# 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:

2022-11-14 19:33:51.334478

### 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 or <https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_codebook.jsp?PUFId=H224> for the full varibale list.

### 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 > 25  
 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 > 25  
 AND Y.AGELAST < 65  
 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 > 25  
 AND AGELAST < 65  
 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\_20221114193351//analytical\_Q1.csv

W (ID variables): PERSON\_ID X (Predictor variables): NON\_WHITE AGE SEX FPL\_PERCENT ICD10\_TOTAL Y (Outcome variables): PAID\_TOTAL Z (Subgroup variables): YEAR

#### Regression Step 2: Test for OLS Assumptions

##### Results for Subgroup: 2018

##### 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)

#### Regression Step 2: Test for OLS Assumptions

##### Results for Subgroup: 2019

##### 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)

#### Regression Step 2: Test for OLS Assumptions

##### Results for Subgroup: 2020

##### 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)

#### Regression Step 3: Create Generalized Linear Models

##### Linear

Generalized model for DV = Y, regression = linear

##### Log Transform Y

Generalized model for DV = log(Y), regression = linear

##### Polynomial

Generalized model for DV = Y^2, regression = linear

##### Logistic

Generalized model for DV = Y > 0, regression = binomial

##### Poisson

Generalized model for DV = Y, regression = poisson

##### Negative Binomial

Generalized model for DV = Y, regression = negative binomial

#### Regression Step 4: Hierarchical Linear Models

##### Fixed Efects

Hierarchical model for DV = Y\_log regression = linear with varying intercepts by RACE

##### Random Efects

Hierarchical model for DV = Y\_log regression = linear with varying coeffeicints of ICD10\_TOTAL by RACE

### Machine Learning Result Summary

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)]

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

Source: \_data//Race\_MEPS//alpha\_dev\_20221114193351//analytical\_Q2.csv

W (ID variables): PERSON\_ID X (Predictor variables): RACE, AGE, SEX, ICD10\_TOTAL, ICD10\_YN, VISITS\_TOTAL Y (Outcome variables): PAID\_TOTAL Z (Subgroup variables): YEAR

Reference group: Non-Hispanic White (RACETH = 2) Focus group: Hispanic, Black, Asian, or Other (RACETH <> 2)

#### Learn Step 2: Manual Feature Selection Assisted with Unsupervised Learning

Unsupervised learning models are used to review predictors for inclusion in a regression model. The regression model is trained on the reference group and predicts values for the focus group. The difference in predicted to actual values represents what is explained by group identififcation independent of the predictors.

##### Principal Component Analysis

See \_fig//Race\_MEPS//alpha\_dev\_20221114193351//results.xlsx

##### K-Means

See \_fig//Race\_MEPS//alpha\_dev\_20221114193351//results.xlsx

##### Linear Regression using ACA Predictors and Visits

Regression Model using hand selected variables:

Absolute difference between groups: 0.37579456318369964 Difference attributable to groups: 0.37579456318369964

Regression Model using hand selected variables:

Absolute difference between groups: 0.37579456318369964 Difference attributable to groups: 0.37579456318369964

#### Learn Step 3: Automated Feature Selection Assisted with Supervised Learning

Supervised algorithms are used to automatically identify relevant features and predict outcomes. These models allow for the inclusion of more data in closer to raw form than OLS. The models are trained on the reference group and then predict values for the focus group. The difference in predicted to actual values represents what is explained by group identififcation independent of the predictors. For feature selection results, see \_fig//Race\_MEPS//alpha\_dev\_20221114193351//results.xlsx

##### Random Forests

Reference Group Rsq: 0.9940878249061551 Absolute difference between groups: 0.37579456318369964 Difference attributable to groups: 1.2547799151155603

##### Recursive feature Elimination

Reference Group Rsq: 0.6344018320582205 Absolute difference between groups: 0.37579456318369964 Difference attributable to groups: 0.36390643700851655

##### Support Vector Machines

Reference Group Rsq: 0.5360354970252728 Absolute difference between groups: 0.37579456318369964 Difference attributable to groups: 0.6621475761750837

#### Learn Step 4: Deep Learning with Expanded predictors

Deep learning algorithms are used for an expanded set of predictors in raw format. These models allow for virtually all structured data without processing and can handle complex interactions not yet understood. The models are trained on the reference group and then predict values for the focus group. The difference in predicted to actual values represents what is explained by group identififcation independent of the predictors.

For training results, see \_fig//Race\_MEPS//alpha\_dev\_20221114193351//results.xlsx

##### MLP Using ACA Data

Absolute difference between groups: 1.0951673001343485 Difference attributable to groups: 1.9656806325199438

##### MLP Using ACA and Diagnosis Data

Absolute difference between groups: 1.0951673001343485 Difference attributable to groups: -0.02307793147541748

##### MLP Using ACA, Diagnosis, and Office Visit Data

Absolute difference between groups: 1.7487125406269852 Difference attributable to groups: -1.710168221108268

##### MLP Using ACA, Diagnosis, and Hospital Visit Data

Absolute difference between groups: 2.6319321827268647 Difference attributable to groups: -1.4987375999913484

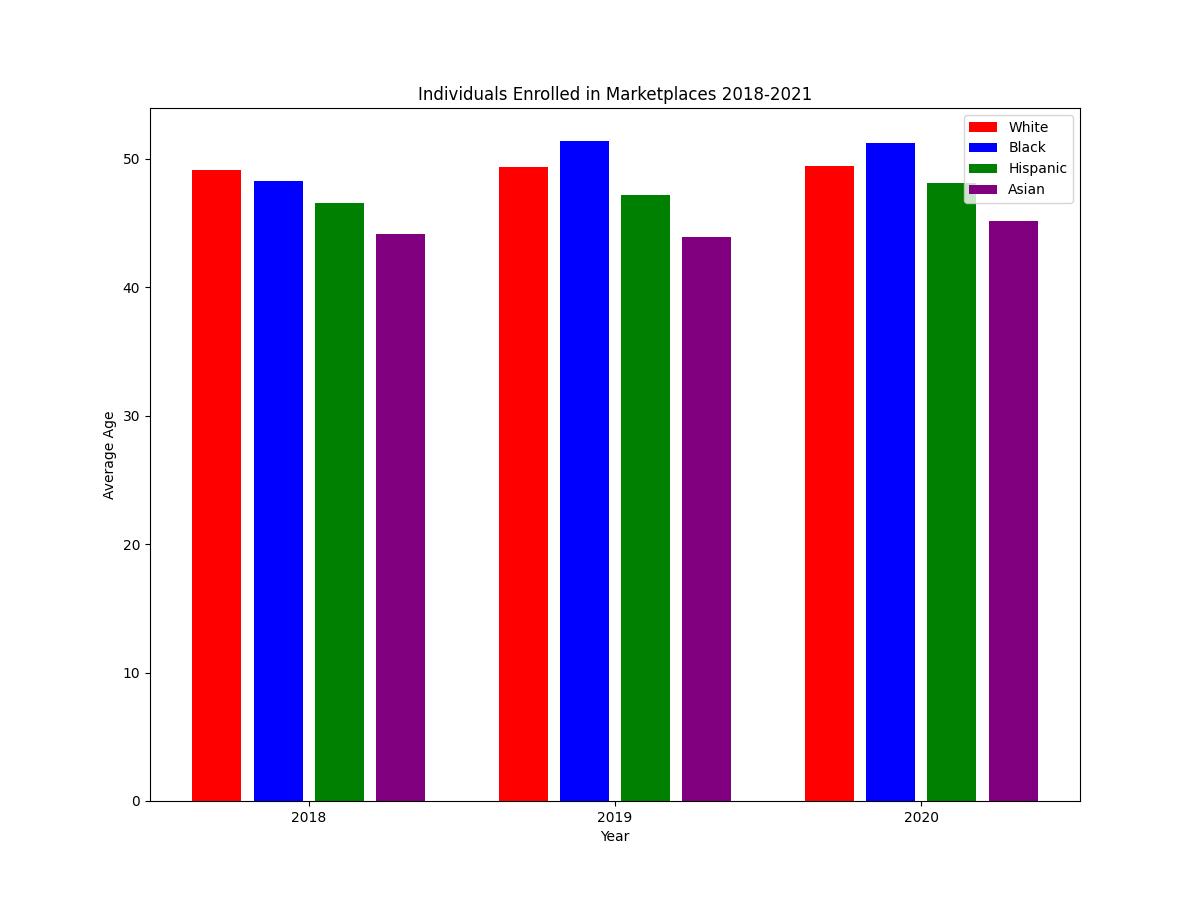
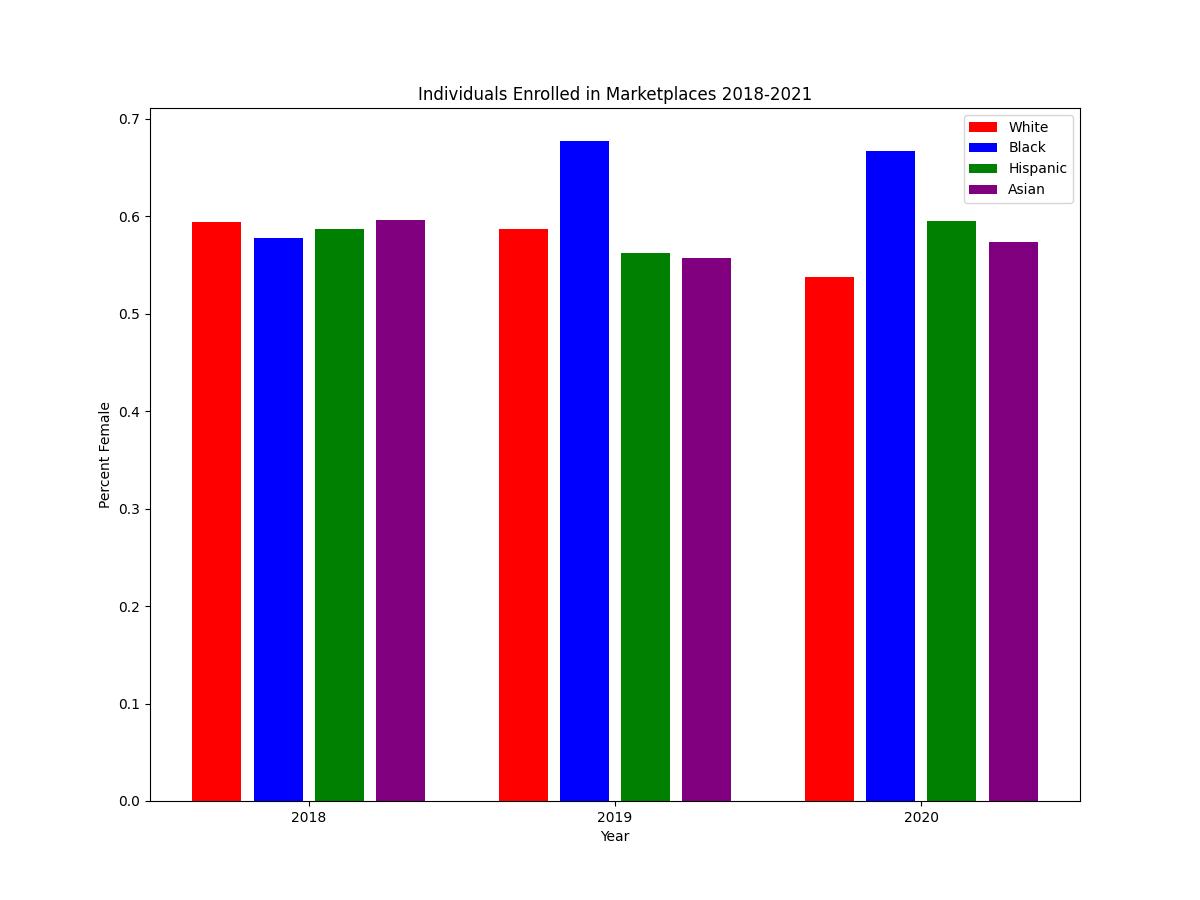
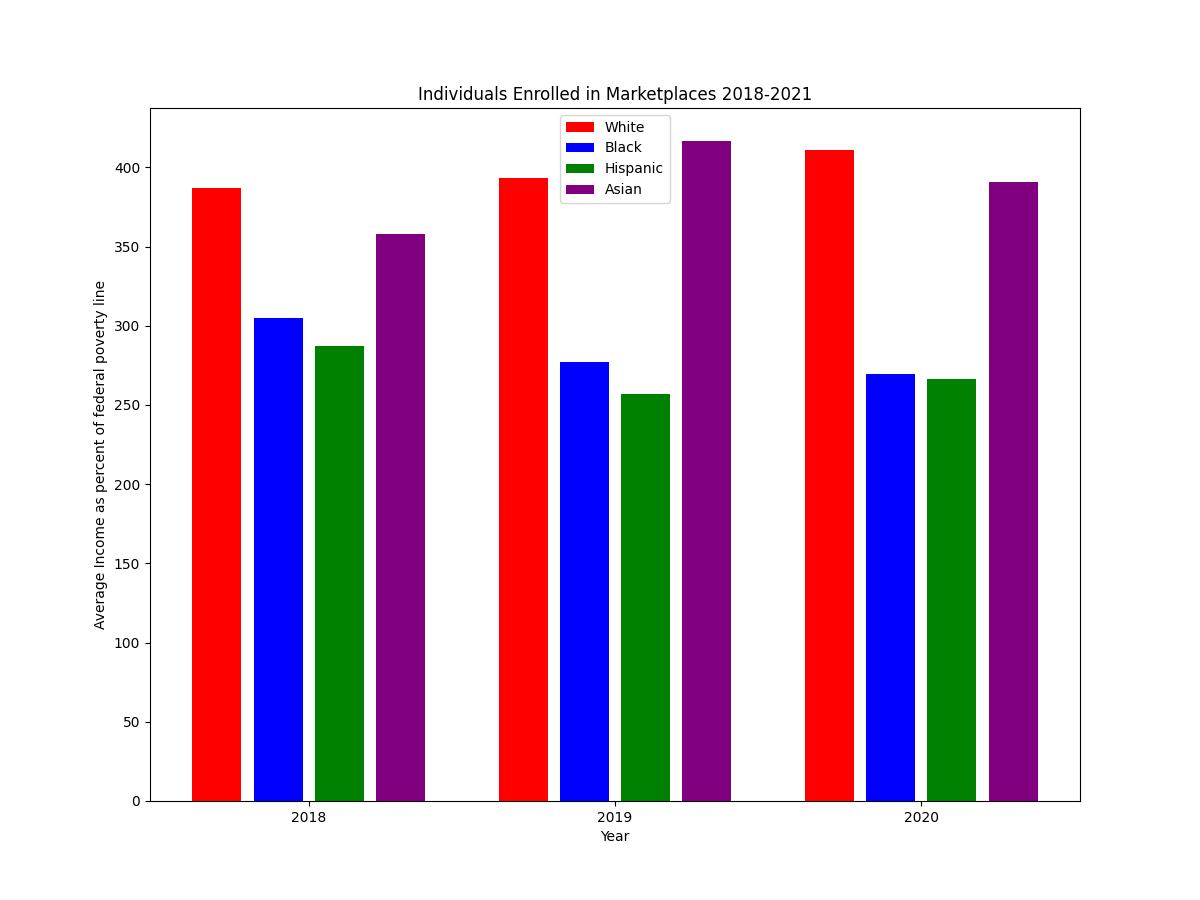
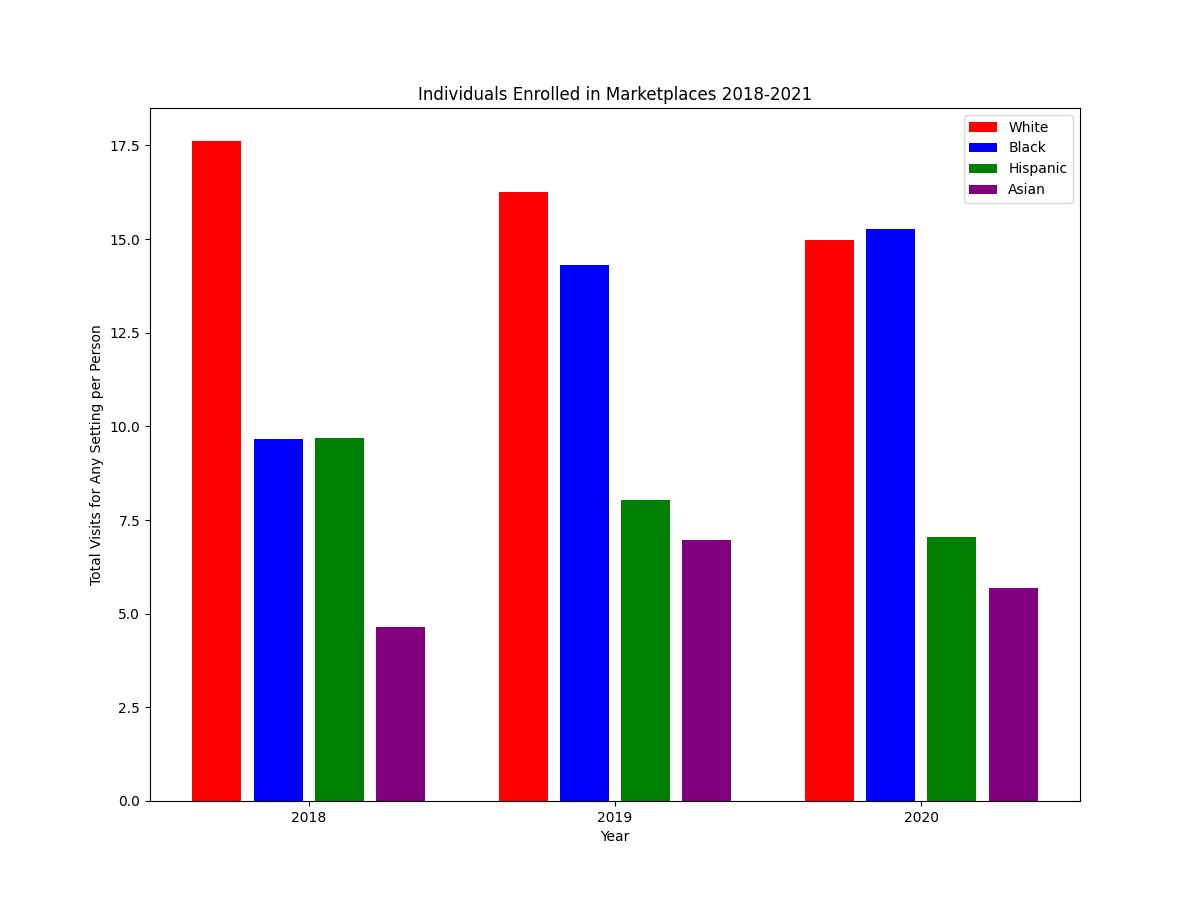
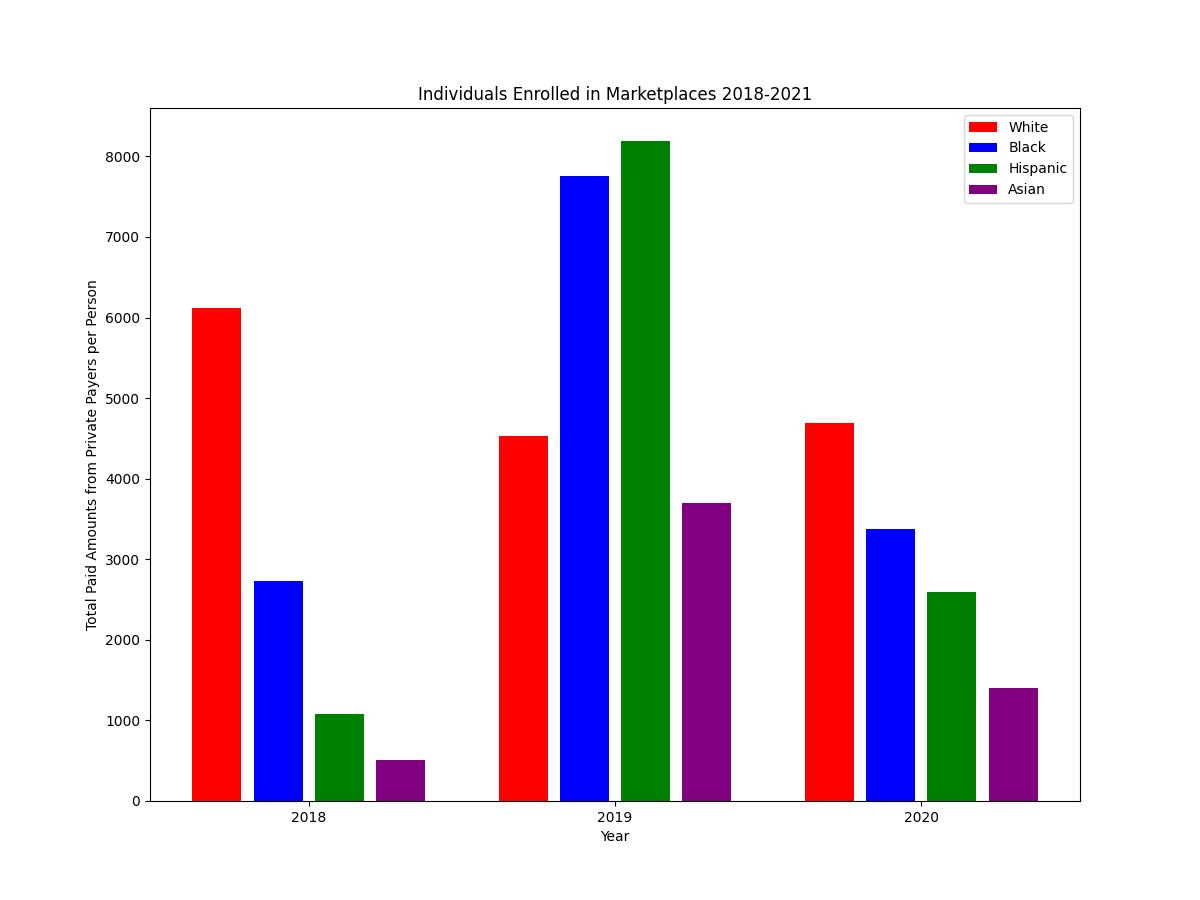
##### MLP Using ACA, Diagnosis, and ER Visit Data

Absolute difference between groups: 1.0013763205564947 Difference attributable to groups: -2.241712604400777

### Tables and Figures

Files can be found at: \_fig//Race\_MEPS//alpha\_dev\_20221114193351//

#### Descriptive Statistics

   Number of Diagnoses per Person.jpeg)  

#### Regression Results

2018 QQ Plot 2019 QQ Plot 2020 QQ Plot 2018 Residuals 2019 Residuals 2020 Residuals 2018 Colinearity 2019 Colinearity 2020 Colinearity