An Examination of the Factors that Affect Crime Underreporting

By: Justin Jones

Honors Thesis
Economics Department
The University of North Carolina at Chapel Hill

April 9, 2015

Approved:

Dr. Stephen Lich-Tyler

Acknowledgements

I would like to thank several individuals for their help with this project. First of all, I would like to thank Dr. Stephen Lich-Tyler. As my advisor, he guided me through the thesis process and I certainly could not have done it without him. I would also like to thank Dr. Klara Peter for her guidance throughout the process. She spent countless hours teaching our class and her help on the project was essential. Aside from these two professors, I would also like to thank the other professors that have taught economics to me at UNC. Drs. Tiago Pires, Jeremy Petranka, Richard Froyen, and Simon Alder have taught me much of what I know today about economics. Without their expertise, I would not have had the prior knowledge needed to complete this project.

On a more personal note, I want to thank my two loving parents, Steve and Robin Jones. I cannot describe everything that they have given to me in a thousand pages; so trying to do it in one page is impossible. However, I want to thank them for everything that they have done for me over the past 21 years. I also want to specifically thank my sister, Christine Machado, as well as the rest of my family. I have been raised very well and would not be here today without their help. I would also like to thank all of my friends for the personal wisdom that they have bestowed upon me. While I cannot list all of them here, I would like to particularly thank Joey Gadberry, Kevin Yu, Patrick Komiske, Michael Boner, Tori Hoover, Sam Martin, Ryan Sanzi, Curran Hill, J'woin Harrison, Teerth Brahmbhatt, and James Beacham. Lastly, I want to thank my girlfriend, Caitlin Fisher. Without her help, none of this would be possible.

Abstract

This study explores crime underreporting in the United States in 2013. In analyzing this phenomenon, it seeks to isolate the factors that cause underreporting, as well as certain reasons why underreporting occurs. I use data from the 2013 National Crime Victimization Survey to estimate three separate models. The first model uses a probit regression to isolate several factors that cause crime underreporting. The second model uses a probit regression with inverse probability weighting to correct for selection bias for crime victims. The third model uses a multinomial logit to analyze the most prevalent reasons for underreporting. The results of the first regression suggest that crime type, knowing the perpetrator, age, gender, marital status, and income are the most prevalent factors that affect underreporting. The results of the second regression illustrate the need for selection correction, and point to crime type, gender, marital status, income, and race as the most important factors affecting crime underreporting. The results of the final regression suggest several relationships between underreporting and reasoning, including a link between assault, knowing the perpetrator, and dealing with the crime another way, a link between age and not reporting a crime because it is too insignificant to report, and a link between being Hispanic and not reporting a crime because the police would not help the situation. These findings are presented as a guide to policymakers on the areas they need to address to curb the problem of crime underreporting.

I. Introduction

The underreporting of crime is an extremely important issue in the contemporary United States. While this topic is largely untouched in the media, it deeply affects the lives of many Americans. According to the National Crime Victimization Survey of 2013, only about 38% of crimes in the United States were reported in the past year. While this may not seem like a big deal without context, it greatly impacts many citizens (ICPSR, 1). One may think that most underreported crimes are not ones of a serious nature. However, this is simply not the case. According to the same survey, only 35.7% of rapes were reported (ICPSR, 1). As a serious crime that carries a sentence of life imprisonment in many jurisdictions, this laissez faire approach to rape reporting is simply unacceptable for society. My project seeks to examine why crime underreporting occurs across the United States. In doing this, I examine the factors that impact underreporting.

In examining these factors, there are two main categories to which they belong: criminal and cultural. For the purpose of this analysis, criminal factors are the factors that are directly associated with the crime. Different from factors of the crime, cultural factors also affect crime underreporting. For example, in some communities, which tend to be wealthier, generally, citizens trust the police. These types of communities allow for crime reporting to be high, because citizens trust that their matters will be handled effectively. However, in other communities, the majority of citizens may not trust the police. This problem leads to crime underreporting, because citizens will not entrust many matters with the police. My analysis will seek to examine the factors that cause cultural crime underreporting. While these issues cannot be addressed nearly as much by policy, they could provide possible ideas for cultural changes to address the problem of crime underreporting.

The motivation behind this project comes from a policy and cultural standpoint. Because crime underreporting is such a prevalent issue in today's society, it would be beneficial if policymakers could address it. However, this problem cannot be addresses adequately without isolating the factors that cause it. My analysis provides this isolation by examining the impact that certain factors have on underreporting. It also highlights the specific reasons why citizens do not report crimes. In providing these details on crime underreporting, it provides the necessary information to policymakers so that they can address the issue. It is important to note that this study is designed to provide the necessary information to create policy; it is not a policy recommendation in itself.

While there have been several analyses done on crime underreporting in the past, the literature lacks an updated, comprehensive study. Furthermore, past studies have focused on one type of crime. My thesis focuses on several different types of crime. It also fills the void left by the current literature by providing new results based on updated information from the most recent National Crime Victimization Survey (2013). Because the literature lacks an updated study on the topic of crime underreporting, I study it. This thesis also has larger implications, broader than just crime underreporting. Its information about the relationship between demographics and underreporting provides valuable information on official underreporting in general, not just crime underreporting.

The paper will now present a review of the relevant literature, followed by a presentation of data and theoretical/empirical models. After these sections, the results are presented, followed by a conclusion. Ultimately, after carefully cleaning and analyzing the data, I conclude several things. My findings suggest that crime type, knowing the perpetrator, age, gender, marital status, income, and race are the most prevalent factors in crime underreporting. The strength of the support for each of these variables varies, as some are supported by all of the models, and some are not. The study also suggests several relationships between underreporting and reasoning, including a link between assault, knowing the perpetrator, and dealing with the crime another way, a link between age and not reporting a crime because it is too insignificant to report, and a link between being Hispanic and not reporting a crime because the police would not help the situation.

II. Literature Review

There are several important studies that address crime underreporting. However, there is considerable room for new research in this area. One of the most recent and prevalent studies in crime underreporting was conducted in 2007. This study, done by Dr. W. David Allen, focused on the underreporting of rape crimes. In doing this, Allen used an economic choice model to examine the factors that affect one's decision to report rapes to police. Allen describes information about a crime as an economic resource. He writes, "To understand underreporting, we must understand reporting. A rape victim possesses a scarce resource: information about the crime. Thus, the decision to report, to allocate that information resource, is an economic decision" (Allen, 628). The distinction made here by Allen is very important.

While many see crime reporting as a "black and white" phenomenon with a clear decision, Allen recognizes that there are many factors that affect one's choice to report or not report a crime. Allen (2007) describes the positive factors later in his analysis, when he writes, "A victim who desires social support or legal justice has an incentive to tell others, including police, about the crime; a victim obtains neither by remaining silent" (Allen, 628). Many would think that the support provided by reporting the crime would lead most rape victims to inform the police about what happened. However, as Allen outlines, this is simply not always the case. He writes, "But rape victims who come forward incur real and unique costs. They lose their anonymity, risk retribution by the offender and stigmatization by people they know, and often must participate in an arduous, sometimes openly hostile, legal process" (Allen, 628). These distinctions made by Allen form an excellent choice model, which can be applied to several different realms of underreporting.

As for method, Allen uses dichotomous and multinomial logit models that employ data from the National Crime Survey (Allen, 633). First, he examines the effect of social support and ancillary evidence on rape reporting. He finds that, "victims will more likely report the crime given higher levels of social support and ancillary evidence associated with the crime" (Allen, 635). He also examines the reasons that victims choose not to report rape crimes. He finds that previously knowing the perpetrator significantly affects one's decision to report a rape. If one has been previously acquainted with the perpetrator, they are significantly less likely to report the crime. Overall, Allen's study provides an excellent analysis of the factors that affect crime underreporting in cases of rape.

Stephen D. Levitt completed another analysis of crime reporting with a slightly different focus. In 1998, Levitt examined the relationship between crime reporting and police. In completing this study, Levitt seeks to examine the impact that hiring a marginal officer would have on crime reporting. Levitt does this because he believes that past crime rate policy studies were inaccurate. He writes, "Empirical studies that use reported crime data to evaluate policies for reducing crime will understate the true effectiveness of these policies if crime reporting/recording behavior is also affected by the policies" (Levitt, 61). Later, he gives an example of how this might occur when he writes, "When the size of the police force increases, changes in the perceived likelihood that a crime will be solved may lead a higher fraction of victimizations to be reported to the police" (Levitt, 61). In doing this analysis, Levitt seeks to estimate the exact impact that adding a police officer has on the perceived likelihood that it will be solved (Levitt, 61). He desires to correct the understated numbers from previous reports on crime.

To estimate reporting bias, Levitt uses an instrumental variables approach on data from the National Crime Victimization Survey. Using years of gubernatorial and mayoral elections as instruments for the number of sworn officers, he finds that "the likelihood that a crime will be reported appears to be an increasing function of the number of sworn officers per capita, although the results are by no means definitive" (Levitt, 78). Furthermore, he is able to come up with a succinct marginal effect for each additional police officer. He reports that the addition of one police officer in an urban environment causes the additional reporting of five serious crimes (Levitt, 78). He also estimates the social cost of these crimes, which is about $20,000. Lastly, he illustrates that a traditional cost benefit analysis of adding more police forces underestimates the value of a marginal police officer due to the crime reporting bias illustrated in the paper (Levitt, 78). Overall, Levitt provides an excellent analysis of crime reporting bias.

