Gentiana Shameti

Colorado State University Global

MIS 581: Capstone – Business Intelligence and Data Analytics

Dr. Osama Morad

February 11, 2024

Race-ethnic Disparities in Chronic Conditions and Health Care Utilization among Women of Reproductive Age

Abstract

This capstone project analyzes race-ethnic disparities in chronic conditions and healthcare utilization among women of reproductive age. Using a nationally representative dataset, I analyzed data from the National Health and Nutrition Examination Survey to examine the prevalence of chronic conditions such as diabetes, hypertension, and obesity among women aged 18 to 45 across different racial and ethnic groups. Furthermore, I explored disparities in healthcare utilization, including access to preventive services, regular check-ups, and utilization of healthcare facilities. Preliminary findings suggest significant variations in the prevalence of chronic conditions among women of different racial and ethnic backgrounds. Additionally, disparities in healthcare utilization patterns are observed, with certain groups experiencing barriers to accessing timely and appropriate care. These findings underscore the importance of addressing systemic inequalities in healthcare provision and access to preventive services among women of reproductive age from diverse racial and ethnic backgrounds. By shedding light on these disparities, this capstone project aims to inform policymakers, healthcare providers, and advocacy groups about the urgent need for targeted interventions to improve healthcare access and outcomes for marginalized communities. Addressing these disparities is essential for promoting reproductive health equity and ensuring that all women have equal opportunities to achieve optimal health outcomes regardless of their race or ethnicity.

Introduction

I am an immigrant and came to the USA in December 1999. I have always been in excellent health, have had excellent health insurance, and access to New York City's best doctors and hospitals. In July 2020, after ending up in the emergency room with strong pains, I was diagnosed with a condition for which I had no symptoms and I was not aware that I had. After two weeks, I had emergency surgery for a condition that I didn't know I had it. Despite being always proactive in doing preventive checkups and tests, I still ended up undiagnosed for my condition. My personal experience in July 2020, made me decide that for the capstone project, I will explore race and ethnicity disparities in chronic conditions and health care utilization among women of reproductive age.

Objectives

The objective is to explore racial and ethnicity differences in chronic health conditions and health care utilization among women of reproductive age.

Description of the Dataset

The National Health and Nutrition Examination Survey (NHANES) is a national survey that monitors the health and nutritional status of adults and children across the United States, it has been gathering data from the National Center for Health Statistics (NCHS) since 1960. NHANES gathers data from both subjective interview objective physical examinations and laboratory tests. Kim et al. (2018) found that there is a significant difference in sociodemographic, clinical, and healthcare utilization characteristics among the different racial and ethnic groups. In addition, the study found that Asians had an increased likelihood of having both undiagnosed hypertension and diabetes, while Blacks and Hispanics had a higher likelihood of having undiagnosed diabetes compared to non-Hispanic, Whites after adjusting for sociodemographic characteristics and BMI.

The dataset that will be used for the capstone project is for the 2017-2020 period and includes separated data files on demographic, dietary, medical examination, laboratory, questionaries, and limited access data. The data was collected in participants' homes and on mobile examination centers. The study design is cross-sectional, and the data collection frequency is biennial (Kindratt, 2022). This dataset is collected before the pandemic. The dataset for the project will have data on demographics, blood pressure and cholesterol, diabetes, hospital utilization and access to care. Figure 1. shows the code for taking the files from the website of NHANES. Figure 2. shows the codes for sorting and merging the data files.

Figure 1

Code for taking the files from the website of NHANES

```
/* Step #1 Take the data from the website */

filename xptIn1 url "https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_DEMO.XPT";
libname xptIn1 xport;

filename xptIn4 url "https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_HUQ.XPT";
libname xptIn4 xport;

filename xptIn2 url "https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_MCQ.XPT";
libname xptIn2 xport;

filename xptIn3 url "https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_BPQ.XPT";
libname xptIn3 xport;

filename xptIn5 url "https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_DIQ.XPT";
libname xptIn5 xport;

/* Step #2 Create a permanent dataset on my computer */

libname mydata "/home/u63730569/sasuser.v94";

data mydata.MCQ_P; set xptIn2.P_MCQ;
keep SEQN MCQ035 MCQ160a MCQ160b MCQ160c MCQ160d MCQ160e MCQ160f MCQ160p MCQ220;
run;

data mydata.DIQ_P; set xptIn5.P_DIQ;
keep SEQN DIQ010;
run;

data mydata.BPQ_P; set xptIn3.P_BPQ;
keep SEQN BPQ030 BPQ080;
run;
```

Figure 2

Codes for sorting and merging the data files

```
proc sort data=mydata.DEMO_P; by seqn; run;
proc sort data=mydata.HUQ_P; by seqn; run;
proc sort data=mydata.MCQ_P; by seqn; run;
proc sort data=mydata.BPQ_P; by seqn; run;
proc sort data=mydata.DIQ_P; by seqn; run;

/* Step #4 Merge all analytical dataset in one*/

data mydata.WRA_HE1720;
merge mydata.DEMO_P (in=x) mydata.HUQ_P mydata.MCQ_P mydata.BPQ_P mydata.DIQ_P;
by seqn;
if x;


if RIDEXPRG=2 and WTMECPRP>0 then inAnalysis=1;
chronic=(bpq030=1)+(bpq080=1)+(diq010=1)+(mcq035=1)+(mcq160a=1)+(mcq160b=1)+(mcq160c=1)+(mcq160d=1)+(mcq160e=1)+(mcq160f=1)+(mcq160p=1)+(mcq22

if chronic = 0 than chroncat = 1;
else if chronic = 1 than chroncat = 2;
else if chronic > 1 than chroncat = 3;

if ROUTINECARE = 1 than ROUTINECARE2 = 1;
else if ROUTINECARE = 2 or ROUTINECARE = 3 than ROUTINECARE2 = 2;

if GenHealth = 1 than GenHealth2 = 1;
else if GenHealth = 2 or GenHealth = 3 than GenHealth2 = 2;
else if GenHealth = 4 or GenHealth = 5 than GenHealth2 =3;
run;
```

Variables Types

The demographics data file has the individual's details such as gender, age, marital status, language, preference, race and ethnicity. The file includes also questions about socioeconomic status, including the highest level of education, and income and questions about military services.

Figure 3

Description of variable for demographics file

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
12	AGE1855p	Num	8			
6	DMDEDUC2	Num	8			Education level - Adults 20+
13	EDUC	Num	8			
3	RIAGENDR	Num	8			Gender
4	RIDAGEYR	Num	8			Age in years at screening
7	RIDEXPRG	Num	8			Pregnancy status at exam
5	RIDRETH3	Num	8			Race/Hispanic origin w/ NH Asian
2	SDDSRVYR	Num	8			Data release cycle
10	SDMVPSU	Num	8			Masked variance pseudo-PSU
11	SDMVSTRA	Num	8			Masked variance pseudo-stratum
1	SEQN	Num	8			Respondent sequence number
8	WTINTPRP	Num	8	15.6	15.6	Full sample interview weight
9	WTMECPRP	Num	8	15.6	15.6	Full sample MEC exam weight

The self-reported race/ethnicity and categorized the race and ethnicity indicators into five groups: non- Hispanic Whites, non-Hispanic Blacks, Hispanics, Asians, and Others. Figure 4

shows the details of variable RIDRETH3. Figure 5 shows the code for adjusting the variables for the demographics file.

The data file on hospital utilization and access to care includes categorical variables such as the response on "general health conditions" on the scale from excellent to poor. The other variable of "routine place to go for healthcare" has responses as 'yes, 'there is no place', and 'there is more than one place'. The third variable is "#times received healthcare over the past year" with responses from 'none' up to '16 or more'.

Figure 4

Content of variable RIDRETH3.

