Health Status & Emotional Support of Immigrants

Background

There has been a lot of research showing the influence of social relationships on patterns of health-related behaviors, which impact the mental and physical health of an individual. (1) A survey study conducted in various provinces of China demonstrated that having emotional support was significantly related to mental health, and community support with individual BMI, physical health indicator. (2)

The United States (US) is the top country for representing a diverse immigrant population. As of 2020, 13.7% of the US population is attributed to foreign-born individuals. (4) A study that examined how acculturation status and suicidal ideation is moderated by social support among Asian immigrants in the US demonstrated that social support in reducing the risk of suicidal ideation was gender-specific and favoring those who were assimilated. (3)

Introduction

Being a first-generation immigrant myself and witnessing the process of acculturation in my family, I wanted to examine the impact of social (emotional) support on the mental and physical health of immigrants. I used the US National Health and Nutrition Examination Survey (NHANES) data from 2007 to 2008, which was the last year social support module was utilized, to conduct this analysis.

Purpose

To examine the association between social (emotional) support and self-rated health status (physical and mental) among a nationally representative sample of adults (>=18 years) of immigrant status.

1. Compare the prevalence of social (emotional) support among immigrants: by demographic characteristics and time in the US.
2. Determine association between health and social (emotional) support.

Methodology

NHANES demographics data (demo_e), and two different modules from questionnaire data: current health status (hsq_e), and social support (ssq_e) from the 2007 to 2008 cycle were used. Firstly, the 3 data files were merged by SEQN, participant identifying variable. Next, responses of "Don't Know" and "Refused" from all variables were recoded to missing, and the rest of the responses of select variables were recoded as demonstrated in **Table 1**. The main exposure of interest is emotional (EMOSPRT). The main outcome variable is the rating of overall health (GENHLTH). The final list of variables relevant for project's data analysis as listed in **Table 2**.

Table 1: Overview of original and recoding of select variables.

Original Data File	Original Variable	Original Responses	Recoding	Renamed Variable	Recoded Responses
SSQ_E	SSQ011*	1=Yes 2=No 3=Don't Need	1 = 1 2, 3 = 0	SOCSPRT	1=Has emotional support 0=No emotional support
	~~SSQ021A**~~	~~10=Spouse~~	~~10 = 1~~	~~WHOSPRT~~	~~1=Spouse~~ ~~2=Children~~ ~~3=Siblings~~ ~~4=Parents~~ ~~5=Relatives~~ ~~6=Friends~~ ~~7=Others~~
	~~SSQ021B~~	~~11=Daughter~~	~~11= 2~~		
	~~SSQ021C~~	~~12=Son~~	~~12 = 2~~		
	~~SSQ021D~~	~~13=Sister/brother~~	~~13 = 3~~		
	~~SSQ021E~~	~~14=Parent~~	~~14 = 4~~		
	~~SSQ021F~~	~~15=Other relatives~~	~~15 = 5~~		
	~~SSQ021G~~	~~16=Neighbors~~	~~16 = 7~~		
	~~SSQ021H~~	~~17=Co-workers~~	~~17 = 7~~		

