# Health, Risk, and Equity

#### Risk Heterogeneity between Black and White Individuals Insured through the ACA Marketplaces from 2018-2020

### Andrew S. Cistola, MPH

### About

Previous research have shown signficiant under-utilization of ambulatory care and signficiantly lower costs among black insured individuals 18-64 from 2014-2018. (<doi:10.1001/jamanetworkopen.2022.17383>) If these differences persist through the ACA risk adjustment program, under-utilization represents an arbitrage opportunity for insurance providers in the individual market. This would represent how a history of systematic racism and barriers to care could be used for cost-containment measures with the beenfits of overall lower health care costs transferred among the larger population. Health equity efforts should be focused on neutralizing arbitrage opporunties based on defacto discrimination and instead incentivize issuers to address disparities in access to care. With the addition of race and zip code to the EDGE server in 2025, CMS will have the ability to reform the ACA risk adjustmenr program to improve these disparities. The purpose of this study is to identify how the ACA Risk Adjustment program interacts with known health disparities in health care utilization among raical groups. This study uses publcily available data from the Medical Expenditure Panel Survey (MEPS) from Agency for Healthcare Research and Quality (AHRQ) (<https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp>) from 2018 to 2021 to identify differences in risk that can be attributed to racial group independent of other factors and whether the ACA risk adjustment program excacerbates these diferences. Evidence from this study can identify how racial data can be used by CMS to promote health equity within the ACA marketplaces.

#### Notes:

This analysis serves to fulfill part of the the requirements for the PhD in Public Health-Heath Services Research at the University of Florida. The author is also employed by Blue Cross Blue Shield of Florida.

#### Status:

In development

#### Reference:

github.com/andrewcistola/Health-Risk-and\_Equity

#### Updated:

2022-12-16 15:26:57.640581

### Import Results

The following files were downloaded from <https://meps.ahrq.gov/mepsweb/data_files/pufs/> and saved to a local database:

##### AHRQ MEPS 2020

Household Consolidated File (h224) Medical Conditions File (h222) Prescribed Medicines File (h220a) Dental Visits File (h220b) Other Medical Expenditures File (h220c) Hospital Inpatient Stays File (h220d) Emergency Room Visits File (h220e) Outpatient Department Visits File (h220f) Office-Based Medical Provider Visits File (h220g) Home Health Visits File (h220h) Home Health Visits File (h220h) Appendix - Condition to Event File (h220if1) Appendix - Prescritpion to Condition File (h220if2)

##### AHRQ MEPS 2019

Household Consolidated File (h216) Medical Conditions File (h214) Prescribed Medicines File (h213a) Dental Visits File (h213b) Other Medical Expenditures File (h213c) Hospital Inpatient Stays File (h213d) Emergency Room Visits File (h213e) Outpatient Department Visits File (h213f) Office-Based Medical Provider Visits File (h213g) Home Health Visits File (h213h) Appendix - Condition to Event File (h213if1) Appendix - Prescritpion to Condition File (h213if2)

##### AHRQ MEPS 2018

Household Consolidated File (h209) Medical Conditions File (h207) Prescribed Medicines File (h206a) Dental Visits File (h206b) Other Medical Expenditures File (h206c) Hospital Inpatient Stays File (h206d) Emergency Room Visits File (h206e) Outpatient Department Visits File (h206f) Office-Based Medical Provider Visits File (h206g) Home Health Visits File (h206h) Appendix - Condition to Event File (h206if1) Appendix - Prescritpion to Condition File (h206if2)

See <https://datatools.ahrq.gov/meps-hc#varexpLabel> for variable explorer.

### Data Cleaning Summary

Raw data was subset for the following conditions:

##### Households

Individuals 26-64 with marketplace coverage for full year

SELECT  
 2020 AS YEAR # Repeated for each year  
 , DUPERSID AS PERSON\_ID  
 , AGELAST AS AGE  
 , SEX  
 , RACETHX AS RACE  
 , POVCAT20 AS FPL\_GROUP  
 , POVLEV20 AS FPL\_PERCENT  
FROM h224 W # Repeated for each household year file  
WHERE  
 AGELAST > 18  
 AND AGELAST < 65  
 AND PRSTX20 = 1 # Variable name charges for each year  
 AND INSCOV20 = 1 # Variable name charges for each year

##### Events

Non-Dental events for year individual has marketpalce coverage

SELECT  
 2020 AS YEAR # Repeated for each year  
 , SQ.DUPERSID AS PERSON\_ID  
 , 'OUTPATIENT' AS SETTING  
 , F.EVNTIDX AS EVENT\_ID  
 , F.OPFPV20X + F.OPDPV20X AS PAID # Combined Doctor and facility payments from privtae insurers for settings that provided both (variable name changes each year)  
FROM (  
 SELECT DISTINCT Y.DUPERSID   
 FROM h224 Y # Repeated for each household year file  
 WHERE  
 Y.AGELAST > 18  
 AND Y.AGELAST < 65  
 AND Y.RACETHX IN (1, 2, 3, 4)  
 AND Y.PRSTX20 = 1 # Variable name charges for each year  
 AND Y.INSCOV20 = 1 # Variable name charges for each year  
 ) SQ  
LEFT JOIN h220f F # Repeated for each event file in year  
 ON SQ.DUPERSID = F.DUPERSID

Then paid amounts were summed by person and year and joined to household records (exclduing dental). The setting of the event for each condition was also collected for each event that documented a ICD10 code.

##### Conditions

Any for individual in same year as marketpalce coverage

SELECT  
 2020 AS YEAR # Repeated for each year  
 , SQ.DUPERSID AS PERSON\_ID  
 , Z.CONDIDX AS CONDITION\_ID  
 , Z.EVNTIDX AS EVENT\_ID  
FROM (  
 SELECT DISTINCT DUPERSID   
 FROM h224 # Repeated for each household year file  
 WHERE  
 AGELAST > 18  
 AND AGELAST < 65  
 AND RACETHX IN (1, 2, 3, 4)  
 AND PRSTX20 = 1 # Variable name charges for each year  
 AND INSCOV20 = 1 # Variable name charges for each year  
 ) SQ  
LEFT JOIN h220if1 Z # Repeated for each appendix file  
 ON SQ.DUPERSID = Z.DUPERSID

All distinct conditions were kept and joined to household records.

##### Final Analytical Data

### Data Preparation Summary

The following Columns were derived for this analysis:

VISITS - VISITS\_TOTAL, ER\_VISITS, HOME\_VISITS, INPATIENT\_VISITS, OFFICE\_VISITS, OUTPATIENT\_VISITS, RX\_VISITS  
PAID - PAID\_TOTAL, ER\_PAID, HOME\_PAID, INPATIENT\_PAID, OFFICE\_PAID, OUTPATIENT\_PAID, RX\_PAID  
ICD10 - ICD10\_TOTAL, ICD10 YES/NO (1/0)

##### Descriptive Statistics

The following statistics describe the population used for both analyses:

##### Research Question 1: Analytical File

##### Research Question 2: Analytical File

### Regression Modeling Result Summary

The following results were collected using R version 4.2.2 (2022-10-31 ucrt)

#### Regression Step 1: Import and Clean Data

Source: \_data//Race\_MEPS//alpha\_dev\_20221216152657//analytical\_Q1.csv

W (ID variables): PERSON\_ID X (Predictor variables): RACE\_GRP AGE SEX FPL\_PERCENT CONDITIONS Y (Outcome variables): PAID\_TOTAL ALWD\_TOTAL Z (Subgroup variables): YEAR

#### OLS on Log costs

##### OLS Assumption 0: Sampling (Random sample, observations > predictors, predictor is independent)

##### OLS Assumption 1: Specification (Relationship between predictor and outcome is linear)

##### OLS Assumption 2: Normality (Errors are normal with a mean = 0)

##### OLS Assumption 3: No Autocorrelation (Error terms are not correlated with each other)

##### OLS Assumption 4: Homoskedasticity (Error is even across observations)

##### OLS Assumption 5: No Colinearity (Predictors are not correlated with each other)

### Paid Amount: Visit Based Paid Amounts

#### OLS on Square Root costs

##### OLS Assumption 0: Sampling (Random sample, observations > predictors, predictor is independent)

##### OLS Assumption 1: Specification (Relationship between predictor and outcome is linear)

##### OLS Assumption 2: Normality (Errors are normal with a mean = 0)

##### OLS Assumption 3: No Autocorrelation (Error terms are not correlated with each other)

##### OLS Assumption 4: Homoskedasticity (Error is even across observations)

##### OLS Assumption 5: No Colinearity (Predictors are not correlated with each other)

#### Two Part Model (Logistic for non-zero, then Log link)

##### Two part model: logistic and poisson

### Allowed Amount: Insurance plus out of pocket

#### OLS on Log costs

##### OLS Assumption 0: Sampling (Random sample, observations > predictors, predictor is independent)

##### OLS Assumption 1: Specification (Relationship between predictor and outcome is linear)

##### OLS Assumption 2: Normality (Errors are normal with a mean = 0)

##### OLS Assumption 3: No Autocorrelation (Error terms are not correlated with each other)

##### OLS Assumption 4: Homoskedasticity (Error is even across observations)

##### OLS Assumption 5: No Colinearity (Predictors are not correlated with each other)

#### OLS on Square Root costs

##### OLS Assumption 0: Sampling (Random sample, observations > predictors, predictor is independent)

##### OLS Assumption 1: Specification (Relationship between predictor and outcome is linear)

##### OLS Assumption 2: Normality (Errors are normal with a mean = 0)

##### OLS Assumption 3: No Autocorrelation (Error terms are not correlated with each other)

##### OLS Assumption 4: Homoskedasticity (Error is even across observations)

##### OLS Assumption 5: No Colinearity (Predictors are not correlated with each other)

#### Two Part Model (Logistic for non-zero, then Log link)

##### Two part model: logistic and poisson

### Machine Learning Result Summary (ASIAN)

Various machine learning models were trained on a reference population and then used to predict values from a focus populaiton. The difference in predicted to actual values for the focus group then to reflects the impact of group identification. This is an adaptation of the Kitigawa-Oaxaca-Blinder method. The following results used the scikit-learn and keras libraries for Python version 3.9.13 (tags/v3.9.13:6de2ca5, May 17 2022, 16:36:42) [MSC v.1929 64 bit (AMD64)]

#### Machine Learning Step 1: Data Processing of Predictors and Outcomes

Source: \_data//Race\_MEPS//alpha\_dev\_20221216152657//analytical\_Q2.csv

W (ID variables): PERSON\_ID Z (Subgroup variables): YEAR

Reference group: Non-Hispanic White (RACETH == 2) Focus group: Not Non-Hispanic White (RACETH == 4)

#### Learn Step 2: Decomposition Paid amount Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 -0.010835 -0.770117 0.639617 -0.000826

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.94’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.622’, ‘R-squared + 0.1956’]

Variables Importances

0 SDOH\_FPL 0.584608 4 AGE 0.268133 1 SDOH\_EDUCATION 0.053495 2 SDOH\_MARITAL 0.041899 5 SEX 0.041778 3 SDOH\_FOOD 0.010087

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.2725’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.2895’, ‘R-squared + 0.5717’]

Variables Importances

0 AGE 0.100108 53 ICD10\_E78 0.086701 63 ICD10\_F32 0.062884 201 ICD10\_M19 0.055748 66 ICD10\_F41 0.035373 .. … … 126 ICD10\_I87 0.000000 84 ICD10\_G58 0.000000 203 ICD10\_M23 0.000000 237 ICD10\_N41 0.000000 254 ICD10\_O03 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.1566’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.4054’, ‘R-squared + 0.3279’]

Variables Importances

5 OFFICE\_VISITS 0.303700 7 RX\_VISITS 0.288776 0 AGE 0.227729 6 OUTPATIENT\_VISITS 0.067518 2 ER\_VISITS 0.049853 4 INPATIENT\_VISITS 0.030946 1 SEX 0.030440 3 HOME\_VISITS 0.001040

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.6179’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.9441’, ‘R-squared + 0.8038’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.688149 4 AGE 0.234533 1 SDOH\_EDUCATION 0.025410 2 SDOH\_MARITAL 0.021700 5 SEX 0.016075 3 SDOH\_FOOD 0.014133

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.465’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.097’, ‘R-squared + 0.2893’]

Variables Importances

53 ICD10\_E78 0.082688 63 ICD10\_F32 0.059815 201 ICD10\_M19 0.057758 204 ICD10\_M25 0.040978 0 AGE 0.040074 .. … … 170 ICD10\_K74 0.000000 171 ICD10\_K76 0.000000 172 ICD10\_K80 0.000000 174 ICD10\_K85 0.000000 188 ICD10\_L70 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.0222’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.5398’, ‘R-squared + 0.3005’]

Variables Importances

5 OFFICE\_VISITS 0.358868 7 RX\_VISITS 0.339052 6 OUTPATIENT\_VISITS 0.101971 0 AGE 0.082351 2 ER\_VISITS 0.057792 4 INPATIENT\_VISITS 0.053922 1 SEX 0.005232 3 HOME\_VISITS 0.000812

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.5384’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -1.0236’, ‘R-squared + 0.5905’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

3 SDOH\_FOOD 0.799767 2 SDOH\_MARITAL 0.097803 4 AGE 0.027528 0 SDOH\_FPL -0.001288 1 SDOH\_EDUCATION -0.018546 5 SEX -0.214214

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.3129’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.2491’, ‘R-squared + 0.0201’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.9314’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.6306’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.8223’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.7397’, ‘R-squared + 0.226’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.025987 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX -0.000000 0 SDOH\_FPL -0.001329

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 5.3585’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.2035’, ‘R-squared + 0.017’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.9314’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.6306’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.8223’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.7397’, ‘R-squared + 0.226’]

##### Multi-Layer Perceptron

Loss

0 3.612802 1 3.276658 2 3.001048 3 2.882109 4 2.921885 .. … 495 0.648976 496 0.696601 497 0.626267 498 0.601877 499 0.663583

[500 rows x 1 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.562’, ‘Non-White Predicted = 4.6521’, ‘Difference in Bs = -0.0497’, ‘Difference in Xs = -0.9099’, ‘R-squared + 0.8801’]

#### Learn Step 2: Decomposition Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 0.006754 -0.05776 0.275156 0.000343

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.3687’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0767’, ‘R-squared + 0.1987’]

Variables Importances

0 SDOH\_FPL 0.574317 4 AGE 0.273393 5 SEX 0.052735 1 SDOH\_EDUCATION 0.044669 2 SDOH\_MARITAL 0.042908 3 SDOH\_FOOD 0.011978

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.4953’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.2033’, ‘R-squared + 0.6796’]

Variables Importances

0 AGE 0.209664 133 ICD10\_J09 0.071211 214 ICD10\_M53 0.056722 66 ICD10\_F41 0.051369 43 ICD10\_E11 0.034681 .. … … 15 ICD10\_B49 0.000000 230 ICD10\_N19 0.000000 229 ICD10\_N18 0.000000 203 ICD10\_M23 0.000000 292 ICD10\_R68 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.3235’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0315’, ‘R-squared + 0.4674’]

Variables Importances

5 OFFICE\_VISITS 0.355758 7 RX\_VISITS 0.228209 0 AGE 0.208431 6 OUTPATIENT\_VISITS 0.081424 4 INPATIENT\_VISITS 0.054751 2 ER\_VISITS 0.042737 1 SEX 0.023552 3 HOME\_VISITS 0.005137

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.1268’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = -0.1652’, ‘R-squared + 0.7877’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.586552 4 AGE 0.326545 2 SDOH\_MARITAL 0.030766 1 SDOH\_EDUCATION 0.028003 5 SEX 0.023047 3 SDOH\_FOOD 0.005087

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.4279’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.1359’, ‘R-squared + 0.3261’]

Variables Importances

0 AGE 0.115948 133 ICD10\_J09 0.108863 66 ICD10\_F41 0.057747 43 ICD10\_E11 0.047580 72 ICD10\_F90 0.045197 .. … … 147 ICD10\_J98 0.000000 149 ICD10\_K02 0.000000 150 ICD10\_K04 0.000000 151 ICD10\_K05 0.000000 335 ICD10\_U07 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.334’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.042’, ‘R-squared + 0.4038’]

Variables Importances

5 OFFICE\_VISITS 0.405402 7 RX\_VISITS 0.246435 0 AGE 0.111153 6 OUTPATIENT\_VISITS 0.088422 4 INPATIENT\_VISITS 0.075675 2 ER\_VISITS 0.061196 3 HOME\_VISITS 0.007244 1 SEX 0.004473

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.1179’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = -0.1741’, ‘R-squared + 0.5833’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

2 SDOH\_MARITAL 0.190235 5 SEX 0.168737 3 SDOH\_FOOD 0.078411 4 AGE 0.023777 0 SDOH\_FPL 0.000144 1 SDOH\_EDUCATION -0.049788

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.4952’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.2032’, ‘R-squared + 0.033’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.2981’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0061’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.2981’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0061’, ‘R-squared + 0.2789’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.023030 0 SDOH\_FPL 0.000075 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX 0.000000

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.5408’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.2488’, ‘R-squared + 0.0275’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.2981’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0061’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.2981’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = 0.0061’, ‘R-squared + 0.2789’]

##### Multi-Layer Perceptron

Loss

0 2.783449 1 1.858967 2 1.646672 3 1.560276 4 1.500449 .. … 495 0.344992 496 0.372512 497 0.365971 498 0.332368 499 0.344926

[500 rows x 1 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.292’, ‘Non-White Predicted = 7.0869’, ‘Difference in Bs = 0.3955’, ‘Difference in Xs = -0.2051’, ‘R-squared + 0.8018’]

### Machine Learning Result Summary (BLAKC)

Various machine learning models were trained on a reference population and then used to predict values from a focus populaiton. The difference in predicted to actual values for the focus group then to reflects the impact of group identification. This is an adaptation of the Kitigawa-Oaxaca-Blinder method. The following results used the scikit-learn and keras libraries for Python version 3.9.13 (tags/v3.9.13:6de2ca5, May 17 2022, 16:36:42) [MSC v.1929 64 bit (AMD64)]

#### Machine Learning Step 1: Data Processing of Predictors and Outcomes

Source: \_data//Race\_MEPS//alpha\_dev\_20221216152657//analytical\_Q2.csv

W (ID variables): PERSON\_ID Z (Subgroup variables): YEAR

Reference group: Non-Hispanic White (RACETH == 2) Focus group: Not Non-Hispanic White (RACETH == 3)

#### Learn Step 2: Decomposition Paid amount Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 -0.010835 -0.770117 0.639617 -0.000826

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.5474’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.6667’, ‘R-squared + 0.1956’]

Variables Importances

0 SDOH\_FPL 0.582337 4 AGE 0.267517 1 SDOH\_EDUCATION 0.053870 5 SEX 0.043704 2 SDOH\_MARITAL 0.042331 3 SDOH\_FOOD 0.010241

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.6266’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.7459’, ‘R-squared + 0.5757’]

Variables Importances

0 AGE 0.102219 53 ICD10\_E78 0.086885 63 ICD10\_F32 0.062354 201 ICD10\_M19 0.053543 66 ICD10\_F41 0.035416 .. … … 305 ICD10\_S20 0.000000 307 ICD10\_S29 0.000000 126 ICD10\_I87 0.000000 50 ICD10\_E58 0.000000 227 ICD10\_M89 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.4462’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.5655’, ‘R-squared + 0.3308’]

Variables Importances

5 OFFICE\_VISITS 0.300218 7 RX\_VISITS 0.291120 0 AGE 0.228656 6 OUTPATIENT\_VISITS 0.068973 2 ER\_VISITS 0.048883 4 INPATIENT\_VISITS 0.031166 1 SEX 0.029973 3 HOME\_VISITS 0.001011

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.2448’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.3641’, ‘R-squared + 0.8039’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.688149 4 AGE 0.234533 1 SDOH\_EDUCATION 0.025410 2 SDOH\_MARITAL 0.021700 5 SEX 0.016075 3 SDOH\_FOOD 0.014133

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.6669’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.7861’, ‘R-squared + 0.2893’]

Variables Importances

53 ICD10\_E78 0.082688 63 ICD10\_F32 0.059815 201 ICD10\_M19 0.057758 204 ICD10\_M25 0.040978 0 AGE 0.040074 .. … … 170 ICD10\_K74 0.000000 171 ICD10\_K76 0.000000 172 ICD10\_K80 0.000000 174 ICD10\_K85 0.000000 188 ICD10\_L70 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.495’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.6142’, ‘R-squared + 0.3005’]

Variables Importances

5 OFFICE\_VISITS 0.358868 7 RX\_VISITS 0.339052 6 OUTPATIENT\_VISITS 0.101971 0 AGE 0.082351 2 ER\_VISITS 0.057792 4 INPATIENT\_VISITS 0.053922 1 SEX 0.005232 3 HOME\_VISITS 0.000812

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.1807’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.3’, ‘R-squared + 0.5905’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

3 SDOH\_FOOD 0.799767 2 SDOH\_MARITAL 0.097803 4 AGE 0.027528 0 SDOH\_FPL -0.001288 1 SDOH\_EDUCATION -0.018546 5 SEX -0.214214

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.7645’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.8838’, ‘R-squared + 0.0201’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.4575’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.5768’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.4426’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.5619’, ‘R-squared + 0.226’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.025987 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX -0.000000 0 SDOH\_FPL -0.001329

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.7066’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.8259’, ‘R-squared + 0.017’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.4575’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.5768’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.4426’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.5619’, ‘R-squared + 0.226’]

##### Multi-Layer Perceptron

Loss

0 3.316610 1 2.970459 2 2.872564 3 2.815425 4 2.691707 .. … 495 0.683407 496 0.638818 497 0.700528 498 0.647766 499 0.676767

[500 rows x 1 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 4.8807’, ‘Non-White Predicted = 5.676’, ‘Difference in Bs = 0.6316’, ‘Difference in Xs = 0.7953’, ‘R-squared + 0.8773’]

#### Learn Step 2: Decomposition Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 0.006754 -0.05776 0.275156 0.000343

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6463’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5255’, ‘R-squared + 0.1987’]

Variables Importances

0 SDOH\_FPL 0.574320 4 AGE 0.273728 5 SEX 0.051420 1 SDOH\_EDUCATION 0.044764 2 SDOH\_MARITAL 0.043763 3 SDOH\_FOOD 0.012005

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6724’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5516’, ‘R-squared + 0.6783’]

Variables Importances

0 AGE 0.209214 133 ICD10\_J09 0.070222 214 ICD10\_M53 0.057445 66 ICD10\_F41 0.051118 43 ICD10\_E11 0.033462 .. … … 91 ICD10\_H18 0.000000 15 ICD10\_B49 0.000000 179 ICD10\_L21 0.000000 305 ICD10\_S20 0.000000 289 ICD10\_R59 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.5535’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.4327’, ‘R-squared + 0.467’]

Variables Importances

5 OFFICE\_VISITS 0.355213 7 RX\_VISITS 0.227775 0 AGE 0.204959 6 OUTPATIENT\_VISITS 0.084678 4 INPATIENT\_VISITS 0.055482 2 ER\_VISITS 0.042545 1 SEX 0.024309 3 HOME\_VISITS 0.005040

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.5287’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.4079’, ‘R-squared + 0.7856’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.586552 4 AGE 0.326545 2 SDOH\_MARITAL 0.030766 1 SDOH\_EDUCATION 0.028003 5 SEX 0.023047 3 SDOH\_FOOD 0.005087

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.68’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5592’, ‘R-squared + 0.3261’]

Variables Importances

0 AGE 0.115948 133 ICD10\_J09 0.108863 66 ICD10\_F41 0.057747 43 ICD10\_E11 0.047580 72 ICD10\_F90 0.045197 .. … … 147 ICD10\_J98 0.000000 149 ICD10\_K02 0.000000 150 ICD10\_K04 0.000000 151 ICD10\_K05 0.000000 335 ICD10\_U07 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.5788’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.458’, ‘R-squared + 0.4038’]

Variables Importances

5 OFFICE\_VISITS 0.405402 7 RX\_VISITS 0.246435 0 AGE 0.111153 6 OUTPATIENT\_VISITS 0.088422 4 INPATIENT\_VISITS 0.075675 2 ER\_VISITS 0.061196 3 HOME\_VISITS 0.007244 1 SEX 0.004473

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.5294’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.4086’, ‘R-squared + 0.5833’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

2 SDOH\_MARITAL 0.190235 5 SEX 0.168737 3 SDOH\_FOOD 0.078411 4 AGE 0.023777 0 SDOH\_FPL 0.000144 1 SDOH\_EDUCATION -0.049788

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.7333’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.6125’, ‘R-squared + 0.033’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6028’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.482’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6774’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5566’, ‘R-squared + 0.2789’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.023030 0 SDOH\_FPL 0.000075 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX 0.000000

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.7017’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5809’, ‘R-squared + 0.0275’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6028’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.482’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.6774’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.5566’, ‘R-squared + 0.2789’]

##### Multi-Layer Perceptron

Loss

0 2.997121 1 1.681265 2 1.695459 3 1.642773 4 1.553380 .. … 495 0.382289 496 0.369685 497 0.369936 498 0.391471 499 0.366578

[500 rows x 1 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 7.1208’, ‘Non-White Predicted = 7.5273’, ‘Difference in Bs = 0.5667’, ‘Difference in Xs = 0.4065’, ‘R-squared + 0.7235’]

### Machine Learning Result Summary (HISPANIC)

Various machine learning models were trained on a reference population and then used to predict values from a focus populaiton. The difference in predicted to actual values for the focus group then to reflects the impact of group identification. This is an adaptation of the Kitigawa-Oaxaca-Blinder method. The following results used the scikit-learn and keras libraries for Python version 3.9.13 (tags/v3.9.13:6de2ca5, May 17 2022, 16:36:42) [MSC v.1929 64 bit (AMD64)]

#### Machine Learning Step 1: Data Processing of Predictors and Outcomes

Source: \_data//Race\_MEPS//alpha\_dev\_20221216152657//analytical\_Q2.csv

W (ID variables): PERSON\_ID Z (Subgroup variables): YEAR

Reference group: Non-Hispanic White (RACETH == 2) Focus group: Not Non-Hispanic White (RACETH == 1)

#### Learn Step 2: Decomposition Paid amount Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 -0.010835 -0.770117 0.639617 -0.000826

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.232’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.1252’, ‘R-squared + 0.1956’]

Variables Importances

0 SDOH\_FPL 0.581661 4 AGE 0.269192 1 SDOH\_EDUCATION 0.054670 5 SEX 0.042270 2 SDOH\_MARITAL 0.041793 3 SDOH\_FOOD 0.010415

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.4716’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.3648’, ‘R-squared + 0.5761’]

Variables Importances

0 AGE 0.099252 53 ICD10\_E78 0.088109 63 ICD10\_F32 0.062581 201 ICD10\_M19 0.054852 204 ICD10\_M25 0.035147 .. … … 227 ICD10\_M89 0.000000 282 ICD10\_R50 0.000000 191 ICD10\_L73 0.000000 229 ICD10\_N18 0.000000 70 ICD10\_F80 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.2426’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.1358’, ‘R-squared + 0.3266’]

Variables Importances

5 OFFICE\_VISITS 0.303378 7 RX\_VISITS 0.287594 0 AGE 0.227389 6 OUTPATIENT\_VISITS 0.069737 2 ER\_VISITS 0.048552 4 INPATIENT\_VISITS 0.031277 1 SEX 0.031076 3 HOME\_VISITS 0.000997

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 4.5607’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = -0.5461’, ‘R-squared + 0.8035’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.688149 4 AGE 0.234533 1 SDOH\_EDUCATION 0.025410 2 SDOH\_MARITAL 0.021700 5 SEX 0.016075 3 SDOH\_FOOD 0.014133

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.6446’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.5378’, ‘R-squared + 0.2893’]

Variables Importances

53 ICD10\_E78 0.082688 63 ICD10\_F32 0.059815 201 ICD10\_M19 0.057758 204 ICD10\_M25 0.040978 0 AGE 0.040074 .. … … 170 ICD10\_K74 0.000000 171 ICD10\_K76 0.000000 172 ICD10\_K80 0.000000 174 ICD10\_K85 0.000000 188 ICD10\_L70 0.000000

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.2996’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.1928’, ‘R-squared + 0.3005’]

Variables Importances

5 OFFICE\_VISITS 0.358868 7 RX\_VISITS 0.339052 6 OUTPATIENT\_VISITS 0.101971 0 AGE 0.082351 2 ER\_VISITS 0.057792 4 INPATIENT\_VISITS 0.053922 1 SEX 0.005232 3 HOME\_VISITS 0.000812

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 4.5702’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = -0.5366’, ‘R-squared + 0.5905’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

3 SDOH\_FOOD 0.799767 2 SDOH\_MARITAL 0.097803 4 AGE 0.027528 0 SDOH\_FPL -0.001288 1 SDOH\_EDUCATION -0.018546 5 SEX -0.214214

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.6471’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.5403’, ‘R-squared + 0.0201’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.1699’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.0631’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.1613’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.0545’, ‘R-squared + 0.226’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.025987 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX -0.000000 0 SDOH\_FPL -0.001329

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.6261’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.5193’, ‘R-squared + 0.017’]

Variables Coefficients

21 ICD10\_C50 2.055176 46 ICD10\_E34 1.411262 247 ICD10\_N83 1.256830 72 ICD10\_F90 1.210083 43 ICD10\_E11 1.180790 .. … … 133 ICD10\_J09 -0.739588 184 ICD10\_L50 -0.850364 273 ICD10\_R25 -0.893235 155 ICD10\_K25 -1.066067 198 ICD10\_M10 -1.119535

[336 rows x 2 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.1699’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.0631’, ‘R-squared + 0.2756’]

Variables Coefficients

4 INPATIENT\_VISITS 1.452032 2 ER\_VISITS 0.244232 7 RX\_VISITS 0.139614 6 OUTPATIENT\_VISITS 0.052833 3 HOME\_VISITS 0.039073 5 OFFICE\_VISITS 0.007957 0 AGE -0.000341 1 SEX -0.324354

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.1613’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.0545’, ‘R-squared + 0.226’]

##### Multi-Layer Perceptron

Loss

0 3.420915 1 2.937585 2 2.776081 3 2.932892 4 2.659459 .. … 495 0.695654 496 0.684668 497 0.753422 498 0.765759 499 0.669294

[500 rows x 1 columns]

[‘White Average = 5.5123’, ‘Non-White Average = 5.1068’, ‘Non-White Predicted = 5.3204’, ‘Difference in Bs = 0.4055’, ‘Difference in Xs = 0.2136’, ‘R-squared + 0.8549’]

#### Learn Step 2: Decomposition Using Machine Learning Models

##### Random Forests

AGE SEX CONDITIONS SDOH\_FPL

0 0.006754 -0.05776 0.275156 0.000343

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.4773’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5293’, ‘R-squared + 0.1987’]

Variables Importances

0 SDOH\_FPL 0.571695 4 AGE 0.272567 5 SEX 0.054543 1 SDOH\_EDUCATION 0.045507 2 SDOH\_MARITAL 0.043275 3 SDOH\_FOOD 0.012413

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5096’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5616’, ‘R-squared + 0.6784’]

Variables Importances

0 AGE 0.211733 133 ICD10\_J09 0.069133 214 ICD10\_M53 0.057685 66 ICD10\_F41 0.047845 43 ICD10\_E11 0.034469 .. … … 22 ICD10\_C55 0.000000 254 ICD10\_O03 0.000000 251 ICD10\_N93 0.000000 64 ICD10\_F34 0.000000 286 ICD10\_R55 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5116’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5636’, ‘R-squared + 0.47’]

Variables Importances

5 OFFICE\_VISITS 0.359242 7 RX\_VISITS 0.227093 0 AGE 0.203826 6 OUTPATIENT\_VISITS 0.083524 4 INPATIENT\_VISITS 0.054901 2 ER\_VISITS 0.043582 1 SEX 0.022657 3 HOME\_VISITS 0.005175

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.2965’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.3485’, ‘R-squared + 0.7855’]

##### Gradient Boosting

Variables Importances

0 SDOH\_FPL 0.586552 4 AGE 0.326545 2 SDOH\_MARITAL 0.030766 1 SDOH\_EDUCATION 0.028003 5 SEX 0.023047 3 SDOH\_FOOD 0.005087

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.4908’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5428’, ‘R-squared + 0.3261’]

Variables Importances

0 AGE 0.115948 133 ICD10\_J09 0.108863 66 ICD10\_F41 0.057747 43 ICD10\_E11 0.047580 72 ICD10\_F90 0.045197 .. … … 147 ICD10\_J98 0.000000 149 ICD10\_K02 0.000000 150 ICD10\_K04 0.000000 151 ICD10\_K05 0.000000 335 ICD10\_U07 0.000000

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.538’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.59’, ‘R-squared + 0.4038’]

Variables Importances

5 OFFICE\_VISITS 0.405402 7 RX\_VISITS 0.246435 0 AGE 0.111153 6 OUTPATIENT\_VISITS 0.088422 4 INPATIENT\_VISITS 0.075675 2 ER\_VISITS 0.061196 3 HOME\_VISITS 0.007244 1 SEX 0.004473

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.2926’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.3446’, ‘R-squared + 0.5833’]

##### Ridge Regression (with Cross Validation)

Variables Coefficients

2 SDOH\_MARITAL 0.190235 5 SEX 0.168737 3 SDOH\_FOOD 0.078411 4 AGE 0.023777 0 SDOH\_FPL 0.000144 1 SDOH\_EDUCATION -0.049788

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.6098’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.6618’, ‘R-squared + 0.033’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5056’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5576’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5065’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5585’, ‘R-squared + 0.2789’]

##### Least absolute shrinkage and selection operator

Variables Coefficients

4 AGE 0.023030 0 SDOH\_FPL 0.000075 1 SDOH\_EDUCATION -0.000000 2 SDOH\_MARITAL 0.000000 3 SDOH\_FOOD 0.000000 5 SEX 0.000000

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.6401’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.6921’, ‘R-squared + 0.0275’]

Variables Coefficients

21 ICD10\_C50 0.756602 43 ICD10\_E11 0.743104 72 ICD10\_F90 0.740985 247 ICD10\_N83 0.722187 284 ICD10\_R52 0.662059 .. … … 266 ICD10\_R12 -0.320411 36 ICD10\_D64 -0.323280 198 ICD10\_M10 -0.362099 214 ICD10\_M53 -0.511099 133 ICD10\_J09 -0.983439

[336 rows x 2 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5056’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5576’, ‘R-squared + 0.3079’]

Variables Coefficients

4 INPATIENT\_VISITS 0.720393 2 ER\_VISITS 0.107840 1 SEX 0.078017 7 RX\_VISITS 0.058982 6 OUTPATIENT\_VISITS 0.030091 3 HOME\_VISITS 0.020296 0 AGE 0.013602 5 OFFICE\_VISITS 0.005487

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.5065’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.5585’, ‘R-squared + 0.2789’]

##### Multi-Layer Perceptron

Loss

0 2.874324 1 1.763538 2 1.976484 3 1.636414 4 1.504555 .. … 495 0.358312 496 0.384264 497 0.340395 498 0.348731 499 0.342771

[500 rows x 1 columns]

[‘White Average = 7.6875’, ‘Non-White Average = 6.948’, ‘Non-White Predicted = 7.4465’, ‘Difference in Bs = 0.7395’, ‘Difference in Xs = 0.4985’, ‘R-squared + 0.7809’]