Code or Value	Value Description	Count	Cumulative	Skip to Item
1	Mexican American	1990	1990	
2	Other Hispanic	1544	3534	
3	Non-Hispanic White	5271	8805	
4	Non-Hispanic Black	4098	12903	
6	Non-Hispanic Asian	1638	14541	
7	Other Race - Including Multi-Racial	1019	15560	
.	Missing	0	15560	

Note. Adapted from "National Health and Nutritional Examination Survey, Questionnaires, Datasets, and Related Documentation, NHANES 2017-March 2020". *https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_DEMO.htm*

Figure 5

Codes for modification of demographics data file

```
libname mydata "/home/u63730569/sasuser.v94";
data mydata.DEMO_P; set xptIn1.P_DEMO;
/*if 21 le RIDAGEYR le 65;*/
if 18 <= RIDAGEYR <= 55;
if DMDEDUC2 not in (.,7,9);
*** Adults 20 and over:
  * Age 20-80 and over: 0=20+ years, 1= 20-29 years, 2=30-39 years, 3=40-49 years, 4=50-59 years, 5=60-69 years, 6=70-79 years, 8=80 years and
  if 18<= ridageyr <30 then AGE1855p = 1;
  if 30<= ridageyr <40 then AGE1855p = 2;
  if 40<= ridageyr <55 then AGE1855p = 3;

  if dmdeduc2 in(1,2,3) then EDUC = 1;    /*HS DIPLOMA or LESS*/
  else if dmdeduc2=4 then EDUC=2;       /*Some college*/
  else if dmdeduc2=5  then EDUC =3;     /*COLLEGE*/
RIAGENDR=GENDER;
RIDRETH3=RACE;

keep SEQN SDDSRVYR RIDAGEYR RIDEXPRG GENDER DMDEDUC2 WTINTPRP WTMECPRP SDMVPSU SDMVSTRA RACE EDUC AGE1855p;
run;
```

The modification of the demographic file was done to select the reproductive age for women from age 18 to 55 years old. The other modification was to group the level of education for the women. If EDUC variable is equal to 1 the woman has high school, if it equals to 2 then the woman has some college and if it is equal to 3, the women has a college degree.

Figure 6 shows the code for adjusting data on hospital utilization and access to care.

Figure 6

Code for adjusting the data on hospital utilization and access to care

```
/*
HUQ010 - General health condition
HUQ030 - Routine place to go for healthcare
HUQ051 - #times receive healthcare over past year
HUD062 - How long since last healthcare visit
HUQ071 - Overnight hospital patient in last year
HUQ090 - Seen mental health professional/past yr*/


data mydata.HUQ_P; set xptIn4.P_HUQ;
IF HUQ010=7 OR HUQ010=9 THEN DELETE;
IF HUQ030=7 OR HUQ030=9 THEN DELETE;
IF HUD062=77  OR HUD062=99 THEN DELETE;
IF HUQ090=7 OR HUQ090=9 THEN DELETE;
IF HUQ051=99 THEN DELETE;
IF HUQ071=7 OR HUQ071=9 THEN DELETE;
GenHealth =HUQ010;
RoutineCare = HUQ030;
TimesCarePY = HUQ051;
HowlongCareVisit = HUD061;
HospitalStayPY = HUQ071;
MentalHealthVisitPY = HUQ090;
drop HUQ010 HUQ030 HUQ051 HUD062 HUQ071 HUQ090;
run;

PROC FREQ DATA=mydata.HUQ_P;
TABLES HUQ010 HUQ030 HUQ051 HUD062 HUQ071 HUQ090/LIST;
RUN;

PROC FREQ DATA=mydata.HUQ_P;
TABLES GenHealth RoutineCare TimesCarePY HowlongCareVisit HospitalStayPY MentalHealthVisitPY/LIST;
RUN;

proc contents data=mydata.HUQ_P;
run;
```

The data file on blood pressure and cholesterol includes two categorical variables such as the response to “ever been told by a doctor that you had high blood pressure” with answers as ‘yes’, ‘no’, ‘refused’ and ‘don’t know’. The other categorical variable is the response to response

to “doctor told you – high cholesterol level” with answers as ‘yes’, ‘no’, ‘refused’ and ‘don’t know’.

For this capstone project, the data file on diabetes has one categorical variable such as the response to “doctor told you have diabetes” with answers as ‘yes’, ‘no’, ‘refused’ and ‘don’t know’. The medical conditions data file includes categorical variables shown on Figure 7.

Figure 7

Content of medical conditions file

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Label
2	MCQ035	Num	8	Still have asthma
10	MCQ220	Num	8	Ever told you had cancer or malignancy
3	MCQ160A	Num	8	Doctor ever said you had arthritis
4	MCQ160B	Num	8	Ever told had congestive heart failure
5	MCQ160C	Num	8	Ever told you had coronary heart disease
6	MCQ160D	Num	8	Ever told you had angina/angina pectoris
7	MCQ160E	Num	8	Ever told you had heart attack
8	MCQ160F	Num	8	Ever told you had a stroke
9	MCQ160P	Num	8	Ever told you had COPD, emphysema, ChB
1	SEQN	Num	8	Respondent sequence number

Model of Data

Cross-sectional studies are conducted to examine the prevalence of a particular condition at a certain point in time. These studies are referred to as prevalence studies and they frequently take the form of surveys. Some advantages of these studies are that they are 1) cheap and quick, 2) simple to carry out and analyze, 3) useful for healthcare planning and investigating trends over

time, and 4) useful when routine data are not available. Some disadvantages are 1) not useful for conditions that have a short duration, 2) not the first choice for investigating casualty, and 3) sampling and data collection need great care Stewart (2022). The data set that will be used on the capstone project has the final variables as shown in Figure 8.

Figure 8

Final variables of the dataset

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
10	AGE1855p	Num	8			
30	BPQ030	Num	8			Told had high blood pressure - 2+ times
31	BPQ080	Num	8			Doctor told you - high cholesterol level
32	DIQ010	Num	8			Doctor told you have diabetes
4	DMDEDUC2	Num	8			Education level - Adults 20+
11	EDUC	Num	8			
12	GENDER	Num	8			
14	GenHealth	Num	8			
37	GenHealth2	Num	8			
18	HUD061	Num	8			
19	HospitalStayPY	Num	8			
17	HowlongCareVisit	Num	8			
21	MCQ035	Num	8			Still have asthma
29	MCQ220	Num	8			Ever told you had cancer or malignancy
22	MCQ160A	Num	8			Doctor ever said you had arthritis
23	MCQ160B	Num	8			Ever told had congestive heart failure
24	MCQ160C	Num	8			Ever told you had coronary heart disease
25	MCQ160D	Num	8			Ever told you had angina/angina pectoris
26	MCQ160E	Num	8			Ever told you had heart attack
27	MCQ160F	Num	8			Ever told you had a stroke
28	MCQ160P	Num	8			Ever told you had COPD, emphysema, ChB
20	MentalHealthVisitPY	Num	8			
13	RACE	Num	8			
3	RIDAGEYR	Num	8			Age in years at screening
5	RIDEXPRG	Num	8			Pregnancy status at exam
36	ROUTINECARE2	Num	8			
15	RoutineCare	Num	8			
2	SDDSRVYR	Num	8			Data release cycle
8	SDMVPSU	Num	8			Masked variance pseudo-PSU
9	SDMVSTRA	Num	8			Masked variance pseudo-stratum
1	SEQN	Num	8			Respondent sequence number
16	TimesCarePY	Num	8			
6	WTINTPRP	Num	8	15.6	15.6	Full sample interview weight
7	WTMECPRP	Num	8	15.6	15.6	Full sample MEC exam weight
35	chroncat	Num	8			
34	chronic	Num	8			
33	inAnalysis	Num	8			

Literature Review

DuBois & Eaton (1996) did a detailed categorization of the negro problem" in America. The analysis indicated that the higher the level of poor health for blacks was one important of racial inequality in the United States. The authors saw the racial differences in health as reflecting differences in "social advancements", the "vast conditions" under which blacks and whites lived. The list of contributing factors included poor heredity, neglect of infants, bad dwellings, poor food, and unsanitary living conditions. In addition, the study pointed out that black men had poorer health than black women and that the gender differences in health were larger for blacks than for whites.

Williams & Sternthal (2010) state that racial differences in health date back to some of the earliest health records in the United States, with blacks (or African Americans) having poorer health than whites across a broad range of health status indicators. There is vast evidence of the continued existence of racial differences in health.

