formats */
data project.cleanedData;
     set project.mergeddata(drop=DR1TSUGR DR2TSUGR OHAREC);
     where RIDAGEYR >=18 and RIDAGEYR <=64 and LBXVD2MS is not
missing and LBXVD3MS
          is not missing and LBXVE3MS is not missing and
ORALHEALTH is not missing and
          AVGSUGAR is not missing;
     format ORALHEALTH ohealthformat. RIDAGEYR ageformat. RIDRETH3
          raceformat. RIAGENDR genderformat.;
run;

proc print data=project.cleanedData(obs=10) label;
run;

proc contents data=project.cleanedData;
run;

/* Table 1: Descriptive statistics & data exploration */
proc sort data=project.cleanedData;
     by RIDRETH3;
run;

proc freq data = project.cleanedData;
     tables RIDRETH3*ORALHEALTH /norow nocol;
     by RIDRETH3;
run;

proc sgplot data= project.cleanedData;
histogram AVGSUGAR;
run;
Proc sgplot data= project.cleanedData;
histogram LBXVD2MS;
where ORALHEALTH=1;
run;
Proc sgplot data= project.cleanedData;
histogram LBXVD2MS;
where ORALHEALTH=0;
run;
```

```
proc sgplot data= project.cleanedData;
histogram LBXVD3MS;
run;

proc sgplot data= project.cleanedData;
histogram LBXVE3MS;
run;

proc sgplot data= project.cleanedData;
histogram ORALHEALTH;
run;

PROC MEANS DATA = project.cleanedData N MEDIAN P25 P75 MIN MAX
maxdec=2;
     var RIDAGEYR RIAGENDR RIDRETH3;
     class ORALHEALTH;
RUN;

proc summary data=project.cleanedData min max PRINT NOLABELS;
class ORALHEALTH; /* Binary variable with "yes" and "no" values */
var AVGSUGAR LBXVD2MS LBXVD3MS LBXVE3MS; /* List of continuous
variables */
run;

/*logistic regression */
* check redundancy for logistic regression;
proc sort data=project.cleanedData out=project.cleanedData
nodupkey
          dupout=project.duplicated;
     by seqn;
run;

*check multicollinearity;
proc reg data=project.cleanedData;
model ORALHEALTH = AVGSUGAR LBXVD2MS LBXVD3MS LBXVE3MS / vif tol;
run;

*check Linearity;
data CheckData;
set project.cleanedData;
ln_AVGSUGAR =log(AVGSUGAR);
ln_LBXVD2MS =log(LBXVD2MS);
```

```
ln_LBXVD3MS =log(LBXVD3MS);
ln_LBXVE3MS =log(LBXVE3MS);
run;
proc logistic data=CheckData;
model ORALHEALTH(event="Need medical attention") = AVGSUGAR
LBXVD2MS LBXVD3MS LBXVE3MS
ln_AVGSUGAR ln_LBXVD2MS ln_LBXVD3MS ln_LBXVE3MS;
run;

* Figure 2: logistic regression plots;
proc logistic data=project.cleanedData plots =all;
model ORALHEALTH(event="Need medical attention") = AVGSUGAR
LBXVD2MS LBXVD3MS LBXVE3MS;
output out=results_dataset predicted=PredictedProb;
run;

* Table 2: logistic regression Analysis;
proc logistic data=project.cleanedData;
model ORALHEALTH = AVGSUGAR LBXVD2MS LBXVD3MS LBXVE3MS ;
run;

/* flow chart sas code*/
/* Figure 1: Enrollment and data collection */
data text;
     length text $200;
     text="8858 Participants were assessed for eligibility";
     x_=25; y_=35;output;
     text="Excluded those who did not meet inclusion criteria*Not
a US civilian
     ,Institutionalized,*Does not live in one of the 15 counties
*researchers visited,
     Ages not 18-64,* and Nonresponsive";
     x_=46; y_=28;output;
     text="3721 Participants underwent* data collection";
     x_=25; y_=20;output;
     text="Venipuncture at *Mobile Examination Center";
     x_=25; y_=10;output;
     text="Telephone surveys";
     x_=6; y_=10;output;
     text="Oral examination at *Mobile Examination Center";
     x_=48; y_=10;output;
run;
```

```
data arrow;
     length name1 name2 $200;
     name1="8858 Participants were assessed for eligibility";
     name2="3721 Participants underwent* data collection";
     m1=25;n1=33.5;m2=25;n2=22;output;
     name1="8858 Participants were assessed for eligibility";
     name2="Excluded those who did not meet inclusion criteria*Not
a US civilian,
     *Institutionalized*Does not live in one of the 15 counties*
researchers visited,
     *Ages not 18-64,* and Nonresponsive";
     m1=25;n1=28;m2=28;n2=28;output;
     name1="3721 Participants underwent* data collection";
     name2="Telephone surveys";
     m1=6;n1=15;m2=6;n2=11;output;
     name1="3721 Participants underwent* data collection";
     name2="Venipuncture at *Mobile Examination Center";
     m1=25;n1=18;m2=25;n2=12;output;
     name1="3721 Participants underwent* data collection";
     name2="Oral examination at *Mobile Examination Center";
     m1=48;n1=15;m2=48;n2=12;output;
run;
data treat;
set text arrow;
run;
proc template;
     define statgraph textplot;
          begingraph;
               layout overlay /yaxisopts=(linearopts=(viewmin=0
viewmax=45) display=none)

xaxisopts=(linearopts=(viewmin=0 viewmax=50) display=none)
                                   walldisplay=none;
                    textplot x=x_ y=y_ text=text / name="m"
                         position=center
                         splitpolicy=split
                         splitchar="*"
                         splitchardrop=true
                         vcenter=bbox
                         position=center
                         display=(fill outline)
                         fillattrs=(color=blue transparency=1)
```

```
                        textattrs=(weight=bold color=black);
                    vectorplot xorigin=m1 yorigin=n1 x=m2 y=n2 /
xaxis=x yaxis=y
                        arrowdirection=out arrowheadshape=open
                        lineattrs=(pattern=solid thickness=1px
color=black);
                    drawline x1=6 y1=15 x2=48 y2=15 / xaxis=x
yaxis=y drawspace=datavalue
                    lineattrs=(pattern=solid thickness=1px
color=black);
                endlayout;
            endgraph;
    end;
run;

ods listing close;
proc sgrender data=treat template=textplot;
run;

ods listing;

/* Validation: multiple logistic regression */
title "Oral Health Predictors - Multicollinearity Investigation of
VIF
and Tolerance";

proc reg data=project.cleanedData;
    model ORALHEALTH= AVGSUGAR LBXVD2MS LBXVD3MS LBXVE3MS / vif
tol collin;
run;

ods graphics on;
proc logistic plots=all;
model y=x;
run;

proc logistic data=project.cleanedData outest=project.betas
    covout alpha=0.05 plots=(oddsratio roc);
    model ORALHEALTH(event="Need medical attention") = AVGSUGAR
LBXVD2MS LBXVD3MS LBXVE3MS;
run;
```