While Allen (2007) and Levitt (1998) have completed the most comprehensive analyses on crime underreporting, there are several other studies worth mentioning. Leonard (1954) also addressed this topic in his study of crime reporting as a police management tool. In this analysis, Leonard examines crime records from the Federal Bureau of Investigations. Using these records, Leonard presents summary statistics and emphasizes the importance of uniformity in criminal reporting (Leonard, 129). This study is very important in the context that it provides about crime reporting in the past. By providing recommendations to the Federal Bureau of Investigations on how to ensure accurate reports, this analysis shows the past difficulties faced in finding quality crime records and reports.

Soares (2004) studies crime reporting as a measure of institutional development. In doing this, the author seeks to find the determinants of crime reporting and view their overall effect on institutional development. The study itself focuses on three different types of crimes: thefts, burglaries, and contact crimes (Soares, 859). In doing this, Soares seeks to compare the three types of crime and see if reporting differs among them in the context of institutional development. Ultimately, Soares finds that "rates of crime are strongly related to democratic stability and perceived corruption" (Soares, 867). He also finds that contact crimes are reported less than thefts, which are reported less than burglaries (Soares, 867).

There are also several articles outside of the realm of crime underreporting that are worth mentioning. Fu et. Al (1998) study abortion underreporting in survey data. While this study is not a part of the crime underreporting literature, it is still valuable because of its similarities with my topic. Ultimately, the authors find that women were less likely to report abortions in the interview portion of the National Survey of Family Growth (NSFG) than in the self-reporting section (Fu, 131). From this, they concluded, "the usefulness of the NSFG remains extremely limited for analyses involving unintended pregnancy and abortion" (Fu, 128). Udry et. Al (1996) also provides analysis on this topic with a slightly different focus. The researchers in this analysis sought to identify different demographic characteristics that caused the underreporting of abortions. They found that "nonwhite women were 3.3 times as likely as whites to underreport" (Udry, 228). By using logistic regression, they also found that increased education, as well as an increased number of abortions in the past, caused women to underreport (Udry, 228). Jones and Kost (2002), Philipov et. Al (2004), Moreau et. Al (2002), and Jones and Forrest (1992) have all added to the abortion underreporting literature as well, with varying results.

While much of it is focused on crime and abortion, the underreporting literature also extends beyond the bounds of these two subject areas. Ciofi Degli Atti et. Al (2002) examines the underreporting of varicella in Italy. He isolates different demographic characteristics that lead to underreporting, including old age and living in the southern region of Italy (Ciofi Degli

Atti, 479). Goodkind (2011) did a similar analysis of child underreporting in China. However, instead of focusing on characteristics, Goodkind focuses on the problem as a whole, finding that "about 19% of children 0-4 were underreported in the 2000 census" (Goodkind, 291). In a completely different sector, Heitman and Lissner (1995) studied dietary underreporting by obese individuals. They found that "degree of obesity was positively associated with underreporting of total energy and protein" (Heitman, 986). Holland (1958), Martinelli and Parker (2009), and Van Hest et. Al (2002) have also contributed to the general underreporting literature, in varying subjects.

While these studies cover underreporting in general, my analysis provides a more comprehensive view of crime underreporting than the ones currently available. It does this by providing information on multiple types or crimes while also examining many different outside factors that impact underreporting. It also expands upon these factors by matching them with reasons for underreporting and providing an economic choice model with which to analyze the decision making of crime victims. Overall, I seek to build upon these initial studies that I examined.

III. Theoretical Framework

<u>*Reporting as an Economic Choice*</u>

There are several important costs and benefits that one considers when reporting a crime. The costs of reporting can include possible retaliation from the perpetrator, public humiliation, and wasted time in the courts if there is no conviction. The benefits can include justice for the offense, social support, and protection from further offense. It is important to account for this cost-benefit analysis that one considers when reporting a crime. These factors are the inspiration for my theoretical model. This model seeks to examine the proxies for and reasons behind these costs/benefits. Presented in equation 1, this model functions to describe one's choice when reporting a crime.

Equation 1: Economic Choice Model

$$S^*_i = Z_i'\gamma_1 + X_i'\gamma_2 + u_i$$

$$S_i = 1 \text{ if } U_i^A > U_i^B$$

$$S_i = 0 \text{ otherwise}$$

My theoretical model is a traditional economic choice model. The economic agents in this model are crime victims who are trying to maximize their individual utility. The decision that they are making is whether or not to report a crime to the police. This decision is binary, they can simply choose to report or not report a crime to the police. The decision is represented by S_i. As stated in the equation, S_i takes on one of two values. In this case, a value of one represents a decision to not report a crime, while a value of zero represents a decision to report a crime.

This equation relies on the assumption that the victim will choose to not report the crime if his/her utility is higher than it would be if he/she did report the crime ($U_i^A > U_i^B$). In this case, U_i^A represents the victim's overall utility if he/she does not report the crime. Alternatively, U_i^B represents the victim's utility if he/she reports the crime. U_i^A must be greater than U_i^B for the victim to make the decision to not report the crime. The logit regression, which is outlined later, simply focuses on determining the factors that make U_i^A greater than U_i^B and vice versa. The multinomial logit, which is also outlined later, focuses on the reasoning behind the difference between U_i^A and U_i^B.

The model captures the important factors in deciding to report a crime through Z_i and X_i. Z_i represents a vector of cost shifters that are observed by the model. These are the factors that affect the victim's choice, but not the overall outcome of the crime. This vector is mostly characterized by the demographic/cultural characteristics presented later in the project. On the other hand, X_i is a vector of observed factors that affect both the victim's choice and the outcome of the crime. While I am not focused on modeling criminal outcomes, this vector is still important because it has an impact on the victim's choice through expectation of future outcome. The last part of the equation, u_i, represents the factors that are unobserved by the model, but still known to affect the victim's utility.

Equation 2: Criminal Outcome

$$Y_i = X_i'\beta + \alpha S_i + \varepsilon_i$$

Equation two models all of the factors that affect a crime outcome. This equation is not the concern of this analysis, but it provides important context on equation one. Y_i represents the overall outcome of the criminal process for the victim. X_i represents the important factors that affect both the decision to report a crime, and the outcome for the victim. S_i is the result of the decision-making process as described by equation one. ε_i is a grouping of unobserved factors that are not present in the model. While it may be important to analyze crime outcomes (Y_i) as a function of different factors, this concept is outside of the scope of this project. Instead, I focus on equation one, the decision to report or not report a crime. I provide equation two simply to illustrate the full context behind the choice that one makes as a crime victim. Equation one describes the actual choice, which is the focus of this project.

Important Factors: Criminal and Cultural/Demographic

In examining the relationship between U_i^A and U_i^B, it is important to identify the factors that affect one's utility when reporting a crime. In this analysis, there are two main categories to which these factors belong: criminal and cultural. Criminal factors are the factors that are directly associated with the crime. These include the type of crime, whether the perpetrator of the crime used a weapon, and whether the victim of the crime knew the perpetrator. I selected these characteristics specifically because of their relevance to a victim's decision to report a crime. I also selected them because they are all pieces of information that one may have to provide on a typical 911 call. Because I thought that these factors would be under serious consideration in the victim's mind as a crime is happening, I chose them for my model. These three factors are responsible for isolating criminal characteristics within the model.

Cultural/demographic factors, on the other hand, are different from the criminal act that takes place. Instead, these factors are characteristics of the victim that may influence his/her decision to report a crime. It can be assumed that the personal experiences of a victim play a large role in his/her decision to report. The variables that I use for cultural/demographic factors are designed to group people in accordance with their personal experiences in order to identify common patterns in underreporting among them. I use the victim's age, gender, race, marital status, income, and educational attainment for cultural characteristics. In doing this, I hope to capture the effect of one's culture and personal experiences on his/her decision to report a crime.

IV. Data

Data Source/Sample

The dataset that I use is The National Crime Victimization Survey from 2013. The United States Census Bureau collects this data for Bureau of Justice Statistics. This dataset is designed to provide accurate information about crime in the United States. In doing this, the survey seeks to accomplish three goals: “to develop detailed information about the victims and consequences of crime, to estimate the number and types of crimes not reported to police, and to provide uniform measures of selected types of crime” (ICPSR, 1). The survey samples individuals from randomly selected households across the United States (ICPSR, 1). The data is structured as a cross-section, as it was collected in 2013. It features the answers of each individual to a large survey of crime-related questions. Based on this format, the data is considered to be self-reported. Specifically, the survey asks respondents to report if they were victims of crime in the past six months. It also asks them to report other characteristics of the crime (if the respondent is a victim), as well as various demographic characteristics. The survey comes with a wealth of supplemental materials, including a codebook, the survey as it is presented to individuals, and a detailed description of each individual variable.

The dataset that I use is very strong and well-suited to my topic. Specifically, the dataset is strong because it is the newest available data on the topic and it was collected by a very reputable source. It is more recent than the data used by every study that I reviewed in the literature review. While the data used by Allen (2007) is very detailed and comprehensive, it is not updated to reflect the current landscape of crime underreporting. My dataset contains the most current and relevant information on the topic. Meanwhile, Levitt (1998) and Soares (2004) use very comprehensive datasets, but they are also not updated to reflect current crime underreporting conditions. Because my data is updated for 2013 and the 2014 set has not been released yet, it is the most current and relevant set to use. It is also strong because it was collected by one of the most reputable sources available. The United States Census Bureau is one of the most trusted suppliers of survey data. The data is also very well-suited to my topic. As the most updated and comprehensive source of crime data, the National Crime Victimization Survey is perfectly suited for the topic of crime underreporting.

Because the data was sampled randomly among the entire U.S. population, it does not contain a large amount of limitations. The only major limitation that comes into play is the limited number of observations that are present in the sample of crime victims. From this large dataset (81,875 observations), the sample that I am most concerned with is the victims of crime. While a large majority of the subjects in the dataset were not victims (74,963), it includes 6,912 crime incidents. This smaller number of crime victims is slightly concerning, as it greatly reduces the overall sample size and could increase standard errors. However, despite this reduction in sample size, I still obtain meaningful results.

Another minor limitation is due to the way that the survey was conducted. The complete survey has approximately 160,000 observations. However, the takers of the survey only have complete information for 81,875 observations. This is due to the fact that the survey takers only asked the "household respondent" for all of the information. Defined by the NCVS as the "household member chosen by the interviewer who appears to be the most knowledgeable about the household composition and is able to answer the household screen questions dealing with crimes against the household," the household respondent is the only member of the population who can answer all of the questions necessary for this analysis (ICPSR, 507). For this reason, I only include household respondents in my sample. However, this alteration to the data does not drastically alter the characteristics of the sample. It also reduces the sample to 81,875, which is still sufficiently large enough to complete my analysis. I am only noting this limitation so that the reader is aware of it.

Dependent Variables: "Not Report" and "Reason"

The main two dependent variables in this analysis are "not report" and "reason". A detailed description of how these variables were collected and reported is presented in table one. "Not report" is the dependent variable for my probit regression, which I run first. My theoretical model presents S_i^* as the choice between not reporting and reporting a crime. This variable is perfectly exemplified in my data source by the variable "not report". This variable measures whether or not a victim decided to call the police when he/she was victimized by crime. This variable is measured with a dummy variable in the dataset. A "0" represents a crime that resulted in a police report. On the other hand, a "1" represents a crime that was not reported. These two outcomes, which are presented in figure one, represent the dependent variable for my probit regression. As shown in the graph, the majority of crimes in the dataset were not reported. Out of the 6,912 crimes that I analyze, 2,820 (40.8%) were reported and 4,092 (59.2%) were not reported. A more detailed description of this variable and its interaction with my independent variables is presented in table three.

For the multinomial logit section of the project, I use a slightly different dependent variable, called "reason". The variable seeks to explain $S*_I$ from my theoretical framework in a slightly different manner. Instead of looking at just the victim's decision to report or not report, this variable looks at the reasoning behind not reporting. The base value (0) for this variable is the same as the "not report' variable. A value of "0" represents a victim's decision to report the crime (outcome zero). The other five outcomes represent specific reasons for why crimes were not reported. Outcome one is that the crime was dealt with another way (occurred in 7.1% of crimes). This means that the victim reported the crime to some other official source, or the victim dealt with the matter personally. Outcome two is that the crime was not important enough to report (18.9% of crimes). Outcome three is that the respondent felt that the police couldn't do anything about the crime (9.4% of crimes). Outcome four is that the police wouldn't have helped the crime situation (13.2% of crimes). Outcome five is any other reason that the respondent did not report the crime (10.3% of crimes). This category covers all other reasons presented for not reporting crimes in the dataset. These six categories form the dependent variable for my multinomial logit. They are presented with numerical values in table three.

Independent Variables

As for the independent variables, I incorporate factors of the crime as well as cultural factors. The crime-related factors that I include are: the type of crime, whether a weapon was used in the act of committing the crime, and whether or not the victim knew the perpetrator of the crime. For the last two variables mentioned, the NCVS only takes data for assault victims. It is important to recognize this fact while interpreting the results. I argue that this set of variables has the most impact on the decision-making of a victim in a crime. The crime variables are all coded as dummy variables.

The first of these variables, type of crime, is divided into five different categories. These categories are theft, burglary, auto theft, assault, and other. I use auto theft as the reference category for this section, so the other crimes are presented in relation to auto theft. As stated before, all of these variables are coded as individual dummy variables. For example, a value of

"0" for theft indicates that a person has not experienced theft within the past six months. In contrast, a value of "1" for theft indicates that a person has been a victim of theft within the past six months. The other variables for the type of crime are also presented in this manner. The exact definitions and survey questions for each of these variables are presented in figure one.

The other crime variables that I use for this analysis deal with other characteristics of the crime other than the type of crime. The first of these variables is "weapon used". This variable, which is detailed in table one, is designed to measure the assault crimes in which the perpetrator uses a weapon. I hypothesize that the presence of a weapon greatly affects one's decision to report an offender of a crime. It is important to note that this data was only gathered for assault victims by the NCVS. This makes sense because the large majority of weapon-involved crimes are assaults. This variable can be viewed as an interaction term between assault and weapon usage. A value of "1" represents an assault crime in which the perpetrator used a weapon. A value of "0" represents an assault crime in which the perpetrator did not use a weapon, or a non-assault crime. As illustrated in table three, this variable occurs 4.7% of the time among crime victims. It is important to keep the details of this variable's coding in mind when evaluating the results.

The second of my other crime variables is "attacker known". This variable represents the situation in which a victim of assault knows the offender. The survey questions that form this variable are presented in table one. Similar to "weapon used", the NCVS only asks this question of assault victims. It does this because it assumes that the large majority of observations in this category are assault victims. This variable can be viewed as an interaction between assault and knowing the attacker. A value of "1" represents an assault crime in which the victim knew the attacker. A value of "0" represents an assault crime in which the victim did not know the perpetrator, or a non-assault crime. As illustrated in table three, this variable occurs 3.2% of the time among crime victims. It is important to keep the details of this variable's coding in mind when evaluating the results. This variable is also important in considering the phenomenon of domestic violence, as I discuss later in the Results section.

As for cultural variables, I focus on the demographics of the victim. These variables include: age, gender, race, marital status, income, and educational attainment. These variables are all coded as single dummy variables except for age, race, and income. Age is a continuous variable. Race and income are divided into a set of four categories. Most of these variables do not require much explanation. If clarification is needed, the detailed coding of the variables is presented in table one. However, I will provide a brief description of each variable. For age, the victim's age at the time of the survey is presented. The mean age for the survey is 50.3 years old with a standard deviation of 0.63. The gender variable is coded so that "female" equals zero and "male" equals one. There are slightly more females in the data, with only 44.7% of the sample being male. Most of the sample is also married, with 52.2% reporting this way (not married equals zero, married equals one). The educational attainment variable is coded so that college graduates take a value of "1", while non-graduates take a value of "0." It is important to note that I define college graduate as someone who has received an associates degree or higher. 41.2% of my sample reported that they were college graduates.

The multi-category demographic variables have more complex breakdowns. First, race is divided into four separate categories. The base category is "White", which comprises 71.8% of the sample. "Black" and "Hispanic" make up the next two largest categories (10.6% and 11.4% of the sample, respectively). It is important to note that the "Hispanic" variable is comprised of those who identify as both White and of Hispanic ethnicity. I coded the variable this way to represent the portion of the population that comes from Latin/South America, and tends to identify as a cultural group. Those of Hispanic origin who identify with races other than White are grouped in their respective racial group. After those groups, the last category is "Other" (6.2% of the sample). This group comprises all other races, including Asians, Native Americans, and biracial people among others. The detailed description of this variable and its respective survey questions are presented in table one.

Lastly, the second multi-category demographic variable is household income. Household income is defined as the amount that all members of the household made in the past year. This variable is coded into four different categories. The first category comprises members of households that made less than $25,000. This category makes up 14.5% of the sample. The next category is made up of members of households that made between $25,000 and $75,000. This is the largest income category at 33.5% of the sample. Next is the highest income category, which is comprised of members of households that make over $75,000 per year. This category contains 21.4% of the sample. Lastly, the "unknown" category represents the respondents that did not report their household income (30.6%). Using these demographic variables, I am able to model the culture surrounding each member of the sample. The detailed descriptions of all of these variables are presented in table one.

Summary Statistics/Analysis

While I have briefly mentioned some of the summary statistics in the previous section, I will now discuss them in depth. The relevant summary statistics for the data are presented in tables two through four. These tables each contain different sets of statistics for different portions of the sample. Table two presents summary statistics for the entire sample. Table three presents summary statistics for only crime victims. Table four is of a slightly different orientation, as it details the specific statistics for the "underreporting reasons" variable. All three of these tables present important statistics that can be used to initially evaluate the data.

As stated before, table two presents the relevant summary statistics for the entire sample of 81,875 observations. By running t-tests between crime non-victims and victims, I am able to illustrate the selection process for crime victims. While I will not mention the summary statistics for every single variable, I will mention the significant differences between the two groups that make them different from each other. The first noticeable difference between crime victims and non-victims is found in the age of the victim. It appears that victims of crime tend to be slightly younger than non-victims in our sample. The mean age for non-victims in our sample is 50.8 years old, while the mean age for victims is 44.8 years old. Based on my t-test for this group, I reject the hypothesis that there is no age difference between non-victims and victims at the 99% confidence level.

I find similar significant results for my race variables. At the 99% confidence level, I reject the hypothesis that there is no difference in racial composition between non-victims and victims. Whites are underrepresented among crime victims. They make up 72.3% of the non-victims, but only 65.7% of the victims. The other three racial groups are overrepresented among crime victims. Blacks make up 10.4% of the non-victim sample, while making up 12.7% of the victim sample. Hispanics represent 11.2% of the non-victim sample, but they make up 14.1% of the victim sample. The same phenomenon is present among the "other" races. They account for 6.1% of non-victims, but 7.5% of victims. Based on these figures, it appears that minority groups are disproportionately impacted by crime in the United States.

Next, I reject the hypothesis that the proportion of married non-victims and victims is equal at the 99% confidence level. It appears that non-married people are disproportionally affected by crime, as they make up 46.9% of non-victims, but 57.9% of victims. The next important variable is household income. Based on my t-tests, I find that the “less than $25,000”, “greater than $75,000”, and “unknown” categories are all significant at the 99% confidence level. This indicates that poor (<$25,000) citizens are disproportionately impacted by crime. They make up 13.9% of the non-victims, but 20.6% of the victims. Alternatively, rich citizens are disproportionately unaffected by crime. According to the statistics, those with incomes over $75,000 comprise 21.7% of the non-victims, while making up only 18.9% of the victims.

The “income unknown” category is also disproportionately affected by crime, as it makes up 31.1% of the non-victims, but only 25.7% of the victims. Lastly, the $25-$75,000 category is statistically significant, but only at the 95% level. Those at the middle-income level are over-affected by crime, comprising 33.4% of non-victims and 34.8% of victims. As illustrated by the statistics, income plays a very large factor in the selection of crime victims. Lastly, it is important to note that the variables for both gender and educational attainment are not significant at any of the traditional confidence levels. This indicates that there are not significant gender or educational biases in the selection of crime victims.

Table three presents the smaller portion of the sample that I am most interested in studying: the victims of crime. This table compares proportions across crimes, as well as demographic factors. For criminal factors, four variables illustrated a significant difference between the proportion of reporters and non-reporters of crime. At the 99% confidence level, I reject the hypothesis that the proportion of crime reporters and non-reporters who are theft victims are equal. I find the same rejection to also be true for burglary victims. However, the effects of the crime types work in different directions. Victims of theft make up a disproportionate amount of crime non-reporters (46.5% versus 43.2% of reporters). On the other hand, victims of burglary comprise a disproportionate amount of reporters (11.3% versus 7.9% of non-reporters). The "other" crime type variable is the third significant criminal factor. At the 95% confidence level, I reject that the proportion of reporters and non-reporters are equal for victims of other crimes. It appears that other crime victims are overrepresented among crime reporters (37.5% of reporters versus 35.0% percent of non-reporters. Lastly among criminal factors, the "attacker known" variable is significant at the 90% confidence level. This indicates that the proportion of crime non-reporters is greater than the proportion of reporters for crime victims (3.5% of non-reporters versus 2.7% of reporters). As for the variables for auto theft, assault, and weapon used, they were not significant at any traditional confidence level. This indicates that they are not disproportionally represented among either the reporting or non-reporting population.

As for the demographic/cultural factors, five different variables came up as significant. First, the gender variable was significant at the 90% confidence level. The statistics indicated that males make up a greater proportion of the reporting population (46.2%) than they do for the underreporting population (43.8%). Out of the race variables, the only significant one is the variable for "Black" (at the 90% confidence level). Blacks make up a greater proportion of the reporting population (13.6%) than they do for the underreporting population (12.1%). Aside from race and gender, the next variable that is significant is the variable for marital status. This variable is also significant at the 90% confidence level. The statistics indicate that married people make up a greater proportion of the reporting population (43.4%) than the underreporting population (41.3%). Lastly among the significant variables, income under $25,000 is significant at the 95% level, and income unknown is significant at the 99% level. Poor citizens are underrepresented among crime reporters (19.4% versus 21.4% of non-reporters). On the other hand, citizens that did not report their income are overrepresented among crime reporters (28.0% versus 24.1% of non-reporters). The variables for age, white race, other race, income between $25-75,000, income greater than $75,000, and college graduate are not significant at any traditional confidence level. This indicates that they are not disproportionately represented among their individual populations. It is important to keep these summary statistics in mind as I move forward to the analysis.

Lastly, in table four I detail the distribution of the "underreporting reasons" variable. While "reported to police" is the largest segment of this variable (41.1%), "not important enough to report" is the most common reason cited for not reporting a crime among my sample (18.9%). The other four underreporting reasons are relatively evenly distributed, ranging from 7.1% to 13.2%. I must keep the distribution of these reasons in mind as I move forward to the multinomial logit section of the project.

V. Empirical Model

Probit Model

For this analysis, I use two separate regression techniques to analyze crime underreporting. The first technique is a probit regression. I use this regression type due to the binary nature of my dependent variable. This dependent variable is whether or not a crime is reported to police. As stated before, this variable is measured with a dummy variable. A “0” represents a crime that resulted in a police report. On the other hand, a “1” represents a crime that was not reported. Because the data are coded in this manner, the coefficients can be interpreted as the impact a certain independent variable has on a whether or not a crime is reported. A positive coefficient means that an independent variable leads to a decrease in the probability that a crime is reported. Conversely, a negative coefficient means that the independent variable has a positive effect on the probability that the crime is reported. The variable for whether or not a crime is reported is the only outcome that I am interested in measuring with my probit regression. The orientation of the variable is slightly counterintuitive, however, it is necessary for the multinomial logit of the project to work successfully. The mathematical representation of this probit model is presented in equation three.

Equation Three: Probit Model

$$Y_i = X_i'\beta + \varepsilon_i$$

P(Crime not Reported) relative to P(Crime Reported)

Probit Model with Inverse Probability Weighting

As illustrated by the summary statistics in the Data section, the population of crime victims is made up of individuals with different characteristics than the general sample population. While the initial probit model provides a decent view of the important factors for reporting among crime victims, it is not fully applicable to the general population. Because there is a clear selection problem with crime victims, I include another regression model that accounts for selection. Called inverse probability weighting (IPW), this method produces marginal effects and p-values that are adjusted to reflect the entire sample, and not just crime victims. The exact method in which IPW weights the observations is presented in equation four. Because this method reduces the selection bias among crime victims, it provides more accurate estimates.

The characteristics from the entire sample are used to determine the probability of response in the model through an initial probit regression on "crime". In this case, I use the cultural/demographic variables to predict "crime". The probabilities derived from this initial estimate are then divided from one, applied to the observations in the crime victim sample, and regressed (ESRC, 1). This method allows me to use all 81,875 observations in my sample in a productive manner instead of only focusing on the 6,912 crime victims. It will be important to compare the results of this method with the original probit model to view the scope of the selection problem.

Equation Four: Probit with Inverse Probability Weighting

$$X_i = Z_i'\gamma_1 + \varepsilon_i$$

Weight for Observation = 1/(Probability of Observation)

$$Y_i = X_i'\beta + Z_i'\gamma_1 + \varepsilon_i$$

P(Crime not Reported) relative to P(Crime Reported)

Multinomial Logit Model

The second part of the project uses a multinomial logit approach. I use this approach due to the complex nature of my dependent variable. This part of the project seeks to compare the probability that a crime was not reported for a specific reason to the probability that it was reported. Instead of being a simple binary variable as the dependent variable is in the first portion of this project, the dependent variable for the second part of this project is a categorical variable that takes on one of eight values. The base (zero) outcome of this multinomial logit is the outcome for "0," which represents a crime that was reported to police. The other five outcomes, which are detailed in the Data section, represent the victims' specific reasons for why crimes were not reported. All coefficients for this model should be interpreted as relative to the base category (crime reported). A positive coefficient indicates that the presence of the choice variable makes it more likely that a victim will cite the given reason for not reporting a crime rather than reporting it. A negative coefficient indicates that the presence of the choice variable makes it less likely that a victim will cite the given reason for not reporting a crime rather than reporting it. It is important to keep these details in mind when evaluating the results. The structure of the empirical model is found in equation five. It is important to note that the sample size (N = 6,864) for the multinomial logit is slightly smaller than for the probit. This is due to the fact that 48 observations have missing data for "underreporting reason".

Equation Five: Multinomial Logit Model

$$Y_i = X_i{'}\beta + \varepsilon_i$$

P(Crime not Reported for Reason 1-5) relative to P(Crime Reported)

VI. Results/Implications

Probit Regression: Results

The probit regression section of the project has identified several key factors that affect crime underreporting. Before discussing the results, it is important to note that they are completely detailed in table five. First, out of the crime types, two are significant at traditional confidence levels. The first of these is burglary. At the 99% confidence level, I reject the hypothesis that burglary has no effect on whether a crime is reported. Based on the negative marginal effect for burglary, I infer that a crime being a burglary makes it less likely to go unreported. The marginal effect on burglary reveals that the fact that a crime is a burglary makes it 11.3 percentage points more likely that it will be reported. The second crime type that was significant was “other crime”. At the 95% confidence level, I reject the hypothesis that a crime being in the “other” category has no effect on whether a crime is reported. The marginal effect on “other” is negative. This indicates that the other crimes are reported at a lower rate relative to the reference group. The marginal effect on rape indicates that the fact that a crime is in this “other” category makes it 3.5 percentage points more likely to be reported.

As for the criminal characteristics outside of crime type, the only one to exhibit significance is the variable for the victim knowing the perpetrator. At the 90% confidence level, I reject the hypothesis that the victim knowing the perpetrator has no effect on whether a crime is reported. As the positive marginal effect on this variable indicates, the victim knowing the perpetrator in an assault crime causes an increased chance that the crime will not be reported. The marginal effect of this variable shows that knowing the perpetrator in an assault crime makes a victim 7.0 percentage points less likely to report a crime. This result is very intuitive, as

I will discuss in the Implications section. This variable is the last of the criminal characteristics that are significant for the probit model.

For cultural/demographic characteristics, there are several significant variables. First, at the 95% confidence level, I reject the hypothesis that age has no effect on the reporting of a crime. As the positive marginal effect on age illustrates, the greater that a person's age is, the less likely he/she is to report a crime. Specifically, each extra year that a person ages causes them to be 0.01 percentage points less likely to report a crime. While this does not represent a large marginal effect, it is still significant at the 95% level. Next, at the 95% confidence level, I reject the hypothesis that gender has no effect on whether a crime is reported. The marginal effect on this variable is negative, indicating that crimes against men are more likely to be reported than crimes against women. The marginal effect on this variable shows that if a victim of a crime is male, it is 2.6 percentage points more likely to be reported.

Marital status is also significant at this level. At the 95% confidence level, I reject the hypothesis that marital status has no effect on crime underreporting. The negative marginal effect on "married" illustrates that being married makes a person more likely to report a crime. Based on this value, being married makes a person 2.6 percentage points more likely to report a crime than not being married. For income, the only variable that is significant for income is the unknown category. At the 99% confidence level, I reject the hypothesis that the presence of an unknown income has no effect on crime underreporting. Specifically, it appears that people who did not report their income were 5.3 percentage points more likely to report the crime (negative marginal effect). For all of the other variables featured in the model, I cannot determine a significant relationship with crime underreporting. At the 95% confidence level, I fail to reject that all of the other variables have an effect on whether a crime is reported. Several of these variables play a larger role in the multinomial logit part of the project.

Probit Regression: Implications

This part of the project contains several important implications that can be used by policymakers to target specific areas. The first of the significant criminal characteristics, the variable for burglary, contains positive information on reporting. As the negative marginal effect indicates, burglaries are reported at an extraordinarily high rate. As stated before, the fact that a crime is a burglary makes it have an 11.3 percentage point greater chance of being reported in relation to the reference group. This type of crime does not require policy intervention when it comes to underreporting. This result also makes intuitive sense. If someone breaks into one's residence (burglary), the victim's immediate reaction is almost always to call the police. When it comes to burglary, crime underreporting is not a significant problem.

While crime underreporting is not a significant problem for burglaries, it seems to be important for other types of crimes that are not specified in the data. According to the model, these "other crimes" are reported at a significantly low rate. Holding all other variables constant, if a crime falls into this category, it has a 3.5 percentage point greater change of going unreported. This is significant for policymakers to consider. Based on the model, it is quite clear that these other crimes are not being reported at a high enough rate. If they want to decrease underreporting for these "other" crimes, policymakers must focus their initiatives on increasing the expected utility of reporting them (U_i^B), and decreasing the expected utility of not reporting them (U_i^A). As I outlined in the theory section, a crime victim will only report to the police if his/her utility is higher for reporting. The second part of this project seeks to isolate the reasons why victims of other crimes report them at a lower rate. This gives policymakers a better idea of how to raise the expected utility of reporting these crimes.

The last criminal characteristic that is significant in the model is the victim's familiarity with the perpetrator in assault crimes. Based on the model, a victim is 7.0 percentage points less likely to report a crime if it is assault and he/she knows the attacker (relative to all other crimes). This is another important area that policymakers must address. While some may deem offenses that were committed by an acquaintance to be personal matters, serious crimes like these assaults are often not reported. Many of these crimes fit the description of domestic violence, which has been an important policy topic in recent memory. If policymakers want to increase the reporting rate for these serious crimes, they must find a way to convince victims to report their acquaintances who perpetrate the acts. The criminal characteristics of the model provide several important implications to policymakers.

The cultural characteristics of the model are also important to examine. As mentioned before, the first statistically significant cultural variable is age. Based on the marginal effect, an increase of one year causes a person to be 0.01 percentage points less likely to report a crime. While this metric is statistically significant, it does not contain a large amount of practical relevance. The marginal effect is so small that it would be difficult for policymakers to target this variable. However, if they were to find an area to target with policy, policymakers should look towards older people. These elderly people tend to report crimes at a marginally lower rate. The next important variable is gender. The marginal effect illustrates that males are 2.6 percentage points more likely to report crimes than females. This indicates that policymakers should target females and implement policy that encourages them to report crimes. Going back to the theoretical model, these policies should increase the utility of reporting crimes for women. It is important for policymakers to find out the reasoning behind this gender bias. I will attempt to address this reasoning in the multinomial logit section of the project.

Other than this slight age bias and gender bias in reporting, only marital status and unknown income are significant. As the model shows, holding everything else constant, being married increases one's likelihood of reporting a crime by 2.6 percentage points. This means that policymakers should target non-married people. One hypothesis for the difference between these groups is that married people have families, and they seek to protect their families by reporting crimes. However, the model does not comment on this; it is up to the policymakers to identify this difference and rectify it. The multinomial logit part of the project seeks to identify some of the reasons behind these differences. Lastly, the variable for "income unknown" is significant. The practical relevance of this finding is unknown. A separate analysis on the reasoning behind people not disclosing their income may have more to report on this topic. These implications lead to the multinomial logit portion of the project, which seeks to build on them.

Comparison between Initial Probit and Probit with IPW

The initial probit results and the probit with IPW results are very similar. For example, the variables for burglary, gender, marital status, and unknown income are all still significant at the same confidence level in the IPW model as they were in the initial probit model. While the marginal effects of these variables have changed slightly between models, the practical results are nearly the same. It is also important to note that nearly all of the variables that are not statistically significant at traditional levels in the initial probit model have remained statistically insignificant.

However, there are some important differences between the two models. These differences reflect the selection issues that I hypothesized were a part of the sample of crime victims. The first variable that exhibits a change is the “other crime” variable. While it is significant at the 95% confidence level in the initial probit model, it is not significant at any traditional level in the probit with IPW model. Based on this model, I fail to reject the hypothesis that a crime being in the “other” category has no effect on a victim’s decision to report it. A similar phenomenon occurs with both the “attacker known” and “reason” variables. In the initial probit model, both of these variables are significant at the 90 and 95 percent levels. However, when corrected for selection bias with IPW, these variables are no longer significant at any traditional confidence level. The probit with IPW model causes several variables from the initial probit model to no longer exhibit statistical significance.

However, this change in statistical significance between the two models also occurs in the opposite direction. Specifically, this phenomenon occurs in two of the variables for race. Both Hispanic race and other race are significant at the 95% confidence level in the probit with IPW model. Both of these variables are not significant at any traditional confidence level in the initial probit model. Based on the positive marginal effect and high relative risk ratio on both of these variables, I find that their presence increases the chance that a crime will not be reported. Specifically, if a Hispanic is victimized by crime, he/she is 4.4 percentage points more likely to not report it to police. Similarly, if a person of “other race” is a victim, he/she is 4.9 percentage points more likely to not report it. These findings have important broader implications. If minority groups such as these are reporting crimes at a lower rate, policymakers must find a way to encourage them to report. The results of the multinomial logit shed more light on the reasons why minorities may report crimes at a lower rate.

While many of the results are the same between the two models, there are some very important differences between them. These differentiated results illustrate the importance of using selection correction. As the second model shows, there are some biases present among the initial probit model. The probit with IPW also added some support for the initial regression, as many of the significant variables from the initial probit model are significant in the IPW model. Overall, the probit model with inverse probability weighting provides an excellent way to check for biases in the first model.

Multinomial Logit: Results

The full regression results for the multinomial logit are presented in tables six and seven. I will briefly summarize the results and their context within the project. The first set of results is for the outcome where a victim dealt with a crime in another way other than reporting it. As explained before, my multinomial logit results present a comparison between the likelihood of this reasoning in not reporting a crime and the likelihood that the crime was reported. For this first regression, the first criminal factor that is significant at the 99% confidence level is burglary. The negative coefficient shows that the fact that a crime is a burglary leads to it being reported at a higher level in relation to this outcome. In other words, burglaries are not often dealt with another way when compared with the other crimes in this set of data. The relative risk ratio also illustrates this notion, as it is below one (0.369). This breakdown of the numbers serves as a guide for the rest of the values that I present. I do not use as much detail from this point to avoid incessant repetition. The other criminal factors that are significant at traditional confidence levels for this outcome are “assault” and “attacker known”. The positive coefficient and high relative risk ratio on “assault” indicate that a victim is more likely to deal with a crime in another manner if it is an assault. The positive coefficient and high relative risk ratio on "attacker known" also indicate that a victim is more likely to deal with a crime in another way if he/she knows the attacker, relative to reporting the crime to police.

For cultural/demographic factors, male gender, Hispanic race, and other race are significant at traditional confidence levels. The coefficients are negative and the relative risk ratios are low for gender and Hispanic race. This indicates that a victim with either of these characteristics is less likely to deal with a crime in another way relative to reporting to police. The effect moves in the opposite direction for the "other race" variable. If someone is in the "other race" category, they are less likely to deal with a crime in another way relative to reporting to police. This trait is exhibited by the negative coefficient and low relative risk ratio on the variable for "other race".

The next outcome deals with the victim thinking that a crime was not important enough to report to police. This outcome has several statistically significant variables. All of the "crime type" variables are significant for this outcome at standard confidence levels. They all exhibit negative coefficients and low relative risk ratios. Based on these coefficients and relative risk ratios, the fact that a crime is a theft, burglary, assault, or other crime (when compared with an auto theft) causes a decreased chance that one may consider it to be too insignificant to report, relative to reporting it to police. As for cultural variables, both the "age" and "income unknown" are significant at the 99% confidence level. However, the two variables work in opposite directions regarding crime underreporting. The positive coefficient and large relative risk ratio on "age" indicates that an increase in the age of the victim makes it more likely that he/she will deem a crime not important enough to report, relative to the reference group. The effect for "income unknown" is in the opposite direction. The negative coefficient and large relative risk ratio on this coefficient indicate that the presence of this variable makes a victim less likely to cite a crime as being unimportant as a reason for not reporting it. It is important to keep these significant variables in mind as I move toward the results section.

The third "reasoning" outcome encompasses the situation where a victim does not report a crime because he/she feels that the police cannot do anything about it. This result does not necessarily contain negative implications about the police, but instead deals with the victim thinking that the incident is simply not within their jurisdiction. The variables that exhibit significance at traditional levels for this outcome are burglary, black race, income unknown, and college graduate. All four of these variables exhibited the same type of effect relative to the base category. Based upon their negative coefficients and low relative risk ratios, these four variables all cause the probability of one not reporting because they feel that the police cannot do anything about it to decrease relative to the base outcome. The implications of these results will be discussed further, later in the project.

The fourth outcome of the regression is that the crime was not reported because the victim felt that police involvement would not help his/her situation. For this outcome, four variables are significant: theft, Hispanic race, income unknown, and marital status. Several of the variables work in different ways. For marital status and unknown income, the coefficient is negative and the relative risk ratio is less than one. This indicates that if a victim is married or does not disclose his/her household income, they are less likely to not report a crime because of a belief that the police will not help, relative to the base outcome. This indicates that married individuals tend to trust the police more than non-married individuals. For the variables of theft and Hispanic race, the effect is in the opposite direction. Victims of theft and crime victims who are Hispanic tend to not report crimes because they believe that the police will not help at a higher rate relative to the base outcome. This result is taken from the positive coefficients and high relative risk ratios on these variables. The consideration that some victims may think the police are biased is important to think about when reviewing the topic of crime underreporting.

Lastly, the fifth outcome of the regression is that the victim did not report the crime for other reasons. This outcome contained several significant variables, including burglary, attacker known, male gender, and Hispanic race. As exemplified by their positive coefficients and high relative risk ratios, victims of assault who know their perpetrator and victims who are Hispanic tend to cite other reasons for not reporting crimes to police. The variables for burglary and male gender work in the opposite direction. Burglary victims and males do not tend to cite other reasons relative to the base outcome. This is exemplified by the negative coefficient and low relative risk ratio on the variable for gender. Overall, this section provides an excellent isolation of certain factors that interact with crime underreporting reasons. In the next section, I dissect these findings and present their greater implications for crime underreporting.

Multinomial Logit: Implications

The implications of this portion of the project are much more focused than the first part. In this part, I am able to isolate the specific reasons why crime underreporting occurs. I am also able to match these reasons with certain criminal and cultural/demographic characteristics. The last part of this process involves taking these isolated reason-factor pairs and describing what they mean for crime underreporting. It is important to note that the most prevalent areas of this concern for this study involve areas in which crime underreporting is a problem. For this reason, it is not necessary to focus greatly on the areas in which underreporting is not a problem. In practical terms, this means that I will not place much emphasis on the negative coefficients from the multinomial logit. These coefficients represent areas in which the reason given is significantly less likely to occur than a report being made. Instead, I focus on the most common reasons given for not reporting (positive coefficients and high relative risk ratios).

From the first few outcomes of the logit, several specific results stand out. The first is the positive coefficient for attacker known on outcome one. Based on the significance and direction of the coefficient, it appears that victims who know their perpetrator generally prefer to deal with their crimes in other ways than the reporting to the police. This makes intuitive sense, as they may try to solve the problem through mediation or some other method than the legal system. However, as I discussed previously, this can create huge issues. If victims choose to report their crimes to other sources, these sources often mishandle the responsibility. A good example of a source other than the police is a university conduct board. As a college student, I have heard countless stories of university conduct boards mishandling cases. If a victim reports a crime to police, this problem can be avoided. Policymakers should create policy that encourages victims to report perpetrators even if they are acquaintances. Once again, it is important to note that the data on this variable is only present for assault victims, so we can only draw refined conclusions about this group. However, it can be assumed that this phenomenon would likely apply to other crime types if data were available on the victim's familiarity with the perpetrator for these crimes.

Aside from the "attacker known" variable, the "assault" variable is also significant for the "dealt with another way" outcome. This part of the logit provides a link between assaults and dealing with crimes another way. The positive coefficient on assault illustrates that assault victims often feel that the crime against them is better handled without police help. This result is very intuitive when thinking about domestic violence issues. In the past year, domestic assault has become one of the most discussed issues in the United States. My findings provide support for the oft-provided assertion that domestic assault victims do not report their perpetrators. Based on the positive, significant coefficient on assault, I conclude that assault victims report their perpetrators to police at a lower than normal rate. When combined with my results for "attacker known", these findings shed tremendous light on domestic violence issues. Policymakers should focus on getting assault victims to report their perpetrators.

Another important implication comes from outcome two. The coefficient for age is both significant and positive for this outcome, indicating that older people are more likely to say that a crime is "not important enough to report" than younger people. The implications of this variable could work in several directions. Older people may be more seasoned and can recognize that some incidents are truly not important enough to report. However, it may also be the case that there are significant barriers to reporting for older people, and they cite this reason as a part of those barriers. The true meaning of this finding is not present in the data. However, it could be important for policymakers to identify which scenario is true, and try to rectify it. Other ambiguous outcomes are present in the data, including the significance of "burglary" on "police wouldn't help", and the significance of "attacker known" and Hispanic race on "another reason". While I cannot draw meaningful implications from these significant findings, it is important for policymakers to recognize that they exist.

The last major finding to come from these results comes from outcome four. Outcome four deals with the victim not reporting a crime due to his/her belief that the police will not help his/her situation. For this outcome, the Hispanic race was significant and the coefficient was positive. This indicates that Hispanics cite "police not helping" as a reason for not reporting at a higher rate relative to the base outcome. This finding is important because it sheds light on a community of people who do not feel that the police adequately serve them. If underreporting is to be addressed by policymakers, they must tackle sensitive issues like these. While I cannot recommend a specific policy, policymakers must work to help mend issues between ethnic communities and law enforcement. These findings contain the most prevalent implications from the multinomial logit section of my project.

VII. Conclusion

The most prevalent results of the initial probit regression suggest that burglary, "other" crimes, knowing the perpetrator, age, gender, marital status, and income are the most important factors in crime underreporting. While "burglary", "other" crimes, being male, and being married increase the chance that one reports a crime, being older decreases the chance that one reports a crime. The effect for income is unclear, as the "unknown" category is significant. The probit with IPW provides support for the conclusions about burglary, gender, marital status, and income from the initial probit regression. However, it does not provide support for the significance of "other" crimes or age. It also lends support to the significance of other variables. These variables are the ones for Hispanic and "other" race. This model illustrates that being of either of these races leads a victim to report crimes at a lower rate. The IPW model successfully illustrates the importance of selection correction. The implications of all of these findings are discussed in detail in the Results/Implications section.

The most significant results of the multinomial logit regression suggest several relationships between underreporting and reasoning, including a link between assaults, knowing the perpetrator, and dealing with the crime another way, a link between older age and thinking that a crime is too insignificant to report, and a link between being Hispanic and not reporting a crime because the police would not help the situation. This information should serve as a vital tool to policymakers who are looking to combat the issue of crime underreporting. If they target these areas with effective policy, these policymakers should be able to make a significant difference.

There is considerable room for other researchers to expand upon my results. First of all, they could address some of the other reasons that were described by outcome five in my multinomial logit model. If a researcher had the resources to personally interview victims and find out what these other reasons might mean, they could find more valuable information on crime underreporting. Secondly, a researcher might try to find more data on sexual assault and rape victims. My dataset was limited in this respect, and I had to include sexual assaults as a part of the "assault" variable. As one of the more prevalent issues facing society today, a comprehensive sexual assault study would add greatly onto my comprehensive study. Lastly, it would be interesting to see the other techniques that researchers could use to analyze the data. While I trust my methods, I believe that other techniques could be used effectively to analyze this data. I greatly encourage other researchers to build upon my results.

VIII. Works Cited

Allen, W. David. "The Reporting and Underreporting of Rape." *Southern Economic Journal* (2011): Southern Economic Association. Web. 14 Nov. 2014.

Ciofi Degli Atti, M. L., M. C. Rota, D. Mandolini, A. Bella, G. Gabutti, P. Crovari, and S. Salmaso. “Assessment of Varicella Underreporting in Italy.” *Epidemiology and Infection* (2002): Cambridge University Press. Web. 18 Mar. 2015.

Fu, Haishan, Jacqueline E. Darroch, Stanley K. Henshaw, and Elizabeth Kolb. "Measuring the Extent of Abortion Underreporting in the 1995 National Survey of Family Growth." *Family Planning Perspectives* (1998): Guttmacher Institute. Web. 18 Mar. 2015.

Goodkind, Daniel M. “China’s Underreporting, Fertility, and Sex Ratio Imbalance in China.” *Demography* (2011): Springer. Web. 18 Mar. 2015.

Heitman, B. L. and L. Lissner. "Dietary underreporting by obese individuals—is it specific or non-specific?" *BMJ* (1995): BMJ Group. Web. 18 Mar. 2015.

Holland, Daniel M. "Dividend Underreporting on Tax Returns." *The Journal of Finance* (1958): Wiley. Web. 18 Mar. 2015.

Jones, Elise F., and Jacqueline Darroch Forrest "Underreporting of Abortion in Surveys of U.S. Women: 1976 to 1988." *Demography* (1992): Springer. Web. 18 Mar. 2015.

Jones, Rachel K., and Kathryn Kost. "Underreporting of Induced and Spontaneous Abortion in the United States: An Analysis of the 2002 National Survey of Family Growth." *Studies in Family Planning* (2007): Population Council. Web. 18 Mar. 2015.

Leonard, A.E. " Crime Reporting as a Police Management Tool." *Annals of the American Academy of Political and Social Science* (1954): Sage Publications. Web. 14 Nov. 2014.

Levitt, Steven D. " The Relationship Between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports." *Journal of Quantitative Criminology* (1998): Journal of Quantitative Criminology. Web. 14 Nov. 2014.

Martinelli, César, and Susan Wendy Parker. "Deception and Misreporting in a Social Program." *Journal of the European Economic Association* (2009): Wiley. Web. 18 Mar. 2015.

Moreau, Caroline, Nathalie Bajos, Jean Bouyer, and The COCON Group. "Question Comprehension and Recall: The Reporting of Induced Abortions in Quantitative Surveys on the General Population." *Population* (2002): Institut National dÈtudes Démographiques. Web. 18 Mar. 2015.

"National Crime Victimization Survey, 2013." ICPSR. University of Michigan, n.d. Web. 14 Nov. 2014.

Philipov, D., E. Andreev, T. Kharkova, and V. Shkolnikov. "Induced Abortion in Russia: Recent Trends and Underreporting in Surveys." *European Journal of Population* (2004): Springer. Web. 18 Mar. 2015.

Soares, Rodrigo R. " Crime Reporting as a Measure of Institutional Development." *Economic Development and Cultural Change* (2004): The University of Chicago Press. Web. 14 Nov. 2014.

"Idea behind inverse probability weighting." *ESRC* (2004): London School of Hygiene & Tropical Medicine. Web. 7 Apr. 2015.

Udry, J. Richard, Monica Gaughan, Pamela J. Schwingl, Bea J. van den Berg. "A Medical Record Linkage Analysis of Abortion Underreporting." *Family Planning Perspectives* (1996): Guttmacher Institute. Web. 18 Mar. 2015.

Van Hest, N. A. H., F. Smit, and J. P. Verhave. "Underreporting of Malaria Incidence in the Netherlands: Results from a Capture-Recapture Study." *Epidemiology and Infection* (2002): Cambridge University Press. Web. 18 Mar. 2015.

IX. Appendix

Table 1: Description of Variables

Dependent Variable	*Survey Questions/Definition of Variables*
Crime Not Reported	• Were the police informed or did they find out about this incident in any way? • = 1 if a crime was not reported to police, = 0 if a crime was reported to police
Reason Not Reported	• Which of these would you say was the most important reason why the incident was not reported to the police? • = 0 if a crime was reported to police, = 1 if a crime was not reported because the respondent dealt with it another way, = 2 if a crime was not reported because the respondent felt that it was not important enough to report, = 3 if a crime was not reported because the respondent felt that the police could not do anything about it, = 4 if a crime was not reported because the respondent felt that the police wouldn't help, = 5 if the crime was not reported for another reason other than the ones specified

Independent Variable	*Survey Questions/Definition of Variables*
Crime Types	• "As I go through them, tell me if any of these happened to you in the last 6 months…" • = 1 if any of these types of incidents happened, = 0 if any of these types of incidents did not happen
Theft	• "Was something belonging to YOU stolen, such as: Things that you carry, like luggage, a wallet, purse, briefcase, book; Clothing, jewelry, or cellphone; Bicycle or sports equipment; Things in your home - like a TV, stereo, or tools; Things outside your home such as a garden hose or lawn furniture; Things belonging to children in the household; Things from a vehicle, such as a package, groceries, camera, or CDs; or did anyone ATTEMPT to steal anything belonging to you?"

Burglary	• "(Other than any incidents already mentioned,) has anyone: Broken in or ATTEMPTED to break into your home by forcing a door or window, pushing past someone, jimmying a lock, cutting a screen, or entering through an open door or window? Has anyone illegally gotten in or tried to get into a garage, shed, or storage room? Has anyone illegally gotten in or tried to get into a hotel or motel room or vacation home where you were staying?"
Auto Theft	• "(Other than any incidents already mentioned,) (was the vehicle/were any) of the vehicle(s) stolen or used without permission? Did anyone steal any parts such as a tire, car stereo, hubcap, or battery? Did anyone steal any gas from (it/them)? Did anyone ATTEMPT to steal any vehicle or parts attached to (it/them)?"
Assault	• "(Other than any incidents already mentioned,) were you attacked or threatened or did you have something stolen from you: At home including the porch or yard; At or near a friend's, relative's, or neighbor's home; At work or school; In places such as a storage shed or laundry room, a shopping mall, restaurant, bank, or airport; While riding in any vehicle; On the street or in a parking lot; At such places as a party, theater, gym, picnic area, bowling lanes, or while fishing or hunting; or did anyone attempt to attack or attempt to steal anything belonging to you from any of these places?" • "(Other than any incidents already mentioned,) have you been forced or coerced to engage in unwanted sexual activity by someone you didn't know, a casual acquaintance, or someone you know well?" • This variable also includes the questioning from "Weapon Used" and "Attacker Known". See these variables for clarification on the survey questions.
Other	• "During the last 6 months, (other than any incidents already mentioned,) did you call the police to report something that happened to you, which you thought was a crime?" • "During the last 6 months, (other than any incidents already mentioned,) did anything which you thought was a crime happen to you, but you did not report to the police?"

Other Crime Factors	• "As I go through them, tell me if any of these happened to you in the last 6 months…" • = 1 if any of these types of incidents happened, = 0 if this any of these types of incidents did not happen
Weapon Used	• "(Other than any incidents already mentioned,) has anyone attacked or threatened you in any of these ways: With any weapon, for instance, a gun or knife; With anything like a baseball bat, frying pan, scissors, or stick; By something thrown, such as a rock or bottle; Include any grabbing, punching, or choking; Any rape, attempted rape or other type of sexual attack; Any face to face threats; Any attack or threat or use of force by anyone at all?"
Attacker Known	• "(Other than any incidents already mentioned,) did you have something stolen from you or were you attacked or threatened by someone at work or school, a neighbor or friend, a relative or family member, or any other person you've met or known?"
Cultural/Demographic Factors	
Age	• "Determined by asking the respondent for month, day, and year of birth." • Coded numerically in the data
Gender	• Determined by asking the respondent his/her • = 1 if respondent is male, = 0 if respondent is female
Race	• Race in dataset: "Respondent's self-reported race. The five categories are: White, Black, American Indian/Aleut/Eskimo, Asian/Pacific Islander, and Other." • I use the Race variable in the dataset as well as the Ethnicity variable to create my Race variable. • Ethnicity in dataset: "A household respondent's self-reported statement of the national, cultural, or linguistic group with which each member of the household identifies. Currently, the NCVS gathers data on Hispanic ethnicity only." • = 0 (White, reference group) if respondent's race is White and ethnicity is not Hispanic, = 1 (Black) if respondent's race is Black, = 2 (Hispanic) if respondent's race is White and ethnicity is Hispanic, = 3 if (Other) if respondent has any other racial/ethnicity

Marital Status	• "The NCVS defines five categories of marital status: married, widowed, divorced, separated, never married." • = 1 if respondent is married, = 0 if respondent is not married
Income	• "The sum of income received by all household members (14 years of age or older) living in a sample housing unit. The income may include wages, salaries, net income from business, farm or rent, pension, dividends, interest, social security payments, alimony, public assistance, child support, and any other money received by household members age 14 or older, and other money income (not identified as income are such things as room and board, insurance payments, lump sum inheritances, occasional gifts, money from selling property, withdrawals from savings accounts, or tax refunds) for the 12-month period immediately preceding the month of interview." • = 0 if household income is less than $25,000, = 1 if household income is between $25,000 and $75,000 (reference group), = 2 if household income is $75,000, = 3 if household income is unknown/not reported
College Graduate	• "Respondents are asked the highest grade or year of regular school attended, and whether that year was completed. Regular schools include all graded public, private, parochial schools, colleges, universities, and professional schools, which advance a person toward an elementary or high school diploma, or a college degree. Regular schools do not include vocational, trade, business, correspondence, or other specialized schools, unless credits obtained are accepted in the regular school system. For persons still attending regular school, highest grade attended is the one in which they are currently enrolled." • = 1 if respondent has received an associate's degree or any higher level of education, = 0 if respondent has received less than an associate's degree.

Source: National Crime Victimization Survey Codebook, 2013 (ICPSR)

Table 2: Summary Statistics for Entire Sample

Variable	*Non-Victims*	*Victims*	*Total Sample*
Theft	--	45.2%	4.3%
Burglary	--	9.3%	1.0%
Auto Theft	--	12.8%	1.2%
Assault	--	16.4%	1.8%
Other	--	36.0%	4.4%
Weapon Used	--	4.7%	0.5%
Attacker Known	--	3.2%	0.4%
Age***	50.8 (0.63)	44.8 (0.19)	50.3 (0.60)
Male Gender	44.7%	44.8%	44.7%
White Race***	72.3%	65.7%	71.8%
Black Race***	10.4%	12.7%	10.6%
Hispanic Race***	11.2%	14.1%	11.4%
Other Race***	6.1%	7.5%	6.2%
Married***	53.1%	42.1%	52.2%
Income <$25,000***	13.9%	20.6%	14.5%
Income $25-75,000**	33.4%	34.8%	33.5%
Income >$75,000***	21.7%	18.9%	21.4%
Income Unknown***	31.1%	25.7%	30.6%
College Graduate	41.2%	40.4%	41.2%
N =	***74,963***	***6,912***	***81,875***
%	***91.6***	***8.4***	***100.0***

I used two-tailed t-tests to compare the characteristics of crime victims and non-victims.
Standard deviations are in parentheses.
***Indicates significance at .01 level
**Indicates significance at .05 level
*Indicates significance at .10 level

Table 3: Summary Statistics for Crime Victims

Variable	*Report*	*Not Report*	*Total Sample*
Theft***	43.2%	46.5%	45.2%
Burglary***	11.3%	7.9%	9.3%
Auto Theft	12.8%	12.7%	12.8%
Assault	16.1%	16.6%	16.4%
Other**	37.5%	35.0%	36.0%
Weapon Used	4.8%	4.6%	4.7%
Attacker Known*	2.7%	3.5%	3.2%
Age	44.5 (0.30)	45.1 (0.25)	44.8 (0.19)
Male Gender*	46.2%	43.8%	44.8%
White Race	65.9%	65.5%	65.7%
Black Race*	13.6%	12.1%	12.7%
Hispanic Race	13.4%	14.5%	14.1%
Other Race	7.0%	7.9%	7.5%
Married*	43.4%	41.3%	42.1%
Income <$25,000**	19.4%	21.4%	20.6%
Income $25-75,000	33.7%	35.4%	34.8%
Income >$75,000	18.8%	19.0%	18.9%
Income Unknown***	28.0%	24.1%	25.7%
College Graduate	41.5%	39.6%	40.4%
N =	***2,820***	***4,092***	***6,912***
%	***40.8***	***59.2***	***100.0***

I used two-tailed t-tests to compare the characteristics of crime reporters and non-reporters.
Standard deviations are in parentheses.
***Indicates significance at .01 level
**Indicates significance at .05 level
*Indicates significance at .10 level

Table 4: Distribution of Underreporting Reasons Variable

Reason	*Distribution (%)*	*N =*
Reported to Police	41.1	2,820
Dealt with Another Way	7.1	487
Not Important Enough to Report	18.9	1,300
Police Couldn't Do Anything	9.4	645
Police Wouldn't Help	13.2	905
Not Reported for Another Reason	10.3	707

Table 5: Probit Models with Marginal Effects

Variable	*Probit* *dy/dx*	*Probit with IPW* *dy/dx*
Theft	0.010 (0.016)	0.010 (0.017)
Burglary	-0.113*** (0.022)	-0.107*** (0.024)
Assault	-0.011 (0.022)	-0.004 (0.024)
Other	-0.035** (0.017)	-0.026 (0.018)
Weapon Used	-0.006 (0.033)	-0.015 (0.036)
Attacker Known	0.070* (0.036)	0.060 (0.039)
Age	0.001** (0.000)	0.001 (0.000)
Male Gender	-0.026** (0.012)	-0.030** (0.013)
Black Race	-0.026 (0.019)	-0.024 (0.020)
Hispanic Race	0.029 (0.018)	0.044** (0.019)
Other Race	0.032 (0.023)	0.049** (0.024)
Married	-0.026** (0.013)	-0.030** (0.014)
Income <$25,000	0.003 (0.017)	0.007 (0.018)
Income >$75,000	-0.002 (0.018)	-0.002 (0.019)
Income Unknown	-0.053*** (0.016)	-0.059*** (0.017)
College Graduate	-0.021 (0.013)	-0.002 (0.014)
Dependent Variable: Crime Not Reported to Police Probit: N = 6,912; $P > X^2 = 0.000$; Pseudo $R^2 = 0.008$ Probit with IPW: N = 6,912; $P > X^2 = 0.000$; Pseudo $R^2 = 0.008$		

Robust standard errors are in parentheses.
***Indicates significance at .01 level
**Indicates significance at .05 level
*Indicates significance at .10 level

Table 6: Multinomial Logit Model with Coefficients

	Underreporting Reasons				
Variable	***Dealt with Another Way***	***Not Important Enough to Report***	***Police Couldn't Do Anything***	***Police Wouldn't Help***	***Not Reported for Another Reason***
Theft	-0.102 (0.119)	-0.155* (0.091)	0.029 (0.123)	0.326*** (0.102)	0.072 (0.106)
Burglary	-0.997*** (0.224)	-0.523*** (0.128)	-0.438*** (0.168)	-0.218 (0.134)	-0.403*** (0.153)
Assault	0.396** (0.160)	-0.303** (0.133)	-0.309 (0.174)	-0.002 (0.141)	0.127 (0.152)
Other	-0.056 (0.130)	-0.285*** (0.098)	-0.212 (0.131)	-0.034 (0.111)	-0.083 (0.115)
Weapon Used	0.154 (0.225)	-0.003 (0.203)	-0.381 (0.297)	-0.140 (0.223)	0.181 (0.214)
Attacker Known	0.713*** (0.239)	0.166 (0.239)	-0.514 (0.397)	-0.076 (0.270)	0.677*** (0.230)
Age	-0.000 (0.003)	0.007*** (0.002)	0.003 (0.003)	0.001 (0.002)	0.003 (0.003)
Male Gender	-0.352*** (0.101)	-0.024 (0.068)	-0.062 (0.089)	-0.037 (0.078)	-0.206** (0.086)
Black Race	0.016 (0.147)	-0.126 (0.108)	-0.408*** (0.150)	0.033 (0.117)	-0.133 (0.132)
Hispanic Race	-0.421** (0.180)	0.062 (0.103)	0.014 (0.132)	0.339*** (0.110)	0.294** (0.122)
Other Race	0.297* (0.177)	0.094 (0.131)	0.191 (0.162)	0.156 (0.149)	0.056 (0.169)
Marital Status	-0.135 (0.107)	-0.084 (0.072)	-0.031 (0.096)	-0.229*** (0.085)	-0.080 (0.090)
Income <$25,000	-0.024 (0.145)	-0.153 (0.097)	0.034 (0.121)	0.151 (0.107)	0.144 (0.118)
Income >$75,000	0.092 (0.146)	0.118 (0.098)	-0.126 (0.132)	-0.194 (0.121)	-0.019 (0.129)
Income Unknown	-0.064 (0.129)	-0.304*** (0.089)	-0.411*** (0.118)	-0.179* (0.121)	-0.085 (0.111)
College Graduate	-0.041 (0.106)	-0.107 (0.075)	-0.219** (0.095)	-0.014 (0.083)	-0.039 (0.090)
Coefficient	-1.486*** (0.218)	-0.656*** (0.152)	-1.150*** (0.194)	-1.179*** (0.180)	-1.455*** (0.190)

Base Outcome for Dependent Variable: Crime Reported to Police
N = 6,864; $P > X^2 = 0.000$; Pseudo $R^2 = 0.013$

Robust standard errors are in parentheses.
***Indicates significance at .01 level
**Indicates significance at .05 level
*Indicates significance at .10 level

Table 7: Multinomial Logit Model with Relative Risk Ratios

	Underreporting Reasons				
Variable	***Dealt with Another Way***	***Not Important Enough to Report***	***Police Couldn't Do Anything***	***Police Wouldn't Help***	***Not Reported for Another Reason***
Theft	0.903	0.856*	1.030	1.385***	1.075
Burglary	0.369***	0.593***	0.646***	0.804	0.668***
Assault	1.486**	0.739**	0.734	1.002	1.136
Other	0.946	0.752***	0.809	0.966	0.921
Weapon Used	1.167	0.997	0.683	0.869	1.198
Attacker Known	2.040***	1.181	0.598	0.927	1.968***
Age	1.000	1.007***	1.003	1.001	1.003
Male Gender	0.703***	0.976	0.940	0.963	0.814**
Black Race	1.016	0.882	0.665***	1.034	0.875
Hispanic Race	0.657**	1.064	1.014	1.403***	1.342**
Other Race	1.346*	1.098	1.211	1.169	1.057
Married	0.874	0.920	0.969	0.796***	0.923
Income <$25,000	0.977	0.858	1.034	1.163	1.155
Income >$75,000	1.097	1.125	0.882	0.824	0.981
Income Unknown	0.938	0.738***	0.663***	0.836*	0.918
College Graduate	0.960	0.898	0.803**	0.986	0.962
Coefficient	0.226***	0.519***	0.317***	0.308***	0.233***

Base Outcome for Dependent Variable: Crime Reported to Police
N = 6,864; $P > X^2 = 0.000$; Pseudo $R^2 = 0.013$

Robust standard errors are in parentheses.
***Indicates significance at .01 level
**Indicates significance at .05 level
*Indicates significance at .10 level