Maha Pasha
Jeff Parker
8 May 2013

Metabolic and Cardiovascular Disease and Associated Risk Factors in Ethnic South Asians

Introduction

Ethnic South Asians (Asian Indian, Pakistani, Bengali, and Sri Lankan) make up 1% of the United States population and are the third largest Asian minority[1]. They have been the fastest growing minority with the highest immigration rates. The Indian population has grown 130% between the early1990s and the early 2000s. In this time frame, the South Asian population as a whole doubled to over 2 million. Because of the drastic increase in South Asian immigrants understanding their health needs is important in addressing health related problems and making health care as efficient and effective as possible for all subsets of the population.

Within the United States studies have found contradicting conclusions of the prevalence of diabetes, coronary heart disease (CHD), and hypertension in ethnic South Asians. Many studies have found South Asians to be more likely to report higher rates of diabetes, CHD, and hypertension. Similar findings have been found in the UK, but there are also studies that found high rates of CHD and diabetes, but not hypertension[2]. This study analyzes the prevalence of CHD, hypertension, and diabetes with their associated risk factors in South Asian's compared to those of non-Hispanic whites in the United States along with analyzing the impact, if any, of duration of stay in the US. Compared to other studies, we use the most recent data to give the most comprehensive perspective of the health situation of South Asians in the United States.

Data

We used cross-sectional data from the 2011 National Health Interview Survey (NHIS), limited to adults of all races. The sample consisted of 19733 Non-Hispanic Whites, 401 South Asians, 1658 other Asians, 4974 Non-Hispanic Black, 238 American Indians/Alaskan Natives, 5846 Hispanic, and 83 multiracial respondents. For our key independent variable, South Asian, we use the answer to the question asking what race the respondent identifies with, and "Asian Indian", the closest option to South Asian is used to represent the South Asian respondents of the survey.

Two files from the NHIS survey were used: *personsx.dta* and *samadult.dta.* The first file contained all members surveyed in the household and the latter was a sample adult file with one adult from each household. Refer to Appendix I for merge commands and data cleaning information.

---

[1] U.S. Census 2011

[2] Diabetes and Cardiovascular Disease Among Asian Indians in the United States by Sarita Mohanty et. al.

Variables

Dependent Variables

Title	Explanation	Type
DIB	Self reported overt or borderline diabetes (1 = diabetes; 0 = no diabetes)	binary variable
CHD	Self reported Coronary heart disease (1 = CHD; 0 = no CHD)	binary variable
HYP	Self reported Hypertension (1 = hypertension; 0 = no hypertension.	binary variable

Independent Variables

Title	Explanation	Type
age	age of observation	continuous variable
aian	Race coding: Alaskan Indian/American Native (1 = aian, 0 = not aian)	binary variable
black	Race coding: Non-Hispanic Black (1 = nhblack; 0 = not nhblack)	binary variable
bmi	body mass indenx	continuous variable
duration	Living in the United States for ten or more years because ten is the standard time frame given for acculturation (1 = 10 + years in US; 0 = less than 10 years in US)	binary variable
female	male or female (1 = female; 0 = male)	binary variable
gendoc	A general doctor as place usually go for routine care (1 = general doctor; 0 = no general doctor)	binary variable
hispanic	Race coding: Hispanic (1 = Hispanic; 0 = non-hispanic)	binary variable
inMidwest	Region coding: in Midwest of United States (1 = in Midwest; 0 = not in Midwest)	binary variable
inNortheast	Region coding: in Northeast of United States (1 = in Northeast; 0 = not in Northeast)	binary variable
inSouth	Region coding; in South of United States (1 = in South; 0 = not in South)	binary variable
inWest	Region coding: in West of United States (1 = in West; 0 = not in West)	binary variable
incomeGreater25	Income greater than $25000 which is the poverty line (1 = income greater than or equal to $25000; 0 = income less than $25000)	binary variable
insured	Covered by any form of insurance (1 = insured; 0 = uninsured)	binary variable
novigorous	Engages in vigorous activity less than once a week (1 = no vigorous activity; 0 = vigorous activity)	binary variable
otherasian	Race coding: Other Asian (1= other Asian; 0 = not other Asian)	binary variable
otherrace	Race coding: Multiracial (1 = other race; 0 = not other race)	binary variable
smoke	Current or past smoker (1 = not current or past smoker; 0 = never a smoker)	binary variable
southasian	Race coding: South Asian/ Asian Indian (1 = South Asian; 0 = not South Asian)	binary variable
USborn	Born in the United States (1= US born; 0 = not US born)	binary variable
UScitizen	Citizen of the United States (1 = citizen; 0 = not citizen)	binary variable
white	Race coding: Non-Hispanic White (1= nhwhite; 0 =	binary variable

