Whodunnit?: An Analysis of How Race And Gender Impact The Likelihood of Being Victim To Police Brutality

Introduction

Since the precipitation of a series of events in Ferguson, Missouri a few years back, a nebulous term has surfaced through both popular and academic discourse: police brutality. This term specifies the extent to which police officers are engaging in unsolicited or inappropriate usage of force, coercive behaviors, and abuse of powers as it pertains to innocent civilians. This surfacing of the term has arisen around an entire movement regarding reduction of systemic and social inequality that has run amok within our society. Indeed, entire organizations such as "Black Lives Matter" and "Project Zero" along with "Say Her Name" have sparked from the controversy of Ferguson and started a movement not only towards the reduction of police brutality, but engaging in a meaningful dialogue about how inequality pervades this brutality.

As other researchers have indicated and has been the topic of many of these social justice and activist movements, there is reason to believe that pervasive discriminatory forces (such as, but certainly not limited to, sexism and racism) are woven into the very fabric of our society. Under basic sociological theory, this interweaving, as a byproduct of history, therein shapes the very fundamental institutions as we know them today – even perhaps in the alleged absence of "overt" prejudice. Therefore, through this sociological lens, it is believed, and perhaps with good reason, that these forces are so pervasive that they permeate supposedly "objective" systems – specifically that of justice. By its construction, the justice system is believed to be a paragon of objectivity – delivering an abstract ideal free from worldly constructs. However, in light of sociological theory, this analysis may often be reductive in the best case and unambiguously deleterious in the worst.

This analysis therein leads to the belief that such a supposed rise in police brutality is related to a targeted form of brutality that is reflective of the wider discrimination within our society and our history such that those of marginalized groups (specifically women and minorities) are more so targeted by these engagements in violent interactions with police officers. As such, the research question examines a dataset of interactions with police officers, specifically in context of stop and frisk, in the year of 2015 in New York City. The present research examines four dependent variables – whether or not the encounter was explained, whether or not the encounter involved frisking, whether or not the encounter involved force, and whether or not the encounter involved arrest. Furthermore, the primary independent variables are the racial and demographic identities of the victim; the present research is attempting to assess if having a particular identity predicts the likelihood of a given encounter with the police resulting in any of these four outcomes (it is worth noting that while only the usage of force would formally classify as police brutality, the present research is broadening the scope to incorporate all discriminatory behaviors involve in police encounters).

Hypotheses

1) Because of potentially discriminatory police practices, one might expect that those being a minority would increase the probability of having one of these four police outcomes occurring in an encounter.
2) Because of potentially discriminatory police practices, one might expect that those being a woman would decrease the probability of having one of these four police outcomes occurring in an encounter out of a benevolently sexist motive to not harm the woman.
3) The interaction between gender and race may coalesce such that the potential bias observed in hypothesis one may be mitigated by the oppositional bias observed in hypothesis 2.

Data Source, Variables, and Coding

The dataset at hand for the present research is from a publically available database from the New York Police Department on the Stop, Question, and Frisk database. This data is a set that documents all of the street encounters between New York Police Department officers and citizens within New York City resulting in the completion of a NYPD form UF250 (Unified Form 250) during 2015. Therefore, this dataset is quite large in nature because it is a legally mandated documentation under the contract to the New York City Police Foundation. If we are to assume that all of the forms have been properly filled out and all of the instances that occurred were documented properly or at all, which admittedly may be a dubious assumption, one can treat this dataset almost as though it were the entirely population of the police encounters in New York in 2015. For the purposes of analysis, these will be treated as a sample because there still may be some collection and documentation bias – but one may be confident that this is at least a relatively robust dataset with regard to gathering practices.

The major variables here are the four outcomes indicative of mistreatment by police officers, the aforementioned demographic variables of interest as predicting these outcomes, and a series of relevant control variables that may impact the propensity to being treated unfairly or differently by the police officers.