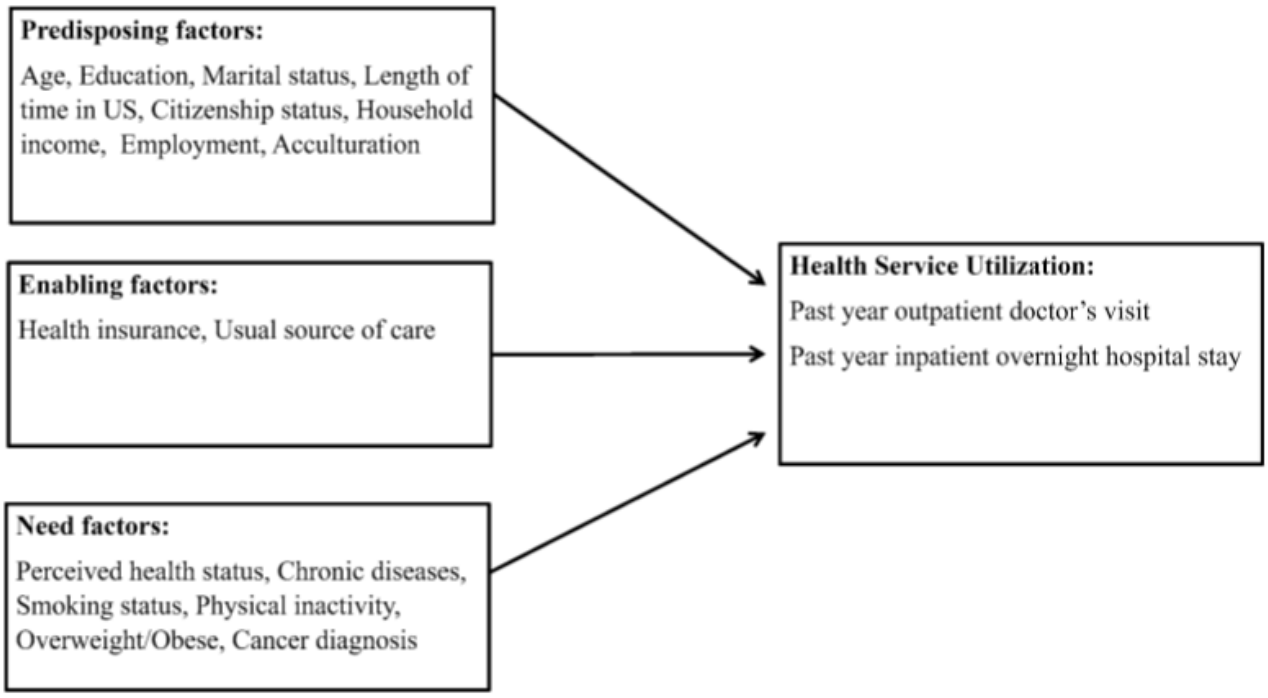
Seo & Strauss (2020) state that in 2018, about 44.7 million people in the United States (U.S.) were immigrants, which account for 14% of the U.S. population, with more than half of it being female. Due to the importance of the health care that immigrant women receive, researchers have compared U.S.–born and immigrant adult women and have found various health disparities. For example, adult Asian immigrant women in California in 2014 and 2015 were significantly less likely to have a past year visit to a doctor than were native-born adult women.

Andersen's behavioral model (BM) of health services use provides a theoretical and methodological framework for my capstone (1995). In the last years, this model has been widely used to explain individual or family characteristics influencing health care utilization. Based on this model, which is shown in Figure 9 the use of health care services by a person can be

predicted by a combination of predisposing characteristics (demographic factors), and enabling resources (individual or community resources to enable or hinder health service use).

Figure 9

Andersen's conceptual model.



Note. Adapted from "Revisiting the behavioral model and access to medical care: does it matter?"

One of the most repeatedly used elements in the literature of this model is having chronic health conditions. This model provides that healthcare use insinuates to the real use of official personal health services incorporating inpatient care, and physician visits. In addition, there are two different types of health services used in the study: a past year outpatient doctor's visit and a past year inpatient overnight hospital stay.

Babitsch et al. (2012) state that healthcare utilization is the point in health systems where patients' needs meet the professional system. Healthcare utilization depends on the quality of the medical facility. In addition, the patient's social characteristics lead to differences in healthcare

utilization. BM is a multilevel model that includes both individual and related elements of health service use. Patient's social characteristic still impacts the outcome of healthcare utilization. For example, women are inclined to utilize outpatient healthcare services more often than men.

Overview of Business Question

Based on my personal health experience, I chose for my capstone to analyze the race-ethnic disparities in chronic conditions and health care utilization among women of reproductive age. My goal is to explore the data information before COVID-19 on race-ethnic disparities in healthcare. My study will be focused on self-reported race/ethnicity and categorize the race and ethnicity indicators into five groups: non-Hispanic Whites, non-Hispanic Blacks, Hispanics, Asians, and Others. The topic I chose is well documented with studies done by professionals specialized in the public health field.

Read et al. (2021) did a study on health disparities among non-Hispanic Whites. The results revealed that health disparity within Whites population is almost as large as disparities within other racial groups. In this study, when the Whites were disaggregated by ancestry, mean health appeared to be more varied among Whites than between Whites and members of other racial/ethnic groups in many cases.

Shortreed et al. (2023) did a study on increased COVID-19 infection risk led to racial and ethnic disparities in severe COVID-19 outcomes.

While the goal of this study is to complete a capstone project for my class, I believe that there is more to be done to address racial-ethnicity problems that are more evident after COVID-19.

Questions and Hypothesis

The hypothesis and testing covered in the capstone project assisted in obtaining a better understanding of race-ethnicity that allowed them to answer the questions raised and to tested them through the use of analytics. Race and ethnicity are represented through the dataset. Below are the questions and hypotheses that were analyzed in the capstone project.

RQ 1: Among women of reproductive age, do women from race and ethnic minority groups have a higher burden (prevalence) of self-reported common chronic conditions (cardiovascular, including hypertension, hyperlipidemia, heart attack, stroke, diabetes, asthma, or any cancer) and self-reported fair or poor general health status than non-Hispanic white women?

H_0 1: Among women of reproductive age, there is no significant difference in burden (prevalence) of self-reported common chronic conditions and self-reported fair/poor general health status in women from race and ethnic minority groups than non-Hispanic white women and

H_α 1: Among women of reproductive age, there is a significant difference in burden (prevalence) of self-reported common chronic conditions and self-reported fair/poor general health status in women from race and ethnic minority groups than non-Hispanic white women. Figure 10 shows the codes for testing question 1 and Figure 11 and Figure 12 show the results of those tests.

Figure 10

Code for testing question 1.

```
/* Race and ethnicity CHRONIC */
proc surveyfreq data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
      tables  RIDRETH3* CHRONIC/row wchisq;
CLUSTER SDMVPSU;
STRATA SDMVSTRA;
WEIGHT WTMECPRP;
   run;

/* Number of Chronic Conditions by Race and enthinicity */
proc surveyfreq data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
      tables  RIDRETH3 * CHRONCAT/row wchisq;
CLUSTER SDMVPSU;
STRATA SDMVSTRA;
WEIGHT WTMECPRP;
   run;
```

The results shown in Figure 10 and Figure 11 show that zero chronic conditions were reported for 255 for non-Hispanic White (value 3 of RIDRETH3), 144 for Mexican-American (value 1 of RIDRETH3), 112 for Other Hispanic (value 2 of RIDRETH3), 261 for non-Hispanic Black (value 4 of RIDRETH3), 169 for non-Hispanic Asian (value 6 of RIDRETH3) and 45 for Other Race (value 7 of RIDRETH3). Therefore, non-Hispanic blacks have the heights number of women with zero conditions.