	nhwhite)	
wtfa_sa	Final Annual Weights as given by NHIS	

Variable Summary for South Asians and Non-Hispanic Whites

	Non-Hispanic White	South Asian
Sample Size	19509	402
Percentage of total population	59.77%	1.21%
Mean age 17-44 years old, % 45+, %	50.55 7700 11809	36.95 309 93
Gender, % Male	46.19%	53.62%
Insurance status, % Uninsured	88.08%	89.25%
Income status, % Greater => 25,000	64.63%	74.51%
Born in the US, % Yes	95.04%	9.27%
If not born in US, % Lived in U.S. 0-10 years Lived in U.S. > 10 years	 16.01% 84%	 46.98% 53.02%
Smoking status, % Current or former smoker	47.19%	14.21%
Activity level (vigorous activity), % Never active or less than Than once a week	57.65%	54.61%
Mean Bmi Normal or low, % Overweight, % Obese, %	27.45 35.69% 33.67% 28.91%	24.67 55.22% 32.84% 8.71%
Has a general doctor, %	26.5%	16.81%
Borderline or overt diabetes, %	9%	6.2%
CHD, %	6.16%	2.99%
Hypertension, %	33.54%	10.47%

We find that South Asians have lower body mass index and smoking levels, but lower activity levels. These three variables are key risk factors for metabolic and cardiovascular conditions.

Analysis & Results

Overview

We discuss the process of analyses and the results for each condition separately below. We run all the analysis with sampling weights as is standard for survey data. We check for multicollinearity by running a regular ordinary least squares regression and using the

vif command to check the variance inflation factor. All values are below 5, so we conclude there is no multicollinearity.

Diabetes

First, we run an unrestricted model that tests all covariates we theoretically determined as impactful. We ran a logit regression, but used to logistic command to give the results outputted as odds ratios for ease of interpreting analysis. We use body mass index instead of dummy variables for overweight and obese because standards for overweight and obese do not account for variability in different races. We realize the limitations of the bmi scale since South Asians might suffer from truncal obesity meaning they bmi is less informative, but due to lack of more specific information we use bmi[3].

```
. logistic DIB aian black hispanic otherasian otherrace southasian smoker bmi USborn
UScitizen age duration female inMidwest inNortheast inWest incomeGreater25 insured
novigorous [pweight = wtfa_sa]

Logistic regression                                     Number of obs   =      17273
                                                        Wald chi2(19)   =     805.28
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood =  -22412856                       Pseudo R2       =     0.1844

------------------------------------------------------------------------------
             |               Robust
         DIB | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   .4186198   .2670602    -1.36   0.172     .1198926    1.461663
       black |   1.814366   .2040539     5.30   0.000     1.455438    2.261808
    hispanic |   1.926214   .2566107     4.92   0.000     1.483568    2.500932
  otherasian |   1.787473   .4439454     2.34   0.019     1.098579    2.908359
   otherrace |   .2813762   .3015452    -1.18   0.237     .0344407     2.29881
  southasian |   2.583265   1.083342     2.26   0.024     1.135537    5.876742
      smoker |   1.187299   .0992022     2.05   0.040     1.007952    1.398558
         bmi |   1.113947   .0069805    17.22   0.000      1.10035    1.127713
      USborn |   .8250143   .1243259    -1.28   0.202     .6140295    1.108495
   UScitizen |   1.039847   .2042387     0.20   0.842     .7075925    1.528115
         age |   1.072471   .0035507    21.13   0.000     1.065534    1.079453
    duration |    1.66021   .5714783     1.47   0.141     .8455942    3.259598
      female |   .7974169   .0684835    -2.64   0.008     .6738804    .9436003
    inMidwest |   1.038452   .1124096     0.35   0.727     .8399357    1.283886
 inNortheast |   .7537834   .0946326    -2.25   0.024     .5893641    .9640719
      inWest |   .8631837   .0955398    -1.33   0.184     .6948477    1.072301
incomeGreater25 |   .8034815   .0705226    -2.49   0.013     .6764953    .9543044
     insured |   1.350467   .1626024     2.50   0.013     1.066585    1.709908
  novigorous |   1.385469   .1212316     3.73   0.000     1.167118     1.64467
       _cons |   .0000411   .0000174   -23.81   0.000     .0000179    .0000943
------------------------------------------------------------------------------

. test inMidwest inNorth inWe

 ( 1)  [DIB]inMidwest = 0
 ( 2)  [DIB]inNortheast = 0
 ( 3)  [DIB]inWest = 0

           chi2(  3) =    7.38
         Prob > chi2 =    0.0606

. test aian black hispanic otherasian otherrace southasian

 ( 1)  [DIB]aian = 0
 ( 2)  [DIB]black = 0
 ( 3)  [DIB]hispanic = 0
 ( 4)  [DIB]otherasian = 0
```