	~~SSQ021I~~	~~18=Church members~~	~~18 =7~~		
	~~SSQ021J~~	~~19=Club members~~	~~19 = 7~~		
	~~SSQ021K~~	~~20=Professionals~~	~~20 = 7~~		
	~~SSQ021L~~	~~21=Friends~~	~~21 = 6~~		
	~~SSQ021M~~	~~22=Other~~	~~22 = 7~~		
HSQ_E	HSD010*	1=Excellent 2=Very good 3=Good 4=Fair 5=Poor	1, 2 = 1 3, 4 = 2 5 = 3	GENHLTH	1=Good 2=Fair 3=Bad
	~~HSQ470**~~	~~0-30 (range of values)~~	~~0 – 10 = 1~~ ~~11 – 20 = 2~~ ~~21 – 30 = 3~~ ~~30 = 4~~	~~BADHLTH_P~~	~~1=Never~~ ~~2=Occasionally~~ ~~3=Often~~ ~~4=Always~~
	~~HSQ480**~~	~~0-30 (range of values)~~	~~0 – 10 = 1~~ ~~11 – 20 = 2~~ ~~21 – 30 = 3~~ ~~30 = 4~~	~~BADHLTH_M~~	~~1=Never~~ ~~2=Occasionally~~ ~~3=Often~~ ~~4=Always~~
	~~HSQ490**~~	~~0-30 (range of values)~~	~~0 – 10 = 1~~ ~~11 – 20 = 2~~ ~~21 – 30 = 3~~ ~~30 = 4~~	~~INACTIVE~~	~~1=Never~~ ~~2=Occasionally~~ ~~3=Often~~ ~~4=Always~~
DEMO_E	DMDCITZN*	1=Citizen by birth/naturalization 2=Not a citizen of the US	2 = 1 1 = 0	IMMIGRANT	1=Immigrant 0=Not Immigrant
	DMDBORN2*	1 = Born in US 50 states/Washington DC 2=Born in Mexico 4=Born in other Spanish speaking country 5=Born in other non-Spanish speaking country	2, 4, 5 = 1 1 = 0		
	DMDYRSUS**	1=<1 year 2=1 years, <5 years 3=5 years, <10 years 4=10 years, <15 years 5=15 years, <20 years	1, 2 = 1 3 = 2 4, 5 = 3 6, 7, 8, 9 = 4	TIMEUS	1=<5 years 2=5-9 years 3=10-19 years 4=20+ years

		6=20 years, <30 years 7=30 years, <40 years 8=40 years, <50 years 9=50 years or more			
	RIDAGEYR	0 to 79 (range of values) 80 >= 80 years of age	0-17 = 0 18-80 = 1	AGEGRP	1= Adults(>=18 years) 0 = Not Adults(<18years)
	DMDEDUC2*	1=<9th grade 2= 9-11th Grade 3=High School Grad/GED 4=Some College/AA degree 5=College Graduate/above	1, 2 = 1 3 = 2 4 = 3 5 = 4	EDUCLVL	1=<High School 2=HS Diploma/GED 3=Some College/AA Degree 4=College Grad/+
	DMDMARTL**	1=Married 2=Widowed 3=Divorced 4= Separated 5=Never married 6=Living with partner	1 = 3 2 = 4 3, 4 = 5 5 = 1 6 = 2	MRTLSTAT	1=Single 2=Live-in w/ Partner 3=Married 4=Widowed 5=Separated/Divorced

*Variables that had 7=Refused and 9=Don't know responses were collapsed as missing responses.
**Variables that had 77=Refused and 99=Don't know responses were collapsed as missing responses.

Table 2: Overview of variables used for data analysis.

Category of Variables	Variable Name	Variable Description
Complex Sample Design Variables	SEQN	Respondent Sequence Number
	WTINT2YR	Full Sample 2 Year Interview Weight
	SDMVSTRA	Masked Variance Pseudo-Stratum
	SDMVPSU	Masked Variance Pseudo-PSU
Domain Variables (sub-population group)	IMMIGRANT	Immigrant Status
	AGEGRP	Adult Age Group (>=18 years)
Exposure Variables	EMOSPRT	Has someone for emotional support
	~~WHOSPRT~~	~~Who gives most emotional support~~
Outcome Variables	GENHLTH	Rating of overall health
	~~BADHLTH_P, HSQ470~~	~~How often physical health was bad (category, numeric)~~
	~~BADHLTH_M, HSQ480~~	~~How often mental health was bad (category, numeric)~~
	~~INACTIVE, HSQ490~~	~~How often inactive due to physical/mental health (category, numeric)~~
Demographic Variables	RIAGENDR	Sex (gender)
	RIDRETH1	Race/Ethnicity Group

	EDUCLVL	Level of Education Completed
	MRTLSTAT	Marital Status
	TIMEUS	Years living in the US