Figure 11

Results for question 1 part I. RIDRETH3 and Chronic.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	24
Number of Clusters	49
Number of Observations	1603
Sum of Weights	49489773.6

Table of RIDRETH3 by chronic

RIDRETH3	chronic	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	0	144	3701550	668024	7.4794	1.2996	67.6658	3.5982
	1	51	1173608	282778	2.3714	0.5587	21.4540	3.5118
	2	19	493047	126135	0.9963	0.2413	9.0131	1.8002
	3	3	73686	47041	0.1489	0.0915	1.3470	0.8299
	4	1	28454	28454	0.0575	0.0572	0.5201	0.5274
	5	0	.	.	.	.	.	.
	6	0	.	.	.	.	.	.
	7	0	.	.	.	.	.	.
	Total	218	5470344	937415	11.0535	1.7874	100.0000	
2	0	112	2998765	420234	6.0594	0.7362	66.4891	1.8412
	1	41	1008052	161366	2.0369	0.3060	22.3507	2.8555
	2	18	458589	139067	0.9266	0.2747	10.1679	2.6072
	3	1	8493	8493	0.0172	0.0171	0.1883	0.1856
	4	1	20420	20420	0.0413	0.0411	0.4528	0.4520
	5	0	.	.	.	.	.	.
	6	1	15843	15843	0.0320	0.0315	0.3513	0.3505
	7	0	.	.	.	.	.	.
	Total	174	4510162	594529	9.1133	1.0337	100.0000	
3	0	255	15243507	1453602	30.8013	2.2808	55.9383	2.5623
	1	133	7984154	875732	16.1329	1.3990	29.2991	2.2658
	2	53	2719383	434688	5.4948	0.8120	9.9792	1.4571
	3	15	798368	264092	1.6132	0.5462	2.9297	0.9212
	4	9	319111	143300	0.6448	0.3046	1.1710	0.5360
	5	3	144618	90312	0.2922	0.1776	0.5307	0.3274
	6	0	.	.	.	.	.	.
	7	2	41414	30824	0.0837	0.0641	0.1520	0.1137
	Total	470	27250554	2174203	55.0630	3.0158	100.0000	
4	0	261	4211555	614597	8.5100	1.0691	62.6733	2.4895
	1	98	1424979	222887	2.8793	0.4004	21.2055	2.1337
	2	49	706022	158617	1.4266	0.2893	10.5065	1.3395
	3	19	259925	83757	0.5252	0.1692	3.8680	1.0020
	4	9	82879	27965	0.1675	0.0576	1.2333	0.4291
	5	2	23702	17865	0.0479	0.0363	0.3527	0.2659
	6	2	10798	7937	0.0218	0.0158	0.1607	0.1154
	7	0	.	.	.	.	.	.
	Total	440	6719859	958972	13.5783	1.6574	100.0000	

Figure 12

Results for question 1 part II. RIDRETH3 and Chronic.

6	0	169	2718627	542727	5.4933	1.0833	82.0716	2.8852
	1	31	472616	119741	0.9550	0.2478	14.2676	2.3647
	2	9	104218	31927	0.2106	0.0661	3.1462	1.2289
	3	1	17048	17048	0.0344	0.0345	0.5146	0.4485
	4	0	.	.	.	.	.	.
	5	0	.	.	.	.	.	.
	6	0	.	.	.	.	.	.
	7	0	.	.	.	.	.	.
	Total	210	3312508	631319	6.6933	1.2723	100.0000	
7	0	45	906814	130775	1.8323	0.2767	40.7310	4.7818
	1	28	834401	218473	1.6860	0.4299	37.4785	6.0177
	2	10	185384	52914	0.3746	0.1128	8.3268	2.2219
	3	3	135037	94049	0.2729	0.1920	6.0654	4.0974
	4	5	164710	80371	0.3328	0.1637	7.3982	3.8203
	5	0	.	.	.	.	.	.
	6	0	.	.	.	.	.	.
	7	0	.	.	.	.	.	.
	Total	91	2226346	309883	4.4986	0.6443	100.0000	
Total	0	986	29780818	1941134	60.1757	1.7101		
	1	382	12897809	1118240	26.0616	1.5578		
	2	158	4666643	492979	9.4295	0.7850		
	3	42	1292556	287085	2.6118	0.5949		
	4	25	615574	179014	1.2438	0.3867		
	5	5	168320	92062	0.3401	0.1808		
	6	3	26641	17720	0.0538	0.0347		
	7	2	41414	30824	0.0837	0.0641		
	Total	1603	49489774	2853278	100.0000			

The Chroncat variable has a value equal to 1 if there is zero chronic condition, a value of 2 if there is one chronic condition, and a value of 3 if there is more than one condition. Figures 13 show the results for the Chroncat variable. Figure 14 totals values 2 and 3 of the Chroncat for each value of the RIDRETH3 variable. The results shown on Figure 14 indicate that non-Hispanic whites have the highest number of women with 1 condition and with more than one

condition with a total of 215, followed by non-Hispanic blacks with 179, Mexican Americans with 74, Other Hispanics with 62, Other Race with 28 and last non-Hispanic Asian with 46.

Figure 13

Results for question 1 part III. RIDRETH3 and Chroncat.

Table of RIDRETH3 by chroncat

RIDRETH3	chroncat	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	144	3701550	668024	7.4794	1.2996	67.6658	3.5982
	2	51	1173608	282778	2.3714	0.5587	21.4540	3.5118
	3	23	595187	144629	1.2026	0.2670	10.8802	2.0576
	Total	218	5470344	937415	11.0535	1.7874	100.0000	
2	1	112	2998765	420234	6.0594	0.7362	66.4891	1.8412
	2	41	1008052	161366	2.0369	0.3060	22.3507	2.8555
	3	21	503345	142384	1.0171	0.2781	11.1602	2.6262
	Total	174	4510162	594529	9.1133	1.0337	100.0000	
3	1	255	15243507	1453602	30.8013	2.2808	55.9383	2.5623
	2	133	7984154	875732	16.1329	1.3990	29.2991	2.2658
	3	82	4022893	584497	8.1287	1.1763	14.7626	1.8964
	Total	470	27250554	2174203	55.0630	3.0158	100.0000	
4	1	261	4211555	614597	8.5100	1.0691	62.6733	2.4895
	2	98	1424979	222887	2.8793	0.4004	21.2055	2.1337
	3	81	1083326	218103	2.1890	0.4084	16.1213	1.5899
	Total	440	6719859	958972	13.5783	1.6574	100.0000	
6	1	169	2718627	542727	5.4933	1.0833	82.0716	2.8852
	2	31	472616	119741	0.9550	0.2478	14.2676	2.3647
	3	10	121266	28732	0.2450	0.0602	3.6608	0.9877
	Total	210	3312508	631319	6.6933	1.2723	100.0000	
7	1	45	906814	130775	1.8323	0.2767	40.7310	4.7818
	2	28	834401	218473	1.6860	0.4299	37.4785	6.0177
	3	18	485131	143945	0.9803	0.3031	21.7905	6.3833
	Total	91	2226346	309883	4.4986	0.6443	100.0000	
Total	1	986	29780818	1941134	60.1757	1.7101		
	2	382	12897809	1118240	26.0616	1.5578		
	3	235	6811147	617469	13.7627	1.1164		
	Total	1603	49489774	2853278	100.0000			

Figure 14

Totals of values 2 and 3 of the Chroncat for each value of RIDRETH3 variable.

	RIDRETH3					
Chroncat	**1**	**2**	**3**	**4**	**6**	**7**
2	51	41	133	98	31	28
3	23	21	82	81	10	18
Total 2+3	**74**	**62**	**215**	**179**	**41**	**46**

Based on the above results there is evidence to reject the H_0 1 or null hypothesis of research question number 1, because the results show that among women of reproductive age, there is a significant difference in burden (prevalence) of self-reported common chronic conditions and self-reported fair/poor general health status in women from race and ethnic minority groups than non-Hispanic white women.

RQ 2: Among women of reproductive age, do women from race and ethnic minority groups have lower healthcare utilization (having a routine place for healthcare, staying in the hospital overnight, number of times visiting doctors or other healthcare professionals during the past year) than non-Hispanic white women?

H_0 2: There is no significant difference in health care utilization by race-ethnicity among women of reproductive age and

$H\alpha$ 2: There is a significant difference in health care utilization by race-ethnicity women of reproductive age.

Figure 15 shows the code of the tests to answer the research question 2. Figures 16 through 21 show the results of those tests.