[3] Mohanty et. al.

```
 ( 5)  [DIB]otherrace = 0
 ( 6)  [DIB]southasian = 0

           chi2(  6) =   66.27
         Prob > chi2 =    0.0000
```

We cannot reject the there is a significant difference between a model including regions and a model that does not include regions. Therefore we exclude regions as potential explanatory variables. Similarly we find nativity status, duration in the United States, and birth in the United States and remove them as explanatory variables. We find that some races are not significant, but in order to capture the effect of ethnic South Asians against Non-Hispanic whites we keep all dummy variables for race.

```
. logistic DIB aian black hispanic otherasian otherrace southasian smoker bmi age female
incomeGreater25 insured novigorous[pweight = wtfa_sa]

Logistic regression                                     Number of obs   =      17300
                                                        Wald chi2(13)   =     805.79
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood =  -22474707                       Pseudo R2       =     0.1825

------------------------------------------------------------------------------
             |               Robust
         DIB | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   .4061639    .260613    -1.40   0.160     .1154867    1.428468
       black |   1.867462   .2052575     5.68   0.000     1.505546    2.316378
    hispanic |   1.999316   .2215178     6.25   0.000     1.609055     2.48423
  otherasian |   1.808983   .4220854     2.54   0.011     1.145053    2.857875
   otherrace |    .288761   .3075711    -1.17   0.244     .0358005    2.329099
  southasian |   2.777645   1.098962     2.58   0.010     1.279095    6.031851
      smoker |   1.183867   .0976572     2.05   0.041     1.007134    1.391613
         bmi |   1.114477   .0068663    17.59   0.000       1.1011    1.128016
         age |   1.072913    .003499    21.58   0.000     1.066077    1.079793
      female |    .798315   .0683858    -2.63   0.009      .674929    .9442575
incomeGreater25 |   .8036316   .0703527    -2.50   0.013     .6769238    .9540566
     insured |   1.347114   .1600962     2.51   0.012     1.067196    1.700452
  novigorous |   1.387164   .1208468     3.76   0.000     1.169427    1.645443
       _cons |   .0000533   .0000164   -31.98   0.000     .0000292    .0000975
------------------------------------------------------------------------------
```

The Wald chi square gives the probability we can reject the null hypothesis, that the covariates in the model are equal to zero (the model fits no better than having no variables). We can reject this because the p-value is 0.0000 meaning the model overall is statistically significant. We are particularly interested in the ethnic South Asian population so we look at interactions between included variables and being South Asian to see if there is any statistically significant effect. We find none of the coefficients to be statistically significant.

We find then that the final model as listed below, but the coefficients are all taken to the power of *e* because binary logit models give the log odds and above with have the odds already converted.

$$DIB = \beta_0 + \beta_1(aian) + \beta_2(black) + \beta_3(hispanic) + \beta_4(otherasian) + \beta_5(otherrace) + \beta_6(southasian) + \beta_7(smoker) + \beta_8(bmi) + \beta_9(age) + \beta_{10}(female) + \beta_{11}(incomeGreater25) + \beta_{12}(aian) + \beta_{13}(insured) + \beta_{14}(novigorous)$$

This is our final model because we can reject the null hypothesis that the model fits not better than all the coefficients being zero. Since we have robust standard errors and differing observations numbers across the full model and the restricted model we cannot run a likelihood ratio test (lrtest command). In lieu of the test we run a Wald Test checking to see if the omitted variables from the full model can be assumed to be zero.

```
. test inMidwest inNortheast inWest USborn UScitizen duration

 ( 1)  [DIB]inMidwest = 0
 ( 2)  [DIB]inNortheast = 0
 ( 3)  [DIB]inWest = 0
 ( 4)  [DIB]USborn = 0
 ( 5)  [DIB]UScitizen = 0
 ( 6)  [DIB]duration = 0

           chi2(  6) =    9.92
         Prob > chi2 =    0.1280
```

We find we cannot reject the hypothesis that they are equal to zero meaning the inclusion of the variables does not improve the model.

We find that South Asians are 2.77 times more likely to have diabetes than non-Hispanic white males (OR = 2.77 CI = 1.27, 6.03) while adjusting for smoking, bmi, no vigorous activity, age, insurance status, income, and gender. Continuous variables have odds ratios close to one, so age and bmi are as expected. Being a female decreases the odds compared to non-Hispanic male whites (the baseline) as does higher income. Insurance status increases the odds and this could be because those with insurance are more likely to go to the doctor and be diagnosed as opposed to those who visit infrequently. South Asians having higher odds of diabetes matches most studies reporting South Asians to be more likely to have diabetes[4].

Coronary Heart Disease (CHD)

The methodology for determining the correct model for coronary heart disease is exactly the same of that for diabetes. We start with the full model and narrow down which variables are significant. We reject that region has an effect and continue to narrow down the model. As mentioned above me maintain keeping all races because the goal is to see the effect on ethnic South Asians even though we find race has no effect on the self-report of CHD.

```
. logistic CHD aian black hispanic otherasian otherrace southasian smoker bmi USborn
UScitizen age duration female inMidwest inNortheast inWest incomeGreater25 insured
novigorous [pweight = wtfa_sa]

Logistic regression                               Number of obs   =      17268
                                                  Wald chi2(19)   =     560.19
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood =  -10637362                 Pseudo R2       =     0.2160

------------------------------------------------------------------------------
             |               Robust
         CHD | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   1.035298   1.087277     0.03   0.974     .1321684    8.109668
       black |    .964565   .1884335    -0.18   0.853     .6577249    1.414551
    hispanic |   .7941085   .1969656    -0.93   0.353     .4883748    1.291238
```

[4] Mohanty et. al.

```
     otherasian |   .7747711   .3661141    -0.54   0.589     .3068616     1.95616
      otherrace |   1.930852   2.159018     0.59   0.556     .2157556    17.27968
     southasian |   1.875911   1.067797     1.11   0.269     .6147459    5.724387
         smoker |   1.718209   .2189491     4.25   0.000      1.33847    2.205685
            bmi |    1.05193   .0099136     5.37   0.000     1.032678    1.071541
         USborn |   1.315344   .3383639     1.07   0.287     .7944617    2.177737
       UScitizen |   1.033477   .4796864     0.07   0.943     .4161189    2.566755
            age |   1.100359   .0057363    18.35   0.000     1.089173     1.11166
       duration |   1.575297   .9786496     0.73   0.464     .4661819    5.323159
         female |   .3678532   .0492578    -7.47   0.000     .2829395    .4782507
      inMidwest |   .9771796   .1600354    -0.14   0.888     .7088768    1.347032
    inNortheast |    .923414   .1685333    -0.44   0.662     .6457177    1.320536
         inWest |   .7873313   .1442834    -1.30   0.192     .5497554    1.127575
incomeGreater25 |   .7869925   .1088524    -1.73   0.083     .6001187    1.032058
        insured |   1.335135   .2866874     1.35   0.178     .8764978    2.033761
     novigorous |   1.392403   .1929238     2.39   0.017     1.061273     1.82685
          _cons |   .0000192   .0000142   -14.67   0.000     4.50e-06     .000082
------------------------------------------------------------------------------
```

```
. test inMidwest inNortheast inWest

 ( 1)  [CHD]inMidwest = 0
 ( 2)  [CHD]inNortheast = 0
 ( 3)  [CHD]inWest = 0

           chi2(  3) =    1.84
         Prob > chi2 =    0.6053
```

```
. test aian black hispanic otherasian otherrace southasian

 ( 1)  [CHD]aian = 0
 ( 2)  [CHD]black = 0
 ( 3)  [CHD]hispanic = 0
 ( 4)  [CHD]otherasian = 0
 ( 5)  [CHD]otherrace = 0
 ( 6)  [CHD]southasian = 0

           chi2(  6) =    3.40
         Prob > chi2 =    0.7577
```

After adjusting the model while including race we find that the odds of CHD are insignificant in relation to race.

```
.  logistic CHD aian black hispanic otherasian otherrace southasian smoker bmi age female
novigorous [pweight = wtfa_sa]


Logistic regression                                     Number of obs   =      31651
                                                        Wald chi2(11)   =    1825.17
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood =  -33828102                       Pseudo R2       =     0.2215

------------------------------------------------------------------------------
             |               Robust
         CHD | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   .9406644    .362703    -0.16   0.874     .4418013    2.002823
       black |   1.161481   .1078085     1.61   0.107     .9682861    1.393223
    hispanic |   .8414107      .0831    -1.75   0.080     .6933318    1.021116
  otherasian |   .6759309   .1297674    -2.04   0.041     .4639658    .9847335
   otherrace |   .4190376   .4937818    -0.74   0.460     .0416123    4.219728
  southasian |   1.521544   .5109699     1.25   0.211      .787828    2.938582
      smoker |   1.590387   .1042452     7.08   0.000      1.39865    1.808408
         bmi |     1.0558   .0050204    11.42   0.000     1.046006    1.065686
         age |   1.079462    .002365    34.90   0.000     1.074836    1.084107
      female |   .4400897   .0288973   -12.50   0.000     .3869451    .5005334
  novigorous |   1.657342   .1383574     6.05   0.000      1.40719    1.951964
       _cons |   .0001121   .0000239   -42.72   0.000     .0000739    .0001702
------------------------------------------------------------------------------
```

We examine potential interactions and find that South Asian smokers have significantly higher odds.

```
Logistic regression                                   Number of obs   =      31651
                                                      Wald chi2(12)   =    1825.81
                                                      Prob > chi2     =     0.0000
Log pseudolikelihood =  -33815049                     Pseudo R2       =     0.2218

------------------------------------------------------------------------------
             |               Robust
         CHD | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   .9409561   .3626413    -0.16   0.875     .4420986    2.002717
       black |    1.16065   .1077097     1.61   0.108     .9676282    1.392174
    hispanic |   .8405361   .0829886    -1.76   0.079     .6926515    1.019995
  otherasian |   .6750198   .1295695    -2.05   0.041     .4633712    .9833406
   otherrace |   .4199921   .4947734    -0.74   0.461      .041733     4.22671
  southasian |   .9142314   .4148005    -0.20   0.843     .3757104    2.224636
      smoker |   1.576843    .103486     6.94   0.000     1.386517    1.793295
         bmi |   1.055875   .0050195    11.44   0.000     1.046083    1.065759
         age |   1.079495   .0023677    34.88   0.000     1.074865    1.084146
      female |   .4407354   .0289684   -12.47   0.000     .3874634    .5013317
  novigorous |    1.65812   .1384786     6.05   0.000     1.407756    1.953009
  southSmoke |   3.808245   2.552524     1.99   0.046     1.023756     14.1662
       _cons |   .0001122   .0000239   -42.70   0.000     .0000739    .0001703
------------------------------------------------------------------------------
```

To confirm the variables omitted will not result in an improved regression we run a Wald Test and cannot reject that the variables have no effect. The test indicates that the coefficients can simultaneously equal zero.

```
. test USborn UScitizen duration inMidwest inNorth inWest income insured

 ( 1)  [CHD]USborn = 0
 ( 2)  [CHD]UScitizen = 0
 ( 3)  [CHD]duration = 0
 ( 4)  [CHD]inMidwest = 0
 ( 5)  [CHD]inNortheast = 0
 ( 6)  [CHD]inWest = 0
 ( 7)  [CHD]incomeGreater25 = 0
 ( 8)  [CHD]insured = 0

           chi2(  8) =    9.65
         Prob > chi2 =    0.2906
```

We find that South Asians have non-signifcantly lower odds of CHD (OR: 0.91, CI: 0.38, 2.22). Because the results are not significant we cannot confirm that South Asian are more likely to be diagnosed. We do find that South Asians who smoke are more likely to be diagnosed with CHD. This is an interesting finding considering South Asians overall smoke less than non-Hispanic whites (as seen through the summary table in Appendix I). We find that insurance status is not significant unlike with diabetes, which might be because CHD immediately effects one's life through chest pain and is less likely to go on unnoticed unlike diabetes which can silently harm a person and take time to be diagnosed[5].

[5] http://www.nhlbi.nih.gov/health/health-topics/topics/cad/

Hypertension

The methodology utilized for the last two conditions will be used to analyze hypertension. We find that region is significant in this model and maintain it while removing other covariates.

```
.  logistic HYP aian black hispanic otherasian otherrace southasian smoker age bmi USborn
UScitizen duration female incomeGreater25 insured  novigorous inNortheast inMidwest
inWest [pweight = wtfa_sa]

Logistic regression                                     Number of obs   =      17263
                                                        Wald chi2(19)   =    1881.13
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood =  -53684632                       Pseudo R2       =     0.1797

------------------------------------------------------------------------------
             |               Robust
         HYP | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   1.132423   .3799392     0.37   0.711      .586708    2.185724
       black |   1.630788   .1134792     7.03   0.000     1.422873    1.869084
    hispanic |   .9607013   .0875742    -0.44   0.660     .8035184    1.148632
  otherasian |   1.487753   .1983044     2.98   0.003     1.145706    1.931918
   otherrace |   1.053574   .5345471     0.10   0.918     .3897579     2.84797
  southasian |   .8199042   .2185762    -0.74   0.456     .4862319    1.382556
      smoker |    1.15648    .058252     2.89   0.004     1.047763    1.276478
         age |   1.066383   .0020706    33.10   0.000     1.062333     1.07045
         bmi |   1.101862   .0045867    23.30   0.000     1.092909    1.110889
      USborn |   1.101737   .1132286     0.94   0.346     .9007357    1.347591
   UScitizen |   1.375367   .1800545     2.43   0.015     1.064105    1.777676
    duration |   1.126908   .2260866     0.60   0.551     .7605286    1.669787
      female |   .7722118   .0383949    -5.20   0.000     .7005097    .8512531
incomeGreater25 |   .9199502    .049628    -1.55   0.122     .8276469    1.022548
     insured |   1.128905   .0803802     1.70   0.089     .9818617     1.29797
  novigorous |   1.281835      .0646     4.93   0.000     1.161273    1.414913
 inNortheast |    .696654   .0524369    -4.80   0.000     .6011011    .8073963
   inMidwest |   .8616069   .0554125    -2.32   0.021     .7595667    .9773552
      inWest |   .8216013   .0547604    -2.95   0.003     .7209877    .9362554
       _cons |   .0005763   .0001451   -29.63   0.000     .0003519    .0009438
------------------------------------------------------------------------------
. test inNortheast inMidwest inWest

 ( 1)  [HYP]inNortheast = 0
 ( 2)  [HYP]inMidwest = 0
 ( 3)  [HYP]inWest = 0

           chi2(  3) =   25.48
         Prob > chi2 =    0.0000

. logistic HYP aian black hispanic otherasian otherrace southasian smoker age bmi
UScitizenfemale novigorous inNortheast inMidwest inWest[pweight = wtfa_sa]

Logistic regression                                     Number of obs   =      31611
                                                        Wald chi2(15)   =    4275.43
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood = -1.048e+08                       Pseudo R2       =     0.2239

------------------------------------------------------------------------------
             |               Robust
         HYP | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        aian |   1.354547   .2617847     1.57   0.116     .9274441    1.978338
       black |   1.665422   .0807204    10.52   0.000     1.514495    1.831389
    hispanic |   .9647933   .0539279    -0.64   0.521     .8646805    1.076497
  otherasian |   1.238096   .1042581     2.54   0.011     1.049726    1.460268
   otherrace |   1.039277    .386598     0.10   0.918     .5012971    2.154605
  southasian |    .669626   .1319261    -2.04   0.042     .4551276     .985216
      smoker |    1.23758   .0434745     6.07   0.000     1.155238     1.32579
         age |   1.064774   .0012294    54.36   0.000     1.062367    1.067186
```

```
         bmi |    1.09963   .0032877    31.77   0.000     1.093205    1.106092
   UScitizen |   1.325859   .1009207     3.71   0.000     1.142105    1.539176
      female |   .8548238   .0295355    -4.54   0.000     .7988519    .9147175
  novigorous |    1.27459   .0470268     6.58   0.000     1.185673    1.370175
inNortheast |   .7069761   .0362022    -6.77   0.000     .6394656     .781614
  inMidwest |   .8561229   .0391911    -3.39   0.001     .7826549    .9364874
      inWest |   .8085657   .0373105    -4.60   0.000     .7386477    .8851018
       _cons |   .0008829   .0001201   -51.68   0.000     .0006763    .0011528
------------------------------------------------------------------------------
```

We find that South Asians have significantly lower odds of hypertension (OR: 0.67, CI: .46, .98). We checked to see if there were any valid interactions between South Asian and the associated risk factors, but none were statistically significant. We then ran a Wald Test to confirm the removed explanatory variables were in fact all zero.

```
. test insured income USborn duration

 ( 1)  [HYP]insured = 0
 ( 2)  [HYP]incomeGreater25 = 0
 ( 3)  [HYP]USborn = 0
 ( 4)  [HYP]duration = 0

           chi2(  4) =    5.58
         Prob > chi2 =    0.2328
```

In this analysis region is significant and this could be because of different obesity levels across regions within the United States impacting obesity or smoking levels. Hypertension could me more sensitive to variations in risk factors. We also find nativity status as significant which might be directly relevant to access of care or perceptions of care, in some cultures hypertension might be concerned a normal occurrence. Further analysis on understanding the difference between ethnic South Asians and Non-Hispanic whites is done in Appendix II for all conditions.

Discussion

We find ethnic South Asians have significantly higher odds of diabetes (OR: 2.77, CI: 1.27, 6.03) and significantly lower odds of hypertension (OR: 0.67, CI: .46, .98) and non-significantly lower odds of CHD (OR: 0.91, CI: 0.38, 2.22) compared to non-Hispanic whites even though they have lower bmi and lower smoking levels. We find that duration of stay in the United States does not have a significant effect on the odds of any of the three conditions. This might be because we have limited respondents who have stayed in the US for less than ten years (under 4% of the dataset). Our findings emulate a similar study using NHIS data[6].

The results from this data show that in the United States current ethnic south Asians despite lower mean bmi and lower smoking rates having higher rates of diabetes. Diabetes is a pressing issue that more time should be invested into to learn why South Asians are so susceptible to it. Furthermore, why South Asians have lower hypertension rates shed light into potential preventive methods for hypertension. The results differ from many studies in the UK, which show higher rates for all conditions. This could be

[6] Mohanty et. al.

because immigration patterns differed in the UK with large waves of South Asian immigrants before the 1960s where it steadily began in the United States. It could also be because the UK has a larger percentage of South Asians meaning more data as well as more variety within the South Asian population. Patterns of immigration would also impact the health of immigrants. The UK had a colonial relationship with South Asia, which could make immigration flow easier than the US. If the US had tighter control only healthier immigrants could come in.

Validity Assessment

Potential problems with our analysis are:

- Endogeniety: Endogeniety occurs when there is feedback between the dependent variable and independent. Those diagnosed with a condition might have higher activity levels, but when diagnosed levels would be different.
- Autocorrelation: Errors not captured in the equation could impact multiple variables.
- Heteroskedasticity: It could be problematic to assume all respondents have the same variance. It could be that people with other health problems have different variance compared to already healthy individuals. This could lead to inefficient estimators.
- Omitted Variable Bias: We are likely to have omitted variable bias because there are several characteristics that were excluded from our analysis and others than theoretically could offer more insight. Having a general doctor is found to be significant in a South Asian and non-Hispanic white analysis, but could not be included in the whole due to limited data. Variables for rural and urban were unavailable, but location impacts the availability and perceptions of care, which would in turn impact diagnosis.

This analysis probably does not have external validity beyond the United States. But the dataset is recent so it speaks to the current population and accurately reflects the US population as it is a national survey.

Appendix I

There were two initial files from the NHIS 2011 survey *personsx.dta* and *samadult.dta* (both converted into STATA files). We merged the two files. The *personx.dta* had many more observations because it's a file of including all adults in the household while sample adult is only one adult in the household. I dropped all observations not in both because I require variables from both datasets. This leaves me with 33,014 observations. In the dataset 78 do not have a listed race and are coded as missing which leaves 32936 observations.

```
. merge 1:m hhx fmx fpx using personsx.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                         68,861
        from master                          0  (_merge==1)
        from using                      68,861  (_merge==2)

    matched                             33,014  (_merge==3)
    -----------------------------------------

. drop if _merge == 1
(0 observations deleted)

. drop if _merge == 2
(68861 observations deleted)
```

Do – Files for data cleaning and original data files are available on request.

Appendix II

Since having a general doctor could not be included in the initial regression due to limited observations we run another logistic regression including only Non-Hispanic Whites and South Asians with the covariates included in the diabetes regression. We control for the effects of only the variables found significant and find that ethnic South Asians compared to non-Hispanic whites odds of diabetes is significantly higher in this analysis. We find that many of the variables statistically significant in the first regression are no longer significant.

```
. logistic DIB southasian smoker bmi age female incomeGreater25 insured novigorous gendoc
if white == 1 | southasian
>  == 1  [pweight = wtfa_sa]

Logistic regression                                     Number of obs   =       2177
                                                        Wald chi2(9)    =      73.40
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood = -1078099.4                       Pseudo R2       =     0.2312

------------------------------------------------------------------------------
             |               Robust
         DIB | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  southasian |   1.630153   1.743076     0.46   0.648     .2004763    13.25543
      smoker |   .8375322   .3487024    -0.43   0.670     .3703462    1.894066
         bmi |   1.173505   .0303033     6.20   0.000      1.11559    1.234427
         age |   1.078749   .0209527     3.90   0.000     1.038454    1.120607
      female |   .4659108   .2550451    -1.40   0.163     .1593474    1.362262
incomeGreater25 |   1.021835   .3747301     0.06   0.953     .4979963    2.096698
     insured |   .8591105   .3231991    -0.40   0.686     .4109844    1.795861
  novigorous |   1.894673   .8012628     1.51   0.131     .8271014    4.340202
      gendoc |   2.459426   .9585352     2.31   0.021     1.145743    5.279351
       _cons |   3.80e-06   5.93e-06    -7.99   0.000     1.78e-07     .000081
------------------------------------------------------------------------------



. logistic DIB southasian bmi age gendoc if white == 1 | southasian == 1  [pweight =
wtfa_sa]

Logistic regression                                     Number of obs   =       3372
                                                        Wald chi2(4)    =     143.05
                                                        Prob > chi2     =     0.0000
Log pseudolikelihood = -2887966.1                       Pseudo R2       =     0.1874

------------------------------------------------------------------------------
             |               Robust
         DIB | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  southasian |    4.88332   2.766542     2.80   0.005     1.608736    14.82332
         bmi |   1.120417   .0161055     7.91   0.000     1.089291    1.152432
         age |   1.068706   .0076124     9.33   0.000      1.05389    1.083731
      gendoc |   1.614279   .3844814     2.01   0.044     1.012146    2.574624
       _cons |   .0000403    .000026   -15.67   0.000     .0000114    .0001429
------------------------------------------------------------------------------
```

We run the same process for hypertension and CHD, which result in the two models below.

```
. logistic HYP southasian age bmi female novigorous inNortheast inMidwest inWest gendoc
if white == 1 | southasian =
> =1 [pweight = wtfa_sa]

Logistic regression                                     Number of obs   =       3343
```

```
                                                Wald chi2(9)    =     274.13
                                                Prob > chi2     =     0.0000
Log pseudolikelihood = -9516087.5               Pseudo R2       =     0.1636

------------------------------------------------------------------------------
             |               Robust
         HYP | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  southasian |   .6223984   .3114696    -0.95   0.343     .2333984    1.659736
         age |   1.055108   .0042036    13.46   0.000     1.046901    1.063379
         bmi |   1.089934   .0102768     9.13   0.000     1.069977    1.110263
      female |   .7617958   .0940125    -2.20   0.027     .5981258     .970252
  novigorous |   1.360599   .1754031     2.39   0.017     1.056809    1.751716
 inNortheast |   .4879305   .0953924    -3.67   0.000     .3326172    .7157663
   inMidwest |   .7179805   .1082837    -2.20   0.028     .5342407    .9649135
      inWest |    .894166   .1452789    -0.69   0.491     .6503091    1.229466
      gendoc |   1.349527    .173676     2.33   0.020     1.048666    1.736705
       _cons |   .0017166   .0006972   -15.68   0.000     .0007744    .0038052
------------------------------------------------------------------------------

. . logistic CHD southasian smoker age female gendoc if white == 1 | southasian == 1
[pweig
> ht = wtfa_sa]

Logistic regression                             Number of obs   =       3461
                                                Wald chi2(5)    =     179.80
                                                Prob > chi2     =     0.0000
Log pseudolikelihood = -2076566.9               Pseudo R2       =     0.2260

------------------------------------------------------------------------------
             |               Robust
         CHD | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  southasian |   1.703266   1.674303     0.54   0.588     .2480589    11.69526
      smoker |   2.349607   .6868997     2.92   0.003     1.324797    4.167169
         age |    1.08686   .0073996    12.23   0.000     1.072453     1.10146
      female |   .3665841   .1121197    -3.28   0.001     .2012953    .6675957
      gendoc |   1.714158   .4775546     1.93   0.053     .9929126    2.959312
       _cons |   .0002125   .0000944   -19.04   0.000      .000089    .0005075
------------------------------------------------------------------------------
```

Diabetes has significantly higher odds for South Asians while hypertension has non-significantly lower odds for South Asians and CHD has non-significantly higher odds compared to non-Hispanic Whites.