Results

Table 3: Count of responses and missing values across variables of interest

The MEANS Procedure

Variable	Label	N	N Miss
SEQN	Respondent sequence number	10149	0
RIAGENDR	Gender	10149	0
RIDRETH1	Race/Ethnicity - Recode	10149	0
WTINT2YR	Full Sample 2 Year Interview Weight	10149	0
SDMVPSU	Masked Variance Pseudo-PSU	10149	0
SDMVSTRA	Masked Variance Pseudo-Stratum	10149	0
HSQ470	no. of days physical health was not good	6333	3816
HSQ480	no. of days mental health was not good	6336	3813
HSQ490	inactive days due to phys./mental hlth	6336	3813
AGEGRP	Adult Age Group (>=18 years)	10149	0
EDUCLVL	Education Level Completed	5928	4221
MRTLSTAT	Marital Status	5931	4218
IMMIGRANT	Immigrant Status	10149	0
TIMEUS	Duration of residence in US	1654	8495
EMOSPRT	Has someone for emotional support	10149	0
WHOSPRT	Who provides emotional support	3222	6927
GENHLTH	Rating of general health	6354	3795
BADHLTH_P	How often physical heath was bad	6333	3816
BADHLTH_M	How often mental health was bad	6336	3813
INACTIVE	How often inactive due to bad health	6336	3813

```
/* tally of obs & missing obs from final data set for analysis */
proc means data=nhanes.analysis n nmiss;
title "Count of Responses & Missing Values Across Variables of Interest";
run;
```

Table 4: Summary of WTINT2YR sampling weight variable.

There are 10149 observations in the analysis data set. Each row of data in NHANES dataset has a value for the sample weight. In this analysis dataset, each row of data represents 29,277 people in the population. The sum of weights, 297,136,095 is the estimated number of people in the population. For reference, the population of the US in the year 2007 was 302,743,399 and in 2008 was 305,694,910. (5)

The MEANS Procedure

Analysis Variable : WTINT2YR Full Sample 2 Year Interview Weight				
N	Minimum	Mean	Maximum	Sum
10149	2359.37	29277.38	186295.51	297136095

```
/* summary of survey design */
proc means data=nhanes.analysis n min mean max sum;
var WTINT2YR;
run;
```

Figure 1: Chi-Square Goodness-of-Fit Test for Dependent Variable: GENHLTH

The-one way table provides the frequency of the sample group and the weighted percent estimates of the population. It is estimated that 51.52% of the population reported feeling "fair" about their health in general. The null hypothesis for Goodness-of-fit test is equal proportions for all the levels of the dependent variable, general health. The Rao-Scott design-adjusted chi-square test for GENHLTH has a p-value of <.0001 indicating that the null hypothesis can be rejected.

Chi-Square Goodness-of-Fit Test: General Health

The SURVEYFREQ Procedure

Rating of general health					
GENHLTH	Frequency	Percent	Std Err of Percent	95% Confidence Limits for Percent	
Good	2413	45.6894	1.9412	41.5742	49.8046
Fair	3697	51.5214	1.6388	48.0474	54.9954
Bad	244	2.7892	0.3639	2.0178	3.5606
Frequency Missing = 3795					

Rao-Scott Chi-Square Test	
Pearson Chi-Square	2699.9843
Design Correction	5.7836
Rao-Scott Chi-Square	466.8353
DF	2
Pr > ChiSq	<.0001
F Value	233.4177
Num DF	2
Den DF	32
Pr > F	<.0001
Sample Size = 6354	

```
/* Descriptive Stats: Goodness-of-fit test of outcome variable */
proc surveyfreq data=nhanes.analysis nosummary;
strata SDMVSTRA;
cluster SDMVPSU;
weight WTINT2YR;
tables GENHLTH / nowt chisq nototal row cl;
format GENHLTH GENHLTHf. ;
title "Chi-Square Goodness-of-Fit Test: General Health";
run;
```

Figure 2: Test of Independent Association between Independent and Dependent Variables

The two-way table provides the frequency of the sample group and the weighted percent estimates of the population. Among individuals who have emotional support, 53.39% reported to have fair health while 49.63% of the counter-part group (no emotional support) reported to have fair health. The null hypothesis that there is no association between emotional support and general health rating can be rejected because the p-value, 0.0001, is less than 0.05.