Figure 15

Code of the tests related to research question 2.

```
/* Race and ethnicity RoutineCare2 */

proc surveyfreq data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
      tables  RIDRETH3 * ROUTINECARE2/row wchisq;
CLUSTER SDMVPSU;
STRATA SDMVSTRA;
WEIGHT WTMECPRP;
   run;

/* General Health status by Race and enthinicity */
proc surveyfreq data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
      tables  RIDRETH3 * GenHealth2/row wchisq;
CLUSTER SDMVPSU;
STRATA SDMVSTRA;
WEIGHT WTMECPRP;
   run;

proc surveyfreq data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
      tables  RIDRETH3*(GenHealth RoutineCare TimesCarePY HowlongCareVisit HospitalStayPY MentalHealthVisitPY)/row wchisq;
CLUSTER SDMVPSU;
STRATA SDMVSTRA;
WEIGHT WTMECPRP;
   run;
```

Figure 16

Results for question 2 part I. RIDRETH3 and ROUTINECARE2.

The SURVEYFREQ Procedure

Data Summary	
Number of Strata	24
Number of Clusters	49
Number of Observations	1603
Sum of Weights	49489773.6

Table of RIDRETH3 by ROUTINECARE2

RIDRETH3	ROUTINECARE2	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	166	4224753	756714	8.6042	1.4582	78.2455	2.8183
	2	49	1174602	240150	2.3922	0.4771	21.7545	2.8183
	Total	215	5399355	927840	10.9965	1.7887	100.0000	
2	1	120	3274155	549131	6.6682	0.9487	73.2443	5.7151
	2	53	1196026	257896	2.4359	0.5558	26.7557	5.7151
	Total	173	4470181	572042	9.1041	1.0010	100.0000	
3	1	387	22539008	2003357	45.9035	3.2565	83.2643	1.9941
	2	80	4530222	584325	9.2264	0.9290	16.7357	1.9941
	Total	467	27069231	2183859	55.1299	3.0365	100.0000	
4	1	376	5677089	801539	11.5621	1.3806	85.2054	2.1630
	2	61	985737	210909	2.0076	0.4085	14.7946	2.1630
	Total	437	6662826	943362	13.5697	1.6376	100.0000	
6	1	161	2540593	491469	5.1742	0.9959	77.3433	2.4890
	2	47	744232	171604	1.5157	0.3517	22.6567	2.4890
	Total	208	3284826	635709	6.6900	1.2923	100.0000	
7	1	77	1837957	285949	3.7432	0.6100	82.9991	5.1277
	2	13	376473	122832	0.7667	0.2444	17.0009	5.1277
	Total	90	2214430	309654	4.5100	0.6487	100.0000	
Total	1	1287	40093556	2397536	81.6555	1.2797		
	2	303	9007292	819745	18.3445	1.2797		
	Total	1590	49100848	2840253	100.0000			

Frequency Missing = 13

Wald Chi-Square Test	
Chi-Square	9.7748
F Value	1.9550
Num DF	5
Den DF	25
Pr > F	0.1207
Adj F Value	1.6422
Num DF	5
Den DF	21
Pr > Adj F	0.1926
Sample Size = 1590	

The ROUTINECARE2 variable has value 1 which means that the women did one routine care and value 2 if the women did 2 or 3 routine cares. The p-value associated with the F value is

0.1207. This p-value represents the probability of obtaining the observed results (or more extreme) if the null hypothesis is true. In other words, it indicates the level of significance of the chi-square test. In this case, with a p-value of 0.1207, we fail to reject the null hypothesis at the typical significance level (such as $\alpha = 0.05$), concluding that there is no significant difference in health care utilization by race-ethnicity among women of reproductive age.

Figure 17

Results for question 2 part II. RIDRETH3 and GenHealth2.

Table of RIDRETH3 by GenHealth2

RIDRETH3	GenHealth2	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	27	801988	181467	1.6333	0.3462	14.8534	1.7874
	2	130	3271443	603336	6.6627	1.1916	60.5895	3.0199
	3	58	1325924	254170	2.7004	0.4909	24.5571	3.1539
	Total	215	5399355	927840	10.9965	1.7887	100.0000	
2	1	19	649895	164793	1.3236	0.3165	14.5385	2.9057
	2	107	2586685	341303	5.2681	0.6934	57.8653	5.0146
	3	47	1233601	262199	2.5124	0.4602	27.5962	3.8460
	Total	173	4470181	572042	9.1041	1.0010	100.0000	
3	1	74	5296418	1035979	10.7868	1.6779	19.5662	2.7929
	2	319	18606955	1380906	37.8954	2.3533	68.7384	2.3640
	3	74	3165858	388807	6.4477	0.8794	11.6954	1.4214
	Total	467	27069231	2183859	55.1299	3.0365	100.0000	
4	1	59	893354	148599	1.8194	0.2694	13.4080	1.4889
	2	298	4658529	728688	9.4877	1.2850	69.9182	2.4682
	3	80	1110944	181399	2.2626	0.3511	16.6738	2.1183
	Total	437	6662826	943362	13.5697	1.6376	100.0000	
6	1	46	758724	190770	1.5452	0.3788	23.0979	2.9711
	2	147	2292013	454894	4.6680	0.9302	69.7758	2.6471
	3	15	234089	50356	0.4768	0.1079	7.1264	1.4876
	Total	208	3284826	635709	6.6900	1.2923	100.0000	
7	1	6	90170	49679	0.1836	0.0975	4.0719	2.1083
	2	67	1741598	254662	3.5470	0.5557	78.6477	5.3126
	3	17	382663	127476	0.7793	0.2564	17.2804	4.9701
	Total	90	2214430	309654	4.5100	0.6487	100.0000	
Total	1	231	8490549	1241583	17.2921	1.7602		
	2	1068	33157221	1609033	67.5288	1.5834		
	3	291	7453078	536156	15.1791	0.9255		
	Total	1590	49100848	2840253	100.0000			

Frequency Missing = 13

Wald Chi-Square Test	
Chi-Square	57.3478
F Value	5.7348
Num DF	10
Den DF	25
Pr > F	0.0002
Adj F Value	3.6703
Num DF	10
Den DF	16
Pr > Adj F	0.0103
Sample Size = 1590	

Figure 18

Results for question 2 part V. RIDRETH3 and TimesCarePY part 1.

Table of RIDRETH3 by TimesCarePY								
RIDRETH3	TimesCarePY	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	0	39	874813	148543	1.7817	0.2914	16.2022	2.2287
	1	50	1254251	261013	2.5544	0.5069	23.2296	2.9811
	2	73	1902850	400692	3.8754	0.8000	35.2422	3.0261
	3	24	651460	187041	1.3268	0.3654	12.0655	2.0648
	4	11	272174	114732	0.5543	0.2307	5.0409	1.8179
	5	5	131345	67338	0.2675	0.1391	2.4326	1.2061
	6	7	192362	59445	0.3918	0.1156	3.5627	1.2454
	7	3	40039	25520	0.0815	0.0508	0.7415	0.4379
	8	3	80060	50216	0.1631	0.1028	1.4828	0.9334
	Total	215	5399355	927840	10.9965	1.7887	100.0000	
2	0	40	813188	137134	1.6562	0.3032	18.1914	3.2469
	1	39	1032270	218633	2.1023	0.4168	23.0924	3.8907
	2	44	1193330	249577	2.4304	0.4918	26.6953	3.7169
	3	23	664331	158354	1.3530	0.2875	14.8614	3.0590
	4	6	139591	55138	0.2843	0.1124	3.1227	1.2337
	5	2	67633	54496	0.1377	0.1081	1.5130	1.2014
	6	10	230053	68160	0.4685	0.1435	5.1464	1.3094
	7	3	87030	51283	0.1772	0.1057	1.9469	1.1269
	8	6	242754	106449	0.4944	0.2015	5.4305	2.0319
	Total	173	4470181	572042	9.1041	1.0010	100.0000	
3	0	66	3716551	597616	7.5692	1.0044	13.7298	1.7790
	1	77	4333182	529559	8.8251	1.0146	16.0078	1.8304
	2	148	8759781	947384	17.8404	1.6936	32.3607	2.2970
	3	75	4602974	828296	9.3745	1.6443	17.0044	2.5103
	4	23	1198727	255235	2.4414	0.5728	4.4284	1.0261
	5	12	724479	192106	1.4755	0.3769	2.6764	0.7263
	6	27	1609575	440577	3.2781	0.8272	5.9461	1.5657
	7	12	632851	273633	1.2889	0.5495	2.3379	0.9731
	8	27	1491112	357546	3.0368	0.6772	5.5085	1.1379
	Total	467	27069231	2183859	55.1299	3.0365	100.0000	
4	0	54	881812	182573	1.7959	0.3502	13.2348	1.6872
	1	93	1346655	198453	2.7426	0.3657	20.2115	1.8674
	2	157	2501358	404297	5.0943	0.7037	37.5420	2.1646
	3	52	769910	151507	1.5680	0.2723	11.5553	1.5587
	4	27	341775	94154	0.6961	0.1879	5.1296	1.1299
	5	8	114438	42725	0.2331	0.0889	1.7176	0.5847
	6	19	299537	88489	0.6100	0.1760	4.4956	1.1068
	7	7	102681	66280	0.2091	0.1328	1.5411	0.9547
	8	20	304660	83990	0.6205	0.1717	4.5725	1.3627
	Total	437	6662826	943362	13.5697	1.6376	100.0000	