Independent Variables

sex: This is one of the primary independent variables to test the hypothesis of how gender predicts interactions with the police. The variable has three values, M (Male), F (Female), and Z (Unknown). For the sake of interpretation, Z entries were removed.

race: This is one of the primary independent variables to test the hypothesis of how race predicts interactions with the police. The variable has eight levels, A (Asian / Pacific Islander), B (Black), I (American Indian / Alaskan Native), P (Black – Hispanic), Q (White – Hispanic), W (White), X

(Unknown), Z (Other). For the sake of interpretation, Z entries and X entries were removed. Additionally, because the argument is that this brutality stems from a racism based off of phenotypic traits (whether or not the victim "looks" a certain race), the categories of Black and Black – Hispanic were collapsed. Additionally, because the hypothesis is that of how minorities are treated as compared to those who are white, White is the reference category.

Dependent Variables

pforce – This is a binary variable is one of the primary dependent variables and documents whether or not a violent encounter occurred. The dataset incorporates a series of binary variables regarding the interaction with 1 meaning the event occurred and 0 meaning the event did not occur. These variables are pf_wall (suspect against a wall), pf_grnd (suspect on ground), pf_drwep (weapon drawn), pf_ptwep (weapon pointed), pf_baton (baton), pf_hcuff (handcuffs), pf_pepsp (pepper spray), pf_other (other). In order to reduce the number of variables present in the model, the pforce variable is created such that if any of these events occurred, this variable is coded as a 1. Because most reports include only one of the constituent components occurring and some of them are mutually exclusive, a reliability analysis was not conducted on the pforce variable.

arstmade – This is one of the primary dependent variables and documents whether or not an arrest was made. This is a binary variable such that 1 means yes and 0 means no.

explnstp -- This variable is one of the primary dependent variables and documents whether or not the victim was given a reason for being stopped. This is a binary variable such that 1 means yes and 0 means no.

frisked -- This variable is one of the primary dependent variables and documents whether or not the victim was frisked. This is a binary variable such that 1 means yes and 0 means no.

Control Variables

build – This is a categorical control variable describing what the physical build of the person who the cop is interacting with. One of the key components that is believed to be related to violent behaviors in police is the extent to which the person is believed or seen as threatening. In many instances, this could be related to body types insofar as those who are largely may naturally be viewed as more threatening. This is a categorical variable coded as H (Heavy), M (Medium), T (Thin), U (Muscular) and Z (Unknown). For the sake of interpretation, observations of Z were removed. Additionally, this is being treated as a categorical variable and not an ordinal variable because it is not necessarily clear which would be more threatening, heavy or muscular, and therefore these two categories cannot be properly ordered.

city -- This is a categorical variable of which of the five boroughs in the city did the event take place. There may be borough wide differences in police forces as well as crime rates that are worth controlling for in the model.

ac_incid – This is a binary variable which codes whether or not the police officer indicating if the event took place in a high crime area. This may additionally help control for crime rates in the area that would potentially affect police behavior. A value of 1 indicates yes and a value of 0 indicates no.

ac_time – This is a binary variable which codes whether or not the police officer indicating if the event took place at an abnormal time. This indicates whether or not the person was acting strangely given the time of day and therefore may have elicited more suspicion on behalf of the police office. A value of 1 indicates yes and a value of 0 indicates no.

armed – This binary variable indicates whether or not the person was armed with any form of weapon. There are six variables in the dataset that ask if the participant had a specific weapon such as pistol (a pistol), riflshot (a rifle), asltweap (an assault weapon), knifcuti (a knife), machgun (a machine gun), othrweap (other weapon). The variable armed is set to 1 (meaning the person was armed) is they had a value of 1 (meaning they were armed) for any of the above weapons. Similar to the calculation of pforce, because many of the victims only had 0 or 1 of the above categories, a reliability analysis was not conducted before collapsing this variable.

offunif --This is a binary variable which codes whether or not the police officer was in uniform. This may be important to control for as it may have impacted the other person's behavior and therefore changed the nature of the altercation. A value of 1 indicates yes and a value of 0 indicates no.

inout – this is a categorical variable which codes for whether or not the altercation took place I (inside) or O (outside). This may have an impact on the police officer's behavior and therefore was controlled for.

Data Analysis

Descriptive Information

All of the variables incorporated in this analysis are binary or categorical. Therefore, the data will be summarized with regard to frequencies as percentages of the sample that fall into a given category. Table 1 includes four subtables giving the percentage frequencies of the five categorical variables: sex, race, city, build, inout.

Table 1: Frequencies of Categorical Variables

Race

White	American Indian / Alaskan Native	Black	Asian / Pacific Islander	White - Hispanic
11.35	.35	60.33	4.98	22.99

It is already apparent that there is a vast difference in the racial dynamics of those who are incorporated in the sample. Specifically, one can see from this chart that there appears to be quite a large population of Black people interacting with the cops (60.33%) which is far above the overall population in New York City. The least represented group appears to be American Indian / Alaskan Native, comprising only .35% of the encounters.

Sex

Female	Male
6.77	93.23

The sample clearly has a heavy skew towards men as they comprise 93.23% of the sample, whereas women comprise only 6.77% of the sample.

City

Bronx	Queens	Staten Island	Manhattan	Brooklyn
21.07	25.34	7.96	17.47	28.16

Overall, there appears to be a relatively even distribution across the boroughs except for the smallest percentage coming from Staten Island (7.96%) and the largest coming from Brooklyn (28.16%) – although given that Staten Island has the smallest population of the boroughs and Brooklyn the largest, this is not surprising.

Build

Heavy	Medium	Muscular	Thin
9.75	49.43	.87	39.95

This distribution shows that the most common body type is Medium, comprising 49.43% of the sample. This is somewhat closely followed by Thin, which comprises 39.95% of the sample. The least represented group, perhaps surprisingly, is Muscular, with .87% of the sample.

Indoor / Outdoor

Inside	Outside
18.72	81.28

Unsurprisingly, quite a large amount of the instances occurred outside (81.28%), but a still somewhat decent amount (perhaps more than one might expect) occurred inside (18.72%).

Table 2 gives the frequencies by percentage of all of the binary variables in the sample. General discussion will be limited to the four dependent variables of interested due to how important they are for both the theory as well as the models run (because the number of successes and failures of the dependent variables is critical for the fitting of a logistic model and can cause trouble if either failure is too small).

Table 2: Frequencies of Binary Variables

Variable	Percentage Yes	Percentage No
Physical Force	33.17	66.83
Arrest Made	17.59	82.41
Reason Given	99.88	.12
Frisked	67.62	32.38
High Crime Area	45.99	54.01
Armed	4.82	95.18
Officer in Uniform	55.73	44.27

From this table, one can observe quite a few striking differences. There is a sizeable amount of instances where physical force was used (33.17%), but it is still a majority compared to those where it was not (66.83%). Likewise, while there were still a decent amount of arrests made (17.59%), many more did not result in arrest (82.41%). One of the most striking differences was the reasons given variable. Only .12% of interactions did not have a reason given, with an overwhelming 99.88% having one given. This difference is so large and the percentage in which this was not the case is so low that analysis of this dependent variable may prove difficult. Last, there were a majority in which the person was frisked (67.62%), but still quite a sizeable portion who were not (32.28%).

Initial Model

The initial analyses are intended as a means of testing the original hypotheses at face value. Specifically, we hypothesize that there may be a propensity towards engaging in violent behavior if one is not white, and a decreased propensity in engaging in violent behavior when the target is a women. The dependent variables in this case are binary. When there are a number of predictors and a binary outcome, there are a few possible models. One of these models is the Linear Probability Model, which treats the outcomes as essentially being linear and therefore the coefficients represented how much a given predictor changes the probability of an event occurring. However, this model may often violate some of the assumptions of linear regression, including homoscedasticity, normality of errors, and providing predictions that are out of the bounds of 0 and 1 – which are impossible values of probability. As such, it may be more advisable to use a method that transforms these linear probabilities into something bounded. In this instance, we may use a logistic regression, which involves a link function relating the linear coefficients to the probability of the event by transforming them into log odds coefficients. This model is frequently advising when working with a binary predictor and avoids many of the problems of the Linear Probability Model. As such, the following initial analyses were conducted using a series of logistic regressions. The tables 3 through 6 will be broken up into four tables for each of the four dependent variable with both sex and race included as predictors. Commentary will be interspersed throughout the tables. As mentioned previously, the race variable was created such that white is the reference category in order to assess if there are racial bias differences in the propensity to behave aggressively or engage in violently behavior towards those who are not white as compared with white people.

Table 3: Race and Sex Predicting Usage of Force

```
Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)    
(Intercept)                        -1.14471    0.07024 -16.297  < 2e-16 ***
raceAmerican Indian / Alaskan Native  0.74052    0.23577   3.141  0.00168 ** 
raceAsian / Pacific Islander        0.19030    0.08078   2.356  0.01848 *  
raceBlack                           0.44699    0.04944   9.041  < 2e-16 ***
raceWhite - Hispanic                0.41065    0.05471   7.506 6.12e-14 ***
sexMale                             0.06762    0.05787   1.169  0.24255    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28044  on 22073  degrees of freedom
Residual deviance: 27944  on 22068  degrees of freedom
  (489 observations deleted due to missingness)
AIC: 27956
```

This model has a series of significant results that will be discussed individually. First, because the reference category for race was White, all of the race categories are interpreted as the average difference in the log odds of being victim to a usage of force, holding the gender of the victim constant. In general, logistic regressions give a term that indicates the amount that the predictor impacts the log odds of the event occurring. For the following coefficient interpretations, it is implied that the increases discussed are relative to white people, as that is the reference category. In this instance, being American Indian increases the log odds of having forced used against you by .74 and the coefficient is significant at the .01 level. Similarly, being Asian American increases the log odds of having forced used against you by .19 and the coefficient is significant at the .05 level. Being Black increases the log odds by .45 and the coefficient is highly statistically significant. Being White Hispanic increase the log odds by .41 and that is also highly significant. Therefore, this model corroborated hypothesis 1 in that being a minority of any kind relative to a white person increases the log odds of having force used against you. It is worth noting that the coefficient are quite large for American Indians. Specifically, if we are to exponentiate the coefficient, we receive a value of 2.097 – which can be interpreted as the odds ratio between American Indians to White people. Therefore, being American Indian increases your odds of having forced used against you by 109.7%! It is worth noting that the sex is not significant in this scenario.

Table 4: Race and Sex Predicting An Arrest Being Made

```
Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)    
(Intercept)                        -1.43426    0.08100 -17.706  < 2e-16 ***
raceAmerican Indian / Alaskan Native -0.30700    0.37824  -0.812   0.4170    
raceAsian / Pacific Islander       -0.24020    0.11293  -2.127   0.0334 *  
raceBlack                           0.28979    0.06248   4.638 3.52e-06 ***
raceWhite - Hispanic                0.55159    0.06749   8.173 3.00e-16 ***
sexMale                            -0.44763    0.06349  -7.050 1.79e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20516  on 22073  degrees of freedom
Residual deviance: 20351  on 22068  degrees of freedom
  (489 observations deleted due to missingness)
AIC: 20363
```

For the sake of brevity, the interpretation and discussion of the coefficients for this model and the models hereafter will be more concise than above. In the model predicting the log odds of an arrest being made, being American Indian does not affect the log odds, being Asian American decreases the log odds by .24 and is statistically significant, and being White – Hispanic and Black both significantly increase the odds, all controlling for gender. This corroborates the hypothesis that there is discrimination against minorities, and furthermore delineates that this may specifically be targeting Hispanic and Black Americans. The coefficient for Male can be interpreted such that being a male relative to being a female on average lowers the log odds of being arrested by -.44 controlling for race and this is statistically significant. This does not corroborate hypothesis 2 as one might expect, but could be a result of the low amount of women in the sample – indicating that women are only approached by the police at all if the case is an extreme one.

Table 5: Race and Sex Predicting A Reason Being Given

```
Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                         7.31004    1.13035   6.467 9.99e-11 ***
raceAmerican Indian / Alaskan Native 11.84682 742.54927   0.016    0.987
raceAsian / Pacific Islander       -0.39987    0.91391  -0.438    0.662
raceBlack                          -0.05569    0.62669  -0.089    0.929
raceWhite - Hispanic                0.20402    0.73081   0.279    0.780
sexMale                            -0.62347    1.01976  -0.611    0.541
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 416.11  on 22073  degrees of freedom
Residual deviance: 414.90  on 22068  degrees of freedom
  (489 observations deleted due to missingness)
AIC: 426.9
```

As the descriptive table depicted, there was incredibly little variance in the reason being given variable. Therefore, unsurprisingly, none of the coefficients were statistically significant – indicating that race nor gender predicts being given a reason for being stopped.

Table 6: Race and Sex Predicting Being Frisked

```
Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)    
(Intercept)                         -0.78611    0.06497 -12.100   <2e-16 ***
raceAmerican Indian / Alaskan Native  0.28476    0.24078   1.183   0.2369    
raceAsian / Pacific Islander         0.15937    0.07408   2.151   0.0314 *  
raceBlack                            0.69556    0.04495  15.472   <2e-16 ***
raceWhite - Hispanic                 0.46968    0.05049   9.303   <2e-16 ***
sexMale                              1.07876    0.05485  19.667   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27777  on 22073  degrees of freedom
Residual deviance: 27107  on 22068  degrees of freedom
  (489 observations deleted due to missingness)
AIC: 27119
```

This model holds similar results to the previous models on the basis of race. Specifically, there is no significant effect for American Indians, being Asian American increases one’s log odds of being frisked, and being Black or Hispanic highly increases one’s log odds of being frisked and is highly statistically significant. Likewise, this corroborates hypothesis 1. Additionally, the sex coefficient indicates that being a male increases one’s log odds of being frisked by 1.08 controlling for gender and this is highly significant. If we are to exponentiate this coefficient, we receive a log odds ratio of 2.94, indicating that being a man relative to a woman results in the odds of being frisked increasing by 194% controlling for race! This does corroborate hypothesis two and may suggest evidence that there is a discomfort in randomly touching women that the officer does not know.

Models Including Controls

One of the major restrictions of the previous models is the extent to which they may be confounded by numerous factors as indicated in the variables description section. Specifically, it is worth considering how factors such as one’s build, the area they are in, if the area has a high crime rate, if the event occurred at a normal time, if the event occurred outside, if the suspect was armed and if the officer approached the suspect in a uniform may be related to the main independent variables (race and gender) in a meaningful way and therefore explain some of the coefficients above. For reasons about why each of those variables specifically was included theoretically as a control, please see the variables section. In this next analysis, we will include all of these variables into the previous models and examine if the new models corroborate any of the previous hypotheses and how the coefficients change in light of the incorporation of these specific control variables. It is worth noting due to the lack of any effects and small variance in the "Reasons Given" model, only the other three models will be examined with controls included. Tables 7 through 9 will summarize these results with commentary following each table.

Table 7: Race and Sex Predicting Usage of Force with Controls

```
Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -0.65109    0.09821  -6.630 3.36e-11 ***
raceAmerican Indian / Alaskan Native  0.67882    0.23925   2.837 0.004549 **
raceAsian / Pacific Islander       0.14190    0.08382   1.693 0.090485 .
raceBlack                          0.34897    0.05147   6.780 1.20e-11 ***
raceWhite - Hispanic               0.24753    0.05693   4.348 1.37e-05 ***
sexMale                            0.13246    0.05994   2.210 0.027104 *
buildMedium                       -0.19166    0.05110  -3.751 0.000176 ***
buildMuscular                      0.44119    0.15617   2.825 0.004727 **
buildThin                          0.01861    0.05172   0.360 0.718964
cityBrooklyn                      -0.69771    0.04280 -16.301  < 2e-16 ***
cityManhattan                     -0.20470    0.04592  -4.458 8.28e-06 ***
cityQueens                        -0.36205    0.04490  -8.063 7.45e-16 ***
cityStaten Island                 -0.78725    0.06625 -11.882  < 2e-16 ***
ac_incid                          -0.19843    0.03515  -5.645 1.65e-08 ***
ac_time                           -0.12901    0.03743  -3.447 0.000567 ***
armed                              0.86433    0.06517  13.262  < 2e-16 ***
offunif                            0.19289    0.03156   6.112 9.86e-10 ***
inoutOutside                      -0.03164    0.03883  -0.815 0.415197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27632  on 21761  degrees of freedom
Residual deviance: 26698  on 21744  degrees of freedom
  (801 observations deleted due to missingness)
AIC: 26734
```

As suspected, many of the control variables are statistically significant. This indicates some evidence that they also predict the outcome measures, and therefore including them in the model may have been a wise choice – especially due to the large reduction of Residual Deviance (this model has a lower Residual Deviance by about 1,000; Residual Deviance is akin to R-squared in regression and therefore this large reduction indicates a significant amount of deviance in the model has been explained by the incorporation of these terms; it is worth noting that a more rigorous test such as an Log-Likelihood Ratio test could be used to significantly determine if this model fits the data better). In terms of the coefficients, there is a similar effect to before such that being of any of the four minority race categories significantly (marginally significant in the case of Asian Americans) increases your log odds of being victim to a usage of force, net of all other factors. It is worth noting that the coefficients, while significant, appear to be quite a large amount lower for both Black and White – Hispanic people, with almost a reduction of the coefficient by half in the latter case. This reduction in coefficient could be evidence for a mediation of sorts such that one or multiple of the control variables may be a mediator in the relationship between race and likelihood of being a victim of force (one possible example could be your racial identity could predict which borough you live in which could predict whether or not force is being used). More rigorous testing would be necessary in order to sufficiently corroborate this hypothesis. Additionally, because the coefficients for race are all still significant, there is still some evidence for hypothesis 1. The coefficient for Male is significant and indicates that being a male increases your log odds net of all other factors of being a victim of force by .13 and this is statistically significant at the .05 level. This provides evidence for hypothesis two insofar as force is being more directed towards men than women. Additionally, it is worth noting that this coefficient was not significant when controls were included – indicating that these other factors may have been confounding this relationship.

Table 8: Race and Sex Predicting an Arrest Being Made with Controls

```
Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                         0.086463   0.122826   0.704 0.481467
raceAmerican Indian / Alaskan Native -0.336938  0.393770  -0.856 0.392180
raceAsian / Pacific Islander        -0.188134   0.124197  -1.515 0.129824
raceBlack                           0.026623   0.069801   0.381 0.702896
raceWhite - Hispanic                0.223993   0.075584   2.963 0.003042 **
sexMale                             -0.334337  0.070525  -4.741 2.13e-06 ***
buildMedium                         -0.085050  0.067883  -1.253 0.210246
buildMuscular                       -0.042621  0.211834  -0.201 0.840543
buildThin                           0.051288   0.068394   0.750 0.453314
cityBrooklyn                        -0.960802  0.053874 -17.834  < 2e-16 ***
cityManhattan                       -0.626546  0.056626 -11.065  < 2e-16 ***
cityQueens                          -0.842012  0.059043 -14.261  < 2e-16 ***
cityStaten Island                   -1.536296  0.104422 -14.712  < 2e-16 ***
ac_incid                            0.084226   0.045754   1.841 0.065641 .
ac_time                             -0.161660  0.048902  -3.306 0.000947 ***
armed                               2.569513   0.073549  34.936  < 2e-16 ***
offunif                             -0.004052  0.042616  -0.095 0.924251
inoutOutside                        -1.198131  0.044672 -26.820  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20244  on 21761  degrees of freedom
Residual deviance: 17208  on 21744  degrees of freedom
  (801 observations deleted due to missingness)
AIC: 17244
```

This model does not have significant coefficients for any of the minority groups except White – Hispanic people such that being White – Hispanic increases the log odds relative to being white and net of all other factors by .22 and this is significant at the .01 level. This does provide small support for hypothesis one, although the lack of significant coefficients for any other racial group does not support this hypothesis. Likewise, the coefficient for Male is still significant, roughly the same effect size, and in the same direction as the model without controls – providing no evidence to support hypothesis 2.

Table 9: Race and Sex Predicting Being Frisked with Controls

```
Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)                      -0.35614    0.09691  -3.675 0.000238 ***
raceAmerican Indian / Alaskan Native  0.25674    0.24584   1.044 0.296338    
raceAsian / Pacific Islander      0.15867    0.07753   2.047 0.040701 *  
raceBlack                         0.65753    0.04715  13.944  < 2e-16 ***
raceWhite - Hispanic              0.42832    0.05288   8.100 5.49e-16 ***
sexMale                           0.98534    0.05693  17.308  < 2e-16 ***
buildMedium                      -0.05332    0.05285  -1.009 0.312982    
buildMuscular                     0.18539    0.17592   1.054 0.291971    
buildThin                         0.05876    0.05408   1.087 0.277174    
cityBrooklyn                     -0.30941    0.04605  -6.719 1.83e-11 ***
cityManhattan                    -0.45841    0.05034  -9.105  < 2e-16 ***
cityQueens                       -0.56148    0.04859 -11.556  < 2e-16 ***
cityStaten Island                -0.75114    0.06340 -11.848  < 2e-16 ***
ac_incid                         -0.06703    0.03586  -1.870 0.061541 .  
ac_time                           0.12931    0.03794   3.408 0.000654 ***
armed                             1.70689    0.11461  14.893  < 2e-16 ***
offunif                          -0.46247    0.03244 -14.258  < 2e-16 ***
inoutOutside                      0.33605    0.03956   8.494  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27372  on 21761  degrees of freedom
Residual deviance: 25877  on 21744  degrees of freedom
  (801 observations deleted due to missingness)
AIC: 25913
```

One gain, this model is quite similar to the model without controls such that there is no significant effect for American Indians, a small effect for Asian Americans, large effects for Black Americans and White – Hispanic Americans, and a large effect for Men net of all other factors. This provides support for both hypotheses 1 and 2. It is worth noting that it does appear that there was a major reduction in any of the main coefficients from the model without controls.

Interaction

Another major failing of the previous models is the lack of recognition of the extent to which various identities may intersect with one another – a huge concept in current activism against police brutality. Therefore, it may be important to create an interaction term between race and gender in order to assess if there are major differences in violence and if including such a term improves the model. Hypothesis three posited that the increase in likelihood posited by hypothesis 1 for minorities may be mitigated by the decrease in likelihood posited by hypothesis 2 for women when an interaction term is considered in the model. In order to test this, we are limiting our analyses to models where there was an observed increased for minorities and an observed decrease for women in the likelihood of being victim to violent actions. This leaves both the models for being a victim of force as well as the model of the likelihood to be frisked. For the sake of brevity and because the model for likelihood to be frisked has larger coefficients, only the latter model will be analyzed again with the incorporation of an interaction term. Table 10 includes this interaction with commentary to follow.

Table 10: The Interaction of Race and Sex Predicting Being Risked with Controls

```
Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                               -0.210088   0.163930  -1.282 0.199993
raceAmerican Indian / Alaskan Native       0.557228   0.939925   0.593 0.553286
raceAsian / Pacific Islander               0.169097   0.335953   0.503 0.614729
raceBlack                                  0.459025   0.163217   2.812 0.004918 **
raceWhite - Hispanic                       0.317633   0.187400   1.695 0.090086 .
sexMale                                    0.828644   0.153572   5.396 6.82e-08 ***
buildMedium                               -0.054664   0.052874  -1.034 0.301210
buildMuscular                              0.181091   0.175998   1.029 0.303509
buildThin                                  0.056273   0.054119   1.040 0.298433
cityBrooklyn                              -0.309624   0.046070  -6.721 1.81e-11 ***
cityManhattan                             -0.458971   0.050362  -9.113  < 2e-16 ***
cityQueens                                -0.561112   0.048600 -11.546  < 2e-16 ***
cityStaten Island                         -0.752048   0.063397 -11.862  < 2e-16 ***
ac_incid                                  -0.067256   0.035862  -1.875 0.060737 .
ac_time                                    0.129769   0.037949   3.420 0.000627 ***
armed                                      1.709738   0.114642  14.914  < 2e-16 ***
offunif                                   -0.461897   0.032449 -14.234  < 2e-16 ***
inoutOutside                               0.335810   0.039581   8.484  < 2e-16 ***
raceAmerican Indian / Alaskan Native:sexMale -0.318610   0.973230  -0.327 0.743385
raceAsian / Pacific Islander:sexMale      -0.005284   0.344380  -0.015 0.987759
raceBlack:sexMale                          0.215580   0.169803   1.270 0.204230
raceWhite - Hispanic:sexMale               0.120198   0.194595   0.618 0.536784
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27372  on 21761  degrees of freedom
Residual deviance: 25875  on 21740  degrees of freedom
  (801 observations deleted due to missingness)
AIC: 25919
```

The model has a number of interesting aspects to it. First, the original coefficients for race have been reduced in significant (resulting in being Asian American not having a significant effect and White – Hispanic becoming only marginally significant). When an interaction term is included in the model, it fundamentally changes the interpretation of the constituent coefficients, making changes in those coefficients not particularly surprising. Additionally, hypothesis three was not corroborated because none of the interaction terms are significant. It is worth noting that the incorporation of the interaction terms only reduced the residual deviance by 2, which is quite a small amount, and therefore not only are the individual interaction terms not significant, the incorporation of the interaction does not significantly fit the data better.

Results and Discussion

The overall results of the analyses provided a mixed series of evidence for the original hypotheses presented in the beginning of the paper. Overall, the research was not particularly interested in the ability to predict being a victim to various forms of violence, and therefore general estimates of how well the model fit the data is not necessarily of interest here. Rather, the research is focused on testing the theory to understand specifics about how race and gender impact the propensity to being a victim of various forms of violence and if the claims of police brutality being specifically targeted towards particular marginalized groups is true.

On the basis of race, there appears to be a strong amount of evidence that minority groups are being treated differently than White Americans. In all of the models (except for the Reason Given model which showed no effects), there was at least one minority group which was significantly being treated differently than White Americans in a manner that increased the log odds (and therefore, by extension of transforming the log odds through the aforementioned link function, the probability as well) of being a victim of a form of violence across the three dependent variables. There are perhaps some mixed cases of evidence about which specific minority groups are being targeted and in what manner they are being targeted (for example the coefficients for Black Americans was not always significant when the prominent controls are added), but in terms of corroborating the overall hypothesis of a general racial bias towards non-white people, the models provided evidence for this. What's more is often that the justification for instances of police brutality involves the suspect being armed – but these effects persisted even when including being armed in the model!

On the basis of sex, the evidence appeared to be more mixed. The hypothesis posited that out of benevolent sexist tendencies, women would be less likely to be victim to aggressive acts by the police. When including all of the control variables into the model, there was evidence for this for both being victim to usage of force and to being frisked such that net of all other factors, being a man increased your log odds and therefore probability of being victim to both of those behaviors. However, in both the model with and without controls, the evidence significantly showed, with a decent sized coefficient in both models, that being a woman increased the likelihood that one would be arrested! One potential explanation for this effect may be the relative frequency of men to women in the sample. In the sample very few women comparatively are involved in interactions with police. As such, it may be the case that the interactions that do occur are more extreme in nature or involve cases that may necessitate an arrest, but perhaps not violence – especially if women are less likely to resist arrest than men. Further research is necessary, however, to corroborate this potential explanation.

Model Specification and Limitations

A final consideration of this research is the usage of the logistic regression model in order to evaluate the relationships above. Given the binary nature of the dependent variables and the wide array of predictors, the usage of a logistic regression may seem like the most logical choice. However, while being distinctive from the linear regression and the aforementioned Linear Probability Model, there are still a number of assumptions to consider in the context of the logistic regression.

One of the critical assumptions is the predictors are independent. This assumption, if violated, could result in multicollinearity, an inflation in variance, and unstable estimates of the coefficients. In this particular model, no rigorous testing was done (especially with the control variables) to examine correlations between the predictors. Part of the reason for this choice was the fact that none of the variables were continuous, and therefore (while obviously not impossible), assessing the inter-correlations and the extent to which the variance may be inflated by the potential lack of independence is quite difficult. As such, this may result in some of the estimates' widening standard error when the control variables were incorporated in the model. Another assumption is that the observations are independent. Theoretically, one could assume that the observations of any given police incidence is independent – but given the geography and density of New York, it is more than possible and likely that any number of these cases in the sample were not independent, which may have impacted the standard errors and resulted likewise in unstable estimates. One other critical assumption is that of linearity between the independent variables and the log odds of the dependent variable. As aforementioned, this transformation makes it so that it is not necessary for the independent variables to be linearly related to the dependent variable, but it is assumed to be linearly related to the log odds. Testing if this assumption holds true is quite difficult – especially when none of the predictors are continuous.

One final limitation is the gathering and non-randomness of the data. The data structure itself implicitly assumes a randomness by virtue of the methods used (as much models assume a random process to generate the observation). However, in many ways there could be a self-selection bias in who is interacting with the police and a targeting bias. The very bias that was investigated here is critical to the structure of the data. Rather, while the research examines if a bias is occurring once the incident with the police office has happened, it does not address the bias with having an event happen with the police officer at all. Therefore, if the research is to suggest a biasing mechanism in police practices, this mechanism then suggests a non-randomness in the creation of the data set (or rather the event of interactions between police and civilians) insofar as these biases “select,” so to speak cases within the dataset. This issue is incredibly deep and complex, and is not necessarily easily solvable by a particular statistical technique. Rather, this issue delves into methodological questions about the appropriate implementation of research and gathering of data as it pertains to more ecological settings, such as police interactions out in the world. As such, the gravity of such an issue is beyond the scope of the research presented.