Test of Independent Association: Health & Emotional Support

The SURVEYFREQ Procedure

Table of EMOSPRT by GENHLTH

EMOSPRT	GENHLTH	Frequency	Percent	Std Err of Percent	95% Confidence Limits for Percent		Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent	
No Support	Good	1321	24.1779	0.9780	22.1046	26.2512	48.5283	1.9535	44.3871	52.6696
	Fair	1757	24.7291	1.3152	21.9410	27.5173	49.6347	1.8315	45.7520	53.5174
	Bad	77	0.9152	0.1230	0.6544	1.1760	1.8370	0.2213	1.3678	2.3061
Has Support	Good	1092	21.5115	1.3435	18.6635	24.3595	42.8706	2.2068	38.1923	47.5488
	Fair	1940	26.7923	0.9657	24.7451	28.8394	53.3947	1.7190	49.7506	57.0389
	Bad	167	1.8740	0.3436	1.1455	2.6025	3.7347	0.6531	2.3501	5.1193

Frequency Missing = 3795

Rao-Scott Chi-Square Test	
Pearson Chi-Square	35.9988
Design Correction	1.4657
Rao-Scott Chi-Square	24.5604
DF	2
Pr > ChiSq	<.0001
F Value	12.2802
Num DF	2
Den DF	32
Pr > F	0.0001

Sample Size = 6354

```
/* Descriptive Stats: Test of Independent Assocation */
proc surveyfreq data=nhanes.analysis nosummary;
strata SDMVSTRA;
cluster SDMVPSU;
weight WTINT2YR;
tables EMOSPRT*GENHLTH/nowt chisq nototal row cl;
format GENHLTH GENHLTHf. EMOSPRT EMOSPRTf. ;
title "Test of Independent Association: Health & Emotional Support";
run;
```

Figure 3: Immigrant Status Domain Estimation using PROC SURVEYMEANS

Unadjusted Prevalence of Health Status of Immigrant Status Individuals

The SURVEYMEANS Procedure

Statistics for IMMIGRANT Domains

IMMIGRANT	Variable	Level	Label	N	Mean	Std Error of Mean	Sum	Std Error of Sum
Not Immigrant	GENHLTH	Good	Rating of general health	2260	0.471403	0.020601	96298644	7896655
		Fair	Rating of general health	3121	0.499920	0.017339	102124221	7644330
		Bad	Rating of general health	221	0.028677	0.003807	5858078	889188
Immigrant	GENHLTH	Good	Rating of general health	153	0.282563	0.023208	4804043	682659
		Fair	Rating of general health	576	0.698972	0.023376	11883686	1693250
		Bad	Rating of general health	23	0.018465	0.003750	313935	82909

```
/* USED IN REPORT METHOD 2: NHANES hypertension example analysis */
/* for categorical variables: use class & var statements */
/* for continuous variables: use just var statement */
/* Unadjusted prevalence of bad mental health days for immigrant, sex & time in us
breakdown */
proc surveymeans data=nhanes.analysis plots=none noobs sum median mean;
strata SDMVSTRA;
cluster SDMVPSU;
weight WTINT2YR;
class GENHLTH;
domain IMMIGRANT;
var GENHLTH;
format IMMIGRANT IMMIGRANTf. GENHLTH GENHLTHf.;
title "Unadjusted Prevalence of Health Status of Immigrant Status Individuals";
run;
```

Figure 4: Immigrant Status Domain Estimation using PROC SURVEYFREQ
Among the 7.68% of individuals being of immigrant status, 5.37% of them reported fair general health. While only 46.15% of the 92.32% US born individuals reported having fair general health.

Unadjusted Prevalence of Health Status of Immigrant Status Individuals

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	16
Number of Clusters	32
Number of Observations	10149
Sum of Weights	297136095

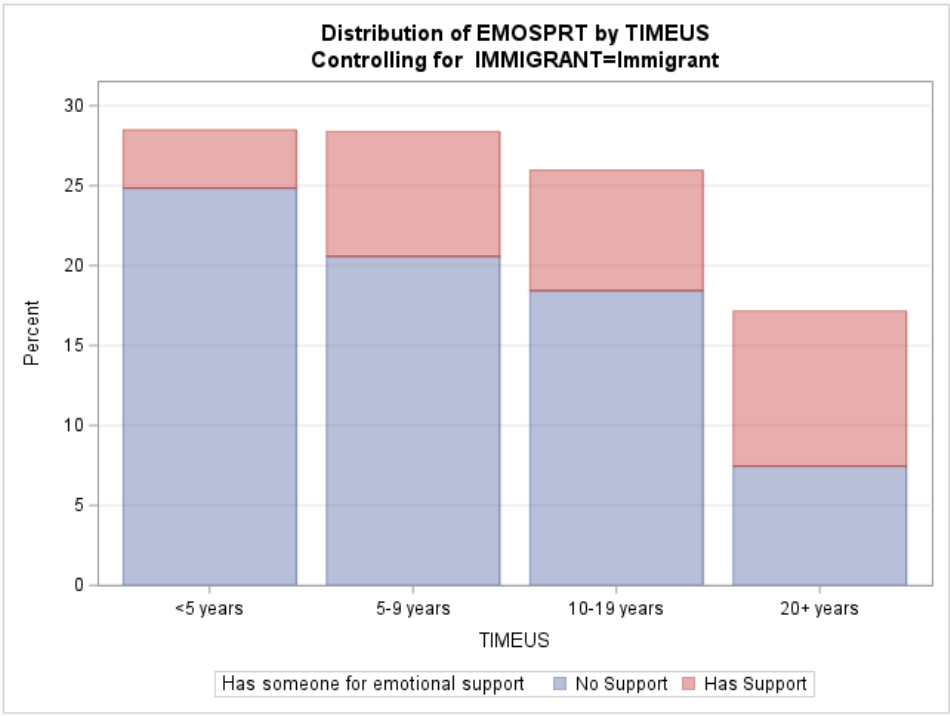
Table of IMMIGRANT by GENHLTH

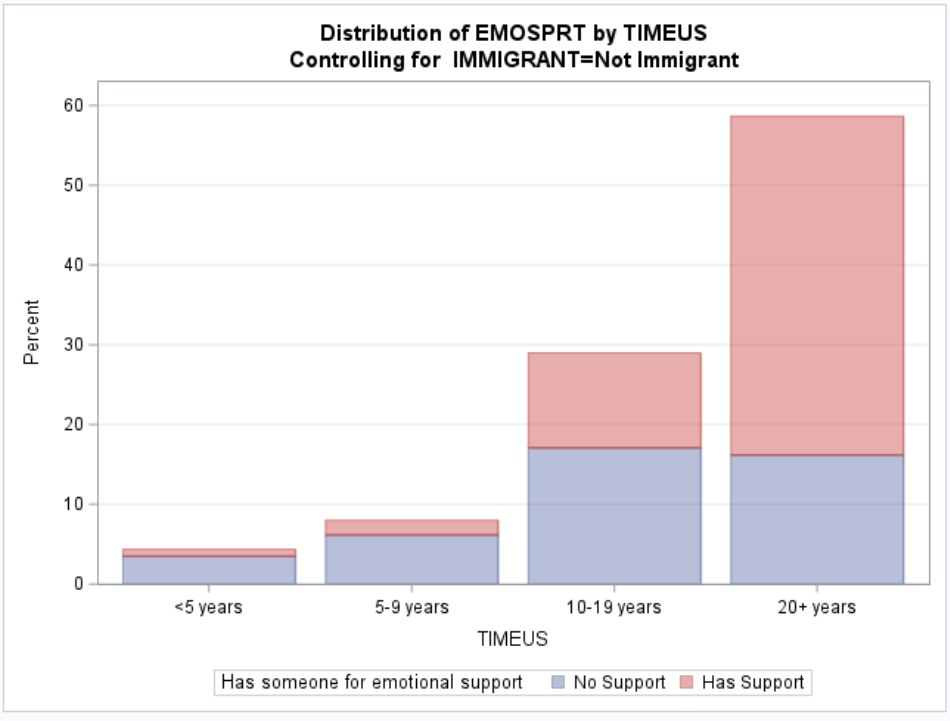
IMMIGRANT	GENHLTH	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent
Not Immigrant	Good	2260	96298644	7896655	43.5184	1.9747
	Fair	3121	102124221	7644330	46.1510	1.7523
	Bad	221	5858078	889188	2.6473	0.3569
	Total	5602	204280942	13891814	92.3168	1.3235
Immigrant	Good	153	4804043	682659	2.1710	0.4079
	Fair	576	11883686	1693250	5.3704	0.9418
	Bad	23	313935	82909	0.1419	0.0424
	Total	752	17001664	2264139	7.6832	1.3235
Total	Good	2413	101102687	7519149	45.6894	1.9412
	Fair	3697	114007907	6887572	51.5214	1.6388
	Bad	244	6172012	886553	2.7892	0.3639
	Total	6354	221282606	12443626	100.0000	

Frequency Missing = 3795

```
/* surveyfreq version of surveymeans */
proc surveyfreq data=nhanes.analysis;
strata SDMVSTRA;
cluster SDMVPSU;
weight WTINT2YR;
tables IMMIGRANT*GENHLTH;
format IMMIGRANT IMMIGRANTf. GENHLTH GENHLTHf.;
title "Unadjusted Prevalence of Health Status of Immigrant Status Individuals";
run;
```

Graph 1: Distribution of Social Support by Time of Residence in the US Among Immigrant vs US Born Status
Individuals with 20+ years of residence in the US, those of immigrant status report less social support, 9.69%, compared to US born individuals, 42.47%.






```
/* Method 2: bar graph*/
ods graphics on;
proc surveyfreq data = nhanes.analysis;
strata SDMVSTRA;
cluster SDMVPSU;
weight WTINT2YR;
tables IMMIGRANT*EMOSPRT*TIMEUS / plots (only) = freqplot (scale=percent
twoway=stacked)nowt;
format IMMIGRANT IMMIGRANTf. EMOSPRT EMOSPRTf. TIMEUS TIMEUSf. ;
title 'Distribution of Social Support by Time of Residence in the US (Immigrant Status)';
```

Figure 5: Logistic Regression: Model Information and Response Profile

Logistic regression predicts the probability of a binary dependent variable. PROC SURVEYLOGISTIC was used to fit a logistic regression model to account for complex survey design features. It is estimated that fair rating of general health was reported roughly 65.73% (20,295,947) of the 30,878,651 distinct individuals reported.

Model Information		
Data Set	NHANES.ANALYSIS	
Response Variable	GENHLTH	Rating of general health
Number of Response Levels	3	
Stratum Variable	SDMVSTRA	Masked Variance Pseudo-Stratum
Number of Strata	16	
Cluster Variable	SDMVPSU	Masked Variance Pseudo-PSU
Number of Clusters	32	
Weight Variable	WTINT2YR	Full Sample 2 Year Interview Weight
Model	Cumulative Logit	
Optimization Technique	Fisher's Scoring	
Variance Adjustment	None	

Variance Estimation	
Method	Taylor Series
Variance Adjustment	None

Number of Observations Read	10149
Number of Observations Used	1333
Sum of Weights Read	2.9714E8
Sum of Weights Used	30878651

Response Profile			
Ordered Value	GENHLTH	Total Frequency	Total Weight
1	Bad	51	781679
2	Fair	955	20295947
3	Good	327	9801025

Probabilities modeled are cumulated over the lower Ordered Values.

Figure 5b: Logistic Regression: Model Convergence Status

The message in the Model Convergence Status section notifies us that the Fisher scoring algorithm was successful in finding a unique solution to the Maximum Likelihood estimates.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Figure 5c: Logistic Regression: Maximum Likelihood Estimates

The parameters: presence of emotional support, adult age group, and Mexican American race/ethnicity group are significant at the alpha level, 0.05.

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept	Bad	-5.2650	0.4084	-12.89	<.0001
Intercept	Fair	-0.5899	0.3614	-1.63	0.1221
EMOSPRT	Has Support	0.5596	0.1416	3.95	0.0011
IMMIGRANT	Immigrant	0.1940	0.2175	0.89	0.3857
AGEGRP	Adults(>=18 years) AGEGRP	0.6649	0.2779	2.39	0.0294
RIAGENDR	Female	0.1909	0.1711	1.12	0.2811
RIDRETH1	Mexican American	0.9294	0.3364	2.76	0.0139
RIDRETH1	Non-Hispanic Black	-0.7303	0.3599	-2.03	0.0594
RIDRETH1	Non-Hispanic White	-0.5570	0.2769	-2.01	0.0614
RIDRETH1	Other Hispanic	0.5247	0.2722	1.93	0.0719
TIMEUS	10-19 years	0.2431	0.3022	0.80	0.4329
TIMEUS	20+ years	0.0544	0.3967	0.14	0.8927
TIMEUS	5-9 years	-0.00551	0.2369	-0.02	0.9817
NOTE: The degrees of freedom for the t tests is 16.					

Figure 5d: Logistic Regression: Overall Model

The Testing Global Null Hypothesis table provides three test statistics for the statistical hypothesis that the true parameters are not significantly different from zero. According to the Wald test statistic, which is appropriate in the presence of complex survey design, the p-value is 0.0033 (< 0.05) indicating that the null hypothesis can be rejected.

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	10.31	5.1629	82.6059	<.0001
Score	8.55	11	6	0.0078
Wald	11.82	11	6	0.0033
NOTE: Second-order Rao-Scott design correction 1.1306 applied to the Likelihood Ratio test.				

Figure 5e: Logistic Regression: Odds Ratio Estimates

The Odds Ratio Estimates table reports an estimated odds ratio for all variables not involved in an interaction. Adults with emotional support have 1.75 (95% CI: 1.30, 2.36) times the odds of having good health than individuals who report no emotional support. There is no association between immigrant status and good general health rating because the 95% confidence interval includes 1.

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
EMOSPRT Has Support vs No Support	1.750	1.296	2.363
IMMIGRANT Immigrant vs Not Immigrant	1.214	0.766	1.925
AGEGRP Adults(>=18 years) AGEGRP vs Not Adults(<18 years)	1.944	1.079	3.505
RIAGENDR Female vs Male	1.210	0.842	1.739
RIDRETH1 Mexican American vs Other/Multi-race	2.533	1.241	5.169
RIDRETH1 Non-Hispanic Black vs Other/Multi-race	0.482	0.225	1.033
RIDRETH1 Non-Hispanic White vs Other/Multi-race	0.573	0.319	1.031
RIDRETH1 Other Hispanic vs Other/Multi-race	1.690	0.949	3.009
TIMEUS 10-19 years vs <5 years	1.275	0.672	2.420
TIMEUS 20+ years vs <5 years	1.056	0.455	2.448
TIMEUS 5-9 years vs <5 years	0.995	0.602	1.643
NOTE: The degrees of freedom in computing the confidence limits is 16.			

```
/*************
**************
Logistic Regression
**************
*************/
proc surveylogistic data = nhanes.analysis;
weight WTINT2YR;
cluster SDMVPSU;
strata SDMVSTRA;
class EMOSPRT (ref='No Support')
IMMIGRANT (ref='Not Immigrant')
RIAGENDR RIDRETH1 AGEGRP TIMEUS/param=ref;
model GENHLTH = EMOSPRT IMMIGRANT AGEGRP RIAGENDR RIDRETH1 TIMEUS/vadjust=none;
format EMOSPRT EMOSPRTf. GENHLTH GENHLTHf. IMMIGRANT IMMIGRANTf.
RIAGENDR RIAGENDRf. AGEGRP AGEGRPf. RIDRETH1 RIDRETH1f. TIMEUS TIMEUSf. ;
title 'Logistic Regression';
run;
```

Summary (answer aim questions using all the results)
Compare the prevalence of social (emotional) support among immigrants: by demographic characteristics and time in the US.
Determine association between health and social (emotional) support.

Topic Resources

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4046043/
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7558190/
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8040330/#bib6
4. https://www.pewresearch.org/short-reads/2020/08/20/key-findings-about-u-s-immigrants/
5. https://www.macrotrends.net/countries/USA/united-states/population

SAS Survey Data Analysis Resources

1. https://stats.oarc.ucla.edu/sas/seminars/sas-survey/
2. https://support.sas.com/resources/papers/proceedings20/4635-2020.pdf
3. https://wwwn.cdc.gov/nchs/nhanes/tutorials/SampleCode.aspx
4. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_surveyfreq_sect002.htm
5. https://www.lexjansen.com/wuss/2017/133_Final_Paper_PDF.pdf
6. https://uta.pressbooks.pub/bigdataforepidemiology/chapter/chapter10-nhanes/
7. https://stats.oarc.ucla.edu/sas/seminars/sas-survey/
8. Complex Survey Data Analysis with SAS Taylor H. Lewis (book)

Didn't use but keep for reference:

1. https://med.nyu.edu/departments-institutes/population-health/divisions-sections-centers/epidemiology/sites/default/files/nyc-hanes-datasets-and-resources-training-part-iii-data-analysis-using-sas-and-sudaan.pdf
2. https://websites.umich.edu/~surveymethod/asda/349-2012.pdf
3. https://support.sas.com/documentation/onlinedoc/stat/131/surveyfreq.pdf