Figure 19

Results for question 2 part VI. RIDRETH3 and TimesCarePY part 2.

6	0	34	513072	127638	1.0449	0.2585	15.6195	2.3444
	1	76	1143667	284643	2.3292	0.5737	34.8167	3.6812
	2	65	1109149	261910	2.2589	0.5334	33.7658	3.9163
	3	22	350029	105201	0.7129	0.2175	10.6559	3.2719
	4	3	41069	23933	0.0836	0.0493	1.2503	0.7971
	5	0	.	.	.	.	.	.
	6	6	87441	37807	0.1781	0.0779	2.6620	1.1445
	7	1	23170	23170	0.0472	0.0466	0.7054	0.6987
	8	1	17229	17229	0.0351	0.0353	0.5245	0.5034
	Total	208	3284826	635709	6.6900	1.2923	100.0000	
7	0	13	325019	89851	0.6619	0.1803	14.6773	3.2170
	1	20	563551	149969	1.1477	0.2879	25.4490	5.6984
	2	24	643694	187016	1.3110	0.4023	29.0682	7.1156
	3	16	325974	101311	0.6639	0.2158	14.7205	4.0391
	4	4	75486	42599	0.1537	0.0887	3.4088	1.8250
	5	3	54210	43942	0.1104	0.0872	2.4481	1.9905
	6	4	83830	44631	0.1707	0.0915	3.7856	1.7128
	7	1	33347	33347	0.0679	0.0677	1.5059	1.5863
	8	5	109318	58274	0.2226	0.1171	4.9366	2.8516
	Total	90	2214430	309654	4.5100	0.6487	100.0000	
Total	0	246	7124455	702116	14.5098	1.0239		
	1	355	9673576	825374	19.7014	1.2845		
	2	511	16110162	1087302	32.8104	1.4600		
	3	212	7364678	911346	14.9991	1.6639		
	4	74	2068823	293279	4.2134	0.6707		
	5	30	1092105	225161	2.2242	0.4275		
	6	73	2502799	472739	5.0973	0.8551		
	7	27	919117	299497	1.8719	0.5925		
	8	62	2245133	419993	4.5725	0.7507		
	Total	1590	49100848	2840253	100.0000			
Frequency Missing = 13								

Figure 20

Results for question 2 part VII. RIDRETH3 and HospitalStayPY.

Table of RIDRETH3 by HospitalStayPY								
RIDRETH3	HospitalStayPY	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	20	498402	103251	1.0151	0.2194	9.2308	2.1056
	2	195	4900952	903668	9.9814	1.7354	90.7692	2.1056
	Total	215	5399355	927840	10.9965	1.7887	100.0000	
2	1	20	472435	141703	0.9622	0.2811	10.5686	2.5557
	2	153	3997746	496211	8.1419	0.8601	89.4314	2.5557
	Total	173	4470181	572042	9.1041	1.0010	100.0000	
3	1	57	3068269	466548	6.2489	0.8684	11.3349	1.4175
	2	410	24000961	1955758	48.8810	2.7685	88.6651	1.4175
	Total	467	27069231	2183859	55.1299	3.0365	100.0000	
4	1	58	849510	165463	1.7301	0.3226	12.7500	2.1023
	2	379	5813317	861972	11.8395	1.5053	87.2500	2.1023
	Total	437	6662826	943362	13.5697	1.6376	100.0000	
6	1	10	138445	49460	0.2820	0.0995	4.2147	1.3518
	2	198	3146380	615045	6.4080	1.2524	95.7853	1.3518
	Total	208	3284826	635709	6.6900	1.2923	100.0000	
7	1	15	377131	117237	0.7681	0.2245	17.0306	5.1432
	2	75	1837299	297928	3.7419	0.6476	82.9694	5.1432
	Total	90	2214430	309654	4.5100	0.6487	100.0000	
Total	1	180	5404192	559056	11.0063	0.9293		
	2	1410	43696656	2558072	88.9937	0.9293		
	Total	1590	49100848	2840253	100.0000			
Frequency Missing = 13								

Wald Chi-Square Test	
Chi-Square	12.9248
F Value	2.5850
Num DF	5
Den DF	25
Pr > F	0.0512
Adj F Value	2.1714
Num DF	5
Den DF	21
Pr > Adj F	0.0963
Sample Size = 1590	

Figure 21

Results for question 2 part VIII. RIDRETH3 and MentalHealthVisitPY.

Table of RIDRETH3 by MentalHealthVisitPY								
RIDRETH3	MentalHealthVisitPY	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	23	565014	182305	1.1507	0.3618	10.4645	2.3172
	2	192	4834341	797510	9.8457	1.5379	89.5355	2.3172
	Total	215	5399355	927840	10.9965	1.7887	100.0000	
2	1	27	835162	224414	1.7009	0.4247	18.6830	3.4412
	2	146	3635019	416594	7.4032	0.7449	81.3170	3.4412
	Total	173	4470181	572042	9.1041	1.0010	100.0000	
3	1	83	4627187	666510	9.4238	1.0753	17.0939	1.6126
	2	384	22442043	1707094	45.7060	2.5808	82.9061	1.6126
	Total	467	27069231	2183859	55.1299	3.0365	100.0000	
4	1	60	786476	141429	1.6018	0.2551	11.8039	1.4246
	2	377	5876351	843660	11.9679	1.4817	88.1961	1.4246
	Total	437	6662826	943362	13.5697	1.6376	100.0000	
6	1	9	156304	53874	0.3183	0.1077	4.7584	1.6583
	2	199	3128522	623828	6.3716	1.2715	95.2416	1.6583
	Total	208	3284826	635709	6.6900	1.2923	100.0000	
7	1	20	495263	170480	1.0087	0.3444	22.3653	6.2866
	2	70	1719167	245376	3.5013	0.5239	77.6347	6.2866
	Total	90	2214430	309654	4.5100	0.6487	100.0000	
Total	1	222	7465405	893157	15.2042	1.2787		
	2	1368	41635443	2232560	84.7958	1.2787		
	Total	1590	49100848	2840253	100.0000			
Frequency Missing = 13								

Wald Chi-Square Test	
Chi-Square	18.1602
F Value	3.6320
Num DF	5
Den DF	25
Pr > F	0.0132
Adj F Value	3.0509
Num DF	5
Den DF	21
Pr > Adj F	0.0318
Sample Size = 1590	

Findings Related to Research Question 1

The analysis of question research number 1 shows that women from race and ethnic minority groups have a higher burden (prevalence) of self-reported common chronic conditions (cardiovascular, including hypertension, hyperlipidemia, heart attack, stroke, diabetes, asthma, or any cancer) and self-reported fair or poor general health status than non-Hispanic white women.

Figure 22

Summary of Race – Ethnicity and Chronic Conditions

Race	Total Number of women with 0, 1+ conditions	Race	Total Number of women with 1+ conditions	Race	Total Number of women with 0 conditions
Non-Hispanic White	470	Non-Hispanic White	215	Non-Hispanic Black	261
Non-Hispanic Black	440	Non-Hispanic Black	179	Non-Hispanic White	255
Mexican - American	218	Mexican - American	74	Non-Hispanic Asian	169
Non-Hispanic Asian	210	Other Hispanic	62	Mexican - American	144
Other Hispanic	174	Other Race Including Multi-Racial	45	Other Hispanic	112
Other Race Including Multi-Racial	91	Non-Hispanic Asian	41	Other Race Including Multi-Racial	45

Figure 22 shows that when it comes to women with zero chronic conditions Non-Hispanic Black women have the highest number among all races. It's slightly higher than Non-Hispanic Whites. In addition, the other races have a high report number of having zero conditions compared with Non-Hispanic Whites versus a low report number of having 1 or more conditions compared with Non-Hispanic Whites.

For example, there is a total of 134 =74+62 reports of having 1 or more conditions for Mexican-Americans and Other Hispanics added together compared to 215 reports for Non-Hispanic Whites, so the difference of these reports is 181=215-134. However, there is a total of 256 =144+112 reports of having zero conditions for Mexican-Americans and Other Hispanics

added together compared to 255 reports for Non-Hispanic Whites, so the difference between these reports is -1= 255-256.

This means that Mexican-Americans and Other Hispanics end up not being diagnosed with chronic conditions like non-Hispanic whites and Blacks.

Findings Related to Research Question 2

The test results of race and ethnicity variables RIDRETH3 and ROUTINECARE2 show that there is no significance between race and ethnicity and routine care. This led to the failure to reject the null hypothesis for research question 2.

A deeper analysis of Chi-square test shows for variables RIDRETH3 and GenHealth2, the p-value associated with the F value is 0.0002. This means if the null hypothesis was associated with these two variables, we would fail to reject the null hypothesis, which is this case would have been there is no significant difference in GenHealth2 by race-ethnicity among women of reproductive age.

Chi-square test shows for variables RIDRETH3 and HospitalStayPY, the p-value associated with the F value is 0.0512. This means if the null hypothesis was associated with these two variables, we would fail to reject the null hypothesis, which is this case would have been there is no significant difference in HospitalStayPY by race-ethnicity among women of reproductive age.

Chi-square test shows for variables RIDRETH3 and MentalHealthVisitPY, the p-value associated with the F value is 0.0132. This means if the null hypothesis was associated with these two variables, we would fail to reject the null hypothesis, which is this case would have been there is no significant difference in MentalHealthVisitPY by race-ethnicity among women of reproductive age.

As a summary, we did find not a significant difference in ROUTINECARE2 variable by race-ethnicity women of reproductive age.

Additional Statistical Tests

Canonical correlation analysis (CCA) is a statistical technique used to examine the relationship between two sets of variables (X and Y). In CCA, one set of variables (X) is considered the predictor or independent variables, while the other set (Y) is considered the outcome or dependent variables.

Afifi et al. (2012) explain that in CCA, the goal is to identify linear combinations of variables from each set (U1 for Y and V1 for X) that maximize the correlation between them. These linear combinations are called the first canonical variables. By selecting appropriate coefficients (a and b), CCA aims to maximize the correlation between U1 and V1.

It's important to note that in CCA, the mean of each variable in both sets (X and Y) is typically subtracted from the original data. This ensures that the sample means of all X and Y variables are zero. This centering step helps in interpreting the results of CCA.

The resulting correlation between the first canonical variables (U1 and V1) is termed the first canonical correlation. This correlation represents the maximum correlation achievable between the two sets of variables by linearly combining them.

In summary, CCA allows researchers to explore the relationship between two sets of variables by identifying linear combinations that maximize their correlation. The resulting canonical variables and correlations provide insights into the underlying structure of the relationship between the sets of variables. Figure 23 has the code for PROC CANCORR. Figures 24 and Figure 25 have the results.

Figure 23

Code for PROC CANCORR

```
434 proc cancorr corr data=mydata.WRA_HE1720 (WHERE=(inAnalysis=1));
435   var CHRONIC ROUTINECARE2 GenHealth2;
436   with RIDRETH3 ridageyr;
437 run;
```

Figure 24

Results for PROC CANCORR part I.

The CANCORR Procedure

Correlations Among the Original Variables

Correlations Among the VAR Variables			
	chronic	ROUTINECARE2	GenHealth2
chronic	1.0000	-0.1020	0.2692
ROUTINECARE2	-0.1020	1.0000	-0.0180
GenHealth2	0.2692	-0.0180	1.0000

Correlations Among the WITH Variables		
	RIDRETH3	RIDAGEYR
RIDRETH3	1.0000	-0.0096
RIDAGEYR	-0.0096	1.0000

Correlations Between the VAR Variables and the WITH Variables		
	RIDRETH3	RIDAGEYR
chronic	-0.0026	0.2579
ROUTINECARE2	-0.0511	-0.1204
GenHealth2	-0.0989	0.0691

Figure 25

Results for PROC CANCORR part II.

The CANCORR Procedure

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)				Test of H0: The canonical correlations in the current row and all that follow are zero				
					Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.275395	0.271915	0.023184	0.075842	0.0821	0.0692	0.8647	0.8647	0.91244392	24.77	6	3170	<.0001
2	0.112584	0.110883	0.024768	0.012675	0.0128		0.1353	1.0000	0.98732475	10.18	2	1586	<.0001

Multivariate Statistics and F Approximations					
S=2 M=0 N=791.5					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.91244392	24.77	6	3170	<.0001
Pillai's Trace	0.08851739	24.48	6	3172	<.0001
Hotelling-Lawley Trace	0.09490420	25.06	6	2111.6	<.0001
Roy's Greatest Root	0.08206623	43.39	3	1586	<.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					

Figure 24 displays the correlations among the variables. The highest correlation among the variables being 0.2692 is between the Chronic and GenHealth2 variables. This means that there is a high correlation between chronic conditions and general health conditions.
The highest correlation "var" with variables being 0.25579 is between Chronic and RIDAGEYER. This means that there is a high correlation between chronic conditions and age of women. Figure 25 shows the first canonical correlation is 0.275395, which would appear to be a little larger than any of the between-set correlations.

Research Method

The data collected from the National Health and Nutrition Examination Survey (NHANES) will be used to address the questions and to test the hypothesis of race-ethnic disparities in common chronic conditions and healthcare utilization among women of reproductive age.

PROC SURVEYFREQ will be used as the primary code for analysis of the relationships of race-ethnic to common chronic conditions and healthcare utilization. Since the majority of the dataset has categorial variables the tests that will be used for statistical analysis with be the Binomial test and chi-square test. The chi-square test is used when we want to see if there is a relationship between two categorical variables. This test assumes that the expected value for each cell is five or higher.

Ethical Considerations

Ballantyne (2019) states that data analytics has the potential to match the health product or service to the planned patient, improve health results, and to decrease costs. The data subject is the individual to whom the data refers, in this case, patients. Stakeholders are those people or

entities that have an interest in the data (they may experience risks, benefits, or opportunity costs). Communities are one type of stakeholder.

World Medical Association (WMA) (2020) states that the use of health information in a specific research study and the active decision by a patient to contribute to a biobank or databank. Figure 26 shows the consideration of WMA related to the use of health information for research.

Figure 26

WMA Declaration of Taipei on ethical considerations regarding databases and biobanks

Paragraph 12: 'If the data or biological material are collected and stored in a Health Database or Biobank for multiple and indefinite uses, consent is only valid if the concerned individuals have been adequately informed about:

- The purpose of the Health Database or Biobank;
- The risks and burdens associated with collection, storage and use of data and material;
- The nature of the data or material to be collected;
- The procedures for return of results including incidental findings;
- The rules of access to the Health Database or Biobank;
- How privacy is protected;
- The governance arrangements as stipulated in paragraph 21;
- That in case the data and material are made non-identifiable the individual may not be able to know what is done with their data/ material and that they will not have the option of withdrawing their consent;
- Their fundamental rights and safeguards established in this Declaration; and
- When applicable, commercial use and benefit sharing, intellectual property issues and the transfer of data or material to other institutions or third countries.'

Note. Adapted from "Declaration of Taipei on ethical considerations regarding health databases and biobanks".

Allen (2016) states that protecting data privacy should be understood as an ethical responsibility of good governments, businesses, and individuals. The methods of data collection

and analysis associated with Big Data represent challenges to individuals' capacity to make significant privacy-protective involvements.

Ienca et al. (2018) did a study reviewing the ethics of big data health research. The ethical and legal challenges include the risk of compromising privacy, personal autonomy, and the solidarity-based approach to healthcare funding, as well as the effects on public demand for transparency, trust, and fairness while using big data.

They concluded that in the future the researchers may be required to rethink what is considered "public" data and what counts as "harm" in data-driven research. The ethics of big should be reduced of big data should not be reduced to a privacy challenge but it should include several positive ethical goals. One example would be the issue of data ownership, group-level ethical harms, and the distinction between academic and commercial use of big data.

While privacy is viewed as both a legal right and a human right, it is also critical to acknowledge that privacy is of ethical importance.

Limitation

Ballantyne (2019) points out that research ethics focuses on individual consent because consent is an important guard against harm, as shown by the history of misuse of research participants on researchers' hands. However, limiting the focus on individual consent shadows the risk of corporate and political interests in data. Data research comes with risks and it is not morally equal to register a subject in clinical without research. Since the communities are not identical or standardized it is a real concern about trivial and controlling 'community engagement', as well as the potential abuse of volunteers in the 'citizen science' movement.

The limitation of the National Health Institute Survey (NHIS) is the inability to calculate reliable statewide estimates. Prior studies have revealed that self-reported receipt of

mammograms for initial finding of breast cancer and other preventive health services are consistent with reports from medical providers and electronic records. A limitation of the Medical Expenditure Panel Survey (MEPS) is that the sample size is much smaller in comparison to prior surveys (Kindratt, 2022, p. 115).

Another limitation is that other medical professionals such as physician assistants and nurse practitioners as advanced practice providers for workforce research. The place of birth questions is limited to US or foreign-born only and there are no data is collected on the country of birth. This limits the ability for data disaggregation among foreign-born groups. As shown in the beginning of this paper, the large number of subsection files requires multiple merges of data files for each survey year. A limitation of self-reported data for cancer screenings is that the data may not represent screening estimates. Self-reported data may not be as accurate as other measures such as electronic medical records.

Conclusion

Kington & Smith, (1997) decades ago did a study to document the differences in health status among racial and ethnic groups across the United States. Their study found that socioeconomic status (SES) plays a much greater role in clarifying racial and ethnic disparities in the capability to function once a person has a chronic illness than in explaining who has chronic illnesses. This suggests that once an individual acquires a chronic condition, their socioeconomic status becomes a critical determinant of their ability to function effectively.

The ongoing evolution of demographics in the United States, including factors such as immigration, interracial marriage, and changes in educational attainment, has implications for the dynamics of race-ethnic disparities in healthcare. These demographic shifts may influence the

composition of racial and ethnic groups, as well as their socioeconomic status, thereby potentially altering patterns of health disparities over time.

The result of this project indicated that there is no significant difference in burden (prevalence) of self-reported common chronic conditions and self-reported fair/poor general health status in women from race and ethnic minority groups than non-Hispanic white women. And among women of reproductive age, there is no significant difference in health care utilization by race-ethnicity among women of reproductive age.

Additional statistical tests showed that there is a high correlation between chronic conditions and general health conditions and that there is a high correlation between chronic conditions and the age of women.

Recommendations

Based on the findings of your project and considering the ongoing demographic evolution in the United States, particularly regarding immigration, interracial marriage, and changes in educational attainment, I would recommend the following:

1. Continued Monitoring and Analysis - As demographic shifts continue to shape the composition of racial and ethnic groups in the United States, it's essential to monitor and analyze healthcare disparities regularly. This includes examining patterns of chronic conditions, general health status, and healthcare utilization among different demographic groups, including race and ethnicity.
2. Exploring Intersectionality – Given the complexity of intersecting factors such as race, ethnicity, socioeconomic status, and age, future research should explore intersectionality in healthcare disparities. This involves understanding how multiple identity factors intersect to influence health outcomes and healthcare access among diverse populations.

3. Longitudinal Studies - Conducting longitudinal studies can provide valuable insights into the dynamic nature of healthcare disparities over time. By tracking changes in health outcomes and healthcare utilization among various demographic groups, researchers can identify trends and factors contributing to disparities.
4. Policy Implication The findings of this project, indicating no significant differences in chronic conditions and healthcare utilization by race and ethnicity among women, have important policy implications. Policymakers should consider these findings when designing interventions and healthcare programs aimed at reducing disparities and promoting health equity.
5. Health Promotion and Education - Given the high correlation between chronic conditions, general health status, and age among women, targeted health promotion and education efforts are crucial. These efforts should focus on preventive measures, early detection, and management of chronic conditions, particularly among vulnerable populations.

By considering these recommendations and continuing to explore the intersection of demographic shifts and healthcare disparities, we can work towards achieving more equitable healthcare outcomes for all individuals, regardless of race, ethnicity, or other demographic factors.

Reference

Afifi, A., May, S., Clark, V.A. (2012). Practical Multivariable Analysis. Fifth Edition. Taylor & Francis Group.

Allen, A. L. (2016). Protecting one's own privacy in a big data economy. *Harv. L. Rev. F.*, *130*, 71.

Andersen, R. M. (1995). Revisiting the behavioral model and access to medical care: does it matter?. *Journal of health and social behavior*, 1-10.

Babitsch, B., Gohl, D., & Von Lengerke, T. (2012). Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998–2011. *GMS Psycho-Social-Medicine*, *9*.

Ballantyne, A. (2019). Adjusting the focus: a public health ethics approach to data research. *Bioethics*, *33*(3), 357-366.

Dieleman, J. L., Cao, J., Chapin, A., Chen, C., Li, Z., Liu, A., ... & Murray, C. J. (2020). US health care spending by payer and health condition, 1996-2016. *Jama*, *323*(9), 863-884.

DuBois, W. E. B., & Eaton, I. (1996). *The Philadelphia Negro: A Social Study*. University of Pennsylvania Press.

Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PloS one*, *13*(10), e0204937.

Kindratt, T. (2022). Big data for epidemiology: Applied data analysis using national health surveys. Peace Ossom-Williamson. Mavs Open Press. Arlington.

NHANES. (2022). National Health and Nutritional Examination Survey, Questionnaires, Datasets, and Related Documentation, NHANES 2017-March 2020". *https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_DEMO.htm*

Read, J. N. G., Lynch, S. M., & West, J. S. (2021). Disaggregating heterogeneity among non-Hispanic Whites: evidence and implications for US racial/ethnic health disparities. *Population research and policy review*, *40*, 9-31.

Reed-Berendt, R., Dove, E. S., Pareek, M., & UK-REACH Study Collaborative Group. (2022). The Ethical Implications of Big Data Research in Public Health: "Big Data Ethics by Design" in the UK-REACH Study. *Ethics & human research*, *44*(1), 2-17.

Seo, J. Y., & Strauss, S. M. (2020). Association of cancer history and health care utilization among female immigrants using NHANES 2007–2016 data. *Journal of Immigrant and Minority Health*, *22*, 1208-1216.

Shortreed, S. M., Gray, R., Akosile, M. A., Walker, R. L., Fuller, S., Temposky, L., ... & Dublin, S. (2023). Increased COVID-19 infection risk drives racial and ethnic disparities in severe COVID-19 outcomes. *Journal of racial and ethnic health disparities*, *10*(1), 149-159.

United States Census Bureau. 2018 American Community Survey (ACS). 2019. Retrieved January 21, 2024. https://www.census.gov/programs-surveys/acs/.

Williams, Dr., Sternthal M. (2010). Understanding racial-ethnic disparities in health: sociological contributions. *Journal of Health Social Behavior 51 Suppl(Suppl): S15-27.*

World Medical Association (June 4, 2020). Declaration of Taipei on ethical considerations regarding health databases and biobanks. Retrieved January 28, 2